

ENCYCLOPEDIA OF

MEDICAL DEVICES AND INSTRUMENTATION

Second Edition

VOLUME 6

Radiotherapy, Heavy Ion – X-Rays, Production of

ENCYCLOPEDIA OF MEDICAL DEVICES AND INSTRUMENTATION, SECOND EDITION

Editor-in-Chief

John G. Webster

University of Wisconsin–Madison

Editorial Board

David Beebe

University of Wisconsin–Madison

Jerry M. Calkins

University of Arizona College of Medicine

Michael R. Neuman

Michigan Technological University

Joon B. Park

University of Iowa

Edward S. Sternick

Tufts–New England Medical Center

Editorial Staff

Vice President, STM Books: **Janet Bailey**

Associate Publisher: **George J. Telecki**

Editorial Director: **Sean Pidgeon**

Director, Book Production and Manufacturing:

Camille P. Carter

Production Manager: **Shirley Thomas**

Illustration Manager: **Dean Gonzalez**

Senior Production Editor: **Kellsee Chu**

Editorial Program Coordinator: **Surlan Murrell**

ENCYCLOPEDIA OF

MEDICAL DEVICES AND INSTRUMENTATION

Second Edition
Volume 6

Radiotherapy, Heavy Ion – X-Rays, Production of

Edited by

John G. Webster

University of Wisconsin–Madison

The *Encyclopedia of Medical Devices and Instrumentation* is available online at
<http://www.mrw.interscience.wiley.com/emdi>

 **WILEY-INTERSCIENCE**

A John Wiley & Sons, Inc., Publication

Copyright © 2006 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222, Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Encyclopedia of medical devices & instrumentation/by John G. Webster,

editor in chief. – 2nd ed.

p. ; cm.

Rev. ed. of: Encyclopedia of medical devices and instrumentation. 1988.

Includes bibliographical references and index.

ISBN-13 978-0-471-26358-6 (set : cloth)

ISBN-10 0-471-26358-3 (set : cloth)

ISBN-13 978-0-470-04071-3 (v. 6 : cloth)

ISBN-10 0-470-04071-8 (v. 6 : cloth)

1. Medical instruments and apparatus—Encyclopedias. 2. Biomedical engineering—Encyclopedias. 3. Medical physics—Encyclopedias. 4. Medicine—Data processing—Encyclopedias. I. Webster, John G., 1932- . II. Title: Encyclopedia of medical devices and instrumentation.

[DNLM: 1. Equipment and Supplies—Encyclopedias—English. W 13

E555 2006]

R856.A3E53 2006

610.2803—dc22

2005028946

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

CONTRIBUTOR LIST

- ABDEL HADY, MAZEN**, *McMaster University, Hamilton, Ontario Canada*, Bladder Dysfunction, Neurostimulation of
- ABEL, L.A.**, *University of Melbourne, Melbourne, Australia*, Ocular Motility Recording and Nystagmus
- ABREU, BEATRIZ C.**, *Transitional Learning Center at Galveston, Galveston, Texas*, Rehabilitation, Computers in Cognitive
- ALEXANDER, A.L.**, *University of Wisconsin–Madison, Madison, Wisconsin*, Magnetic Resonance Imaging
- ALI, ABBAS**, *University of Illinois, at Urbana-Champaign, Bioinformatics*
- ALI, MÜFTÜ**, *School of Dental Medicine, Boston, Massachusetts*, Tooth and Jaw, Biomechanics of
- ALPERIN, NOAM**, *University of Illinois at Chicago, Chicago, Illinois*, Hydrocephalus, Tools for Diagnosis and Treatment of
- ANSON, DENIS**, *College Misericordia, Dallas, Pennsylvania*, Environmental Control
- ARENA, JOHN C.**, *VA Medical Center and Medical College of Georgia*, Biofeedback
- ARIEL, GIDEON**, *Ariel Dynamics, Canyon, California*, Biomechanics of Exercise Fitness
- ARMSTRONG, STEVE**, *University of Iowa, Iowa City, Iowa*, Biomaterials for Dentistry
- ASPDEN, R.M.**, *University of Aberdeen, Aberdeen, United Kingdom*, Ligament and Tendon, Properties of
- AUBIN, C.E.**, *Polytechnique Montreal, Montreal Quebec, Canada*, Scoliosis, Biomechanics of
- AYRES, VIRGINIA M.**, *Michigan State University, East Lansing, Michigan*, Microscopy, Scanning Tunneling
- AZANGWE, G.**, Ligament and Tendon, Properties of
- BACK, LLOYD H.**, *California Institute of Technology, Pasadena, California*, Coronary Angioplasty and Guidewire Diagnostics
- BADYLAK, STEPHEN F.**, *McGowan Institute for Regenerative Medicine, Pittsburgh, Pennsylvania*, Sterilization of Biologic Scaffold Materials
- BANDYOPADHYAY, AMIT**, *Washington State University, Pullman, Washington*, Orthopedic Devices, Materials and Design for
- BANERJEE, RUPAK K.**, *University of Cincinnati, Cincinnati, Ohio*, Coronary Angioplasty and Guidewire Diagnostics
- BARBOUR, RANDALL L.**, *State University of New York Downstate Medical Center, Brooklyn, New York*, Peripheral Vascular Noninvasive Measurements
- BARKER, STEVEN J.**, *University of Arizona, Tucson, Arizona*, Oxygen Monitoring
- BARTH, ROLF F.**, *The Ohio State University, Columbus, Ohio*, Boron Neutron Capture Therapy
- BECCHETTI, F.D.**, *University of Michigan, Ann Arbor, Michigan*, Radiotherapy, Heavy Ion
- BELFORTE, GUIDO**, *Politecnico di Torino – Department of Mechanics*, Laryngeal Prosthetic Devices
- BENKESER, PAUL**, *Georgia Institute of Technology, Atlanta, Georgia*, Biomedical Engineering Education
- BENNETT, JAMES R.**, *University of Iowa, Iowa City, Iowa*, Digital Angiography
- BERSANO-BEGEY, TOMMASO**, *University of Michigan, Ann Arbor, Michigan*, Microbioreactors
- BIGGS, PETER J.**, *Harvard Medical School, Boston, Massachusetts*, Radiotherapy, Intraoperative
- BIYANI, ASHOK**, *University of Toledo, and Medical College of Ohio, Toledo, Ohio*, Human Spine, Biomechanics of
- BLOCK, W.F.**, *University of Wisconsin–Madison, Madison, Wisconsin*, Magnetic Resonance Imaging
- BLUE, THOMAS E.**, *The Ohio State University, Columbus, Ohio*, Boron Neutron Capture Therapy
- BLUMSACK, JUDITH T.**, *Disorders Auburn University, Auburn, Alabama*, Audiometry
- BOGAN, RICHARD K.**, *University of South Carolina, Columbia, South Carolina*, Sleep Laboratory
- BOKROS, JACK C.**, *Medical Carbon Research Institute, Austin, Texas*, Biomaterials, Carbon
- BONGIOANNINI, GUIDO**, *ENT Division Mauriziano Hospital, Torino, Italy*, Laryngeal Prosthetic Devices
- BORAH, JOSHUA**, *Applied Science Laboratories, Bedford, Massachusetts*, Eye Movement, Measurement Techniques for
- BORDEN, MARK**, *Director of Biomaterials Research, Irvine, California*, Biomaterials, Absorbable
- BORTON, BETTIE B.**, *Auburn University Montgomery, Montgomery, Alabama*, Audiometry
- BORTON, THOMAS E.**, *Auburn University Montgomery, Montgomery, Alabama*, Audiometry
- BOSE SUSMITA.**, *Washington State University, Pullman, Washington*, Orthopedic Devices, Materials and Design for
- BOVA, FRANK J.**, *M. D. Anderson Cancer Center Orlando, Orlando, FL*, Radiosurgery, Stereotactic
- BRENNER, DAVID J.**, *Columbia University Medical Center, New York, New York*, Computed Tomography Screening
- BREWER, JOHN M.**, *University of Georgia*, Electrophoresis
- BRIAN, L. DAVIS**, *Lerner Research Institute, The Cleveland Clinic Foundation, Cleveland, Ohio*, Skin, Biomechanics of
- BRITT, L.D.**, *Eastern Virginia Medical School, Norfolk, Virginia*, Gastrointestinal Hemorrhage
- BRITT, R.C.**, *Eastern Virginia Medical School, Norfolk, Virginia*, Gastrointestinal Hemorrhage
- BROZIK, SUSAN M.**, *Sandia National Laboratories, Albuquerque, New Mexico*, Microbial Detection Systems
- BRUNER, JOSEPH P.**, *Vanderbilt University Medical Center, Nashville, Tennessee*, Intrauterine Surgical Techniques
- BRUNSWIG NEWRING, KIRK A.**, *University of Nevada, Reno, Nevada*, Sexual Instrumentation
- BRUYANT, PHILIPPE P.**, *University of Massachusetts, North Worcester, Massachusetts*, Nuclear Medicine, Computers in
- BUNNELL, BERT J.**, *Bunnell Inc., Salt Lake City, Utah*, High Frequency Ventilation
- CALKINS, JERRY M.**, *Defense Research Technologies, Inc., Rockville, Maryland*, Medical Gas Analyzers
- CANNON, MARK**, *Northwestern University, Chicago, Illinois*, Resin-Based Composites

- CAPPELLERI, JOSEPH C.**, *Pfizer Inc., Groton, Connecticut*, Quality-of-Life Measures, Clinical Significance of
- CARDOSO, JORGE**, *University of Madeira, Funchal, Portugal*, Office Automation Systems
- CARELLO, MASSIMILIANA**, *Politecnico di Torino – Department of Mechanics, Laryngeal Prosthetic Devices*
- CASKEY, THOMAS C.**, *Cogene Biotech Ventures, Houston, Texas*, Polymerase Chain Reaction
- CECCIO, STEVEN**, *University of Michigan, Ann Arbor, Michigan*, Heart Valve Prostheses, In Vitro Flow Dynamics of
- CHAN, JACKIE K.**, *Columbia University, New York, New York*, Photography, Medical
- CHANDRAN, K.B.**, *University of Iowa, Iowa City, Iowa*, Heart Valve Prostheses
- CHATZANDROULIS, S.**, *NTUA, Athens, Attiki, Greece*, Capacitive Microsensors for Biomedical Applications
- CHAVEZ, ELIANA**, *University of Pittsburgh, Pittsburgh, Pennsylvania*, Mobility Aids
- CHEN, HENRY**, *Stanford University, Palo Alto, California*, Exercise Stress Testing
- CHEN, JIANDE**, *University of Texas Medical Branch, Galveston, Texas*, Electrogastrogram
- CHEN, YAN**, *Lerner Research Institute, The Cleveland Clinic Foundation, Cleveland, Ohio*, Skin, Biomechanics of
- CHEYNE, DOUGLAS**, *Hospital for Sick Children Research Institute, Biomagnetism*
- CHUI, CHEN-SHOU**, *Memorial Sloan-Kettering Cancer Center, New York, New York*, Radiation Therapy Treatment Planning, Monte Carlo Calculations in
- CLAXTON, NATHAN S.**, *The Florida State University, Tallahassee, Florida*, Microscopy, Confocal
- CODERRE, JEFFREY A.**, *Massachusetts Institute of Technology, Cambridge, Massachusetts*, Boron Neutron Capture Therapy
- COLLINS, BETH**, *University of Mississippi Medical Center, Jackson, Mississippi*, Hyperbaric Medicine
- COLLINS, DIANE**, *University of Pittsburgh, Pittsburgh, Pennsylvania*, Mobility Aids
- CONSTANTINOU, C.**, *Columbia University Radiation Oncology, New York, New York*, Phantom Materials in Radiology
- COOK, ALBERT**, *University of Alberta, Edmonton, Alberta, Canada*, Communication Devices
- COOPER, RORY**, *University of Pittsburgh, Pittsburgh, Pennsylvania*, Mobility Aids
- CORK, RANDALL C.**, *Louisiana State University, Shreveport, Louisiana*, Monitoring, Umbilical Artery and Vein, Blood Gas Measurements; Transcutaneous Electrical Nerve Stimulation (TENS); Ambulatory Monitoring
- COX, JOSEPHINE H.**, *Walter Reed Army Institute of Research, Rockville, Maryland*, Blood Collection and Processing
- CRAIG, LEONARD**, *Feinberg School of Medicine of Northwestern University, Chicago, Illinois*, Ventilators, Acute Medical Care
- CRESS, CYNTHIA J.**, *University of Nebraska, Lincoln, Nebraska*, Communicative Disorders, Computer Applications for
- CUMMING, DAVID R.S.**, *University of Glasgow, Glasgow, United Kingdom*, Ion-Sensitive Field-Effect Transistors
- CUNNINGHAM, JOHN R.**, *Camrose, Alberta, Canada*, Cobalt 60 Units for Radiotherapy
- D'ALESSANDRO, DAVID**, *Montefiore Medical Center, Bronx, New York*, Heart-Lung Machines
- D'AMBRA, MICHAEL N.**, *Harvard Medical School, Cambridge, Massachusetts*, Cardiac Output, Thermodilution Measurement of
- DADSETAN, MAHROKH**, *Mayo Clinic, College of Medicine, Rochester, Minnesota*, Microscopy, Electron
- DALEY, MICHAEL L.**, *The University of Memphis, Memphis, Tennessee*, Monitoring, Intracranial Pressure
- DAN, LOYD**, *Linköping University, Linköping, Sweden*, Thermocouples
- DAS, RUPAK**, *University of Wisconsin, Madison, Wisconsin*, Brachytherapy, High Dosage Rate
- DATTAWADKAR, AMRUTA M.**, *University of Wisconsin, Madison, Madison, Wisconsin*, Ocular Fundus Reflectometry
- DAVIDSON, MICHAEL W.**, *The Florida State University, Tallahassee, Florida*, Microscopy, Confocal
- DE LUCA, CARLO**, *Boston University, Boston, Massachusetts*, Electromyography
- DE SALLES, ANTONIO A.F.**, *UCLA Medical School, Los Angeles, California*, Stereotactic Surgery
- DECAU, SABIN**, *University of Maryland, School of Medicine*, Shock, Treatment of
- DECHOW, PAUL C.**, *A & M University Health Science Center, Dallas, Texas*, Strain Gages
- DELBEKE, JEAN**, *Catholique University of Louvain, Brussels, Belgium*, Visual Prostheses
- DELL'OSSO, LOUIS F.**, *Case Western Reserve University, Cleveland, Ohio*, Ocular Motility Recording and Nystagmus
- DELORME, ARNAUD**, *University of San Diego, La Jolla, California*, Statistical Methods
- DEMENKOFF, JOHN**, *Mayo Clinic, Scottsdale, Arizona*, Pulmonary Physiology
- DEMIR, SEMAHAT S.**, *The University of Memphis and The University of Tennessee Health Science Center, Memphis, Tennessee*, Electrophysiology
- DEMLING, ROBERT H.**, *Harvard Medical School*, Skin Substitute for Burns, Bioactive
- DENNIS, MICHAEL J.**, *Medical University of Ohio, Toledo, Ohio*, Computed Tomography
- DESANTI, LESLIE**, *Harvard Medical School*, Skin Substitute for Burns, Bioactive
- DEUTSCH, STEVEN**, *Pennsylvania State University, University Park, Pennsylvania*, Flowmeters
- DEVINENI, TRISHUL**, *Conemaugh Health System*, Biofeedback
- DI BELLA EDWARD, V.R.**, *University of Utah*, Tracer Kinetics
- DIKIDES, NICHOLAS A.**, *Advanced Concepts Analysis, Inc., Falls Church, Virginia*, Thermography
- DOLAN, PATRICIA L.**, *Sandia National Laboratories, Albuquerque, New Mexico*, Microbial Detection Systems
- DONOVAN, F.M.**, *University of South Alabama*, Cardiac Output, Indicator Dilution Measurement of
- DOUGLAS, WILSON R.**, *Children's Hospital of Philadelphia, Philadelphia, Pennsylvania*, Intrauterine Surgical Techniques
- DRAPER, CRISSA**, *University of Nevada, Reno, Nevada*, Sexual Instrumentation
- DRZEWIECKI, TADEUSZ M.**, *Defense Research Technologies, Inc., Rockville, Maryland*, Medical Gas Analyzers
- DURFEE, W.K.**, *University of Minnesota, Minneapolis, Minnesota*, Rehabilitation and Muscle Testing
- DYRO, JOSEPH F.**, *Setauket, New York*, Safety Program, Hospital

- DYSON, MARY**, *Herts, United Kingdom*, Heat and Cold, Therapeutic
- ECKERLE, JOSEPH S.**, *SRI International, Menlo Park, California*, Tonometry, Arterial
- EDWARDS, BENJAMIN**, *University of Wisconsin-Madison, Madison, Wisconsin*, Codes and Regulations: Radiation
- EDWARDS, THAYNE L.**, *University of Washington, Seattle, Washington*, Chromatography
- EKLUND, ANDERS**, *University of Illinois at Chicago, Chicago, Illinois*, Hydrocephalus, Tools for Diagnosis and Treatment of
- EL SOLH, ALI A.**, *Erie County Medical Center, Buffalo, New York*, Sleep Studies, Computer Analysis of
- ELMAYERGI, NADER**, *McMaster University, Hamilton, Ontario, Canada*, Bladder Dysfunction, Neurostimulation of
- ELSHARYDAH, AHMAD**, *Louisiana State University, Baton Rouge, Louisiana*, Ambulatory Monitoring; Monitoring, Umbilical Artery and Vein, Blood Gas Measurements
- FADDY, STEVEN C.**, *St. Vincents Hospital, Sydney, Darlinghurst, Australia*, Cardiac Output, Fick Technique for
- FAHEY, FREDERIC H.**, *Childrens Hospital Boston*, Computed Tomography, Single Photon Emission
- FAIN, S.B.**, *University of Wisconsin-Madison, Madison, Wisconsin*, Magnetic Resonance Imaging
- FELDMAN, JEFFREY**, *Childrens Hospital of Philadelphia, Philadelphia, Pennsylvania*, Anesthesia Machines
- FELLERS, THOMAS J.**, *The Florida State University, Tallahassee, Florida*, Microscopy, Confocal
- FERRARA, LISA**, *Cleveland Clinic Foundation, Cleveland, Ohio*, Human Spine, Biomechanics of
- FERRARI, MAURO**, *The Ohio State University, Columbus, Ohio*, Drug Delivery Systems
- FONTAINE, ARNOLD A.**, *Pennsylvania State University, University Park, Pennsylvania*, Flowmeters
- FOUST, MILTON J., JR.**, *Medical University of South Carolina Psychiatry and Behavioral Sciences, Charleston, South Carolina*, Electroconvulsive Therapy
- FRASCO, PETER**, *Mayo Clinic Scottsdale, Scottsdale, Arizona*, Temperature Monitoring
- FRAZIER, JAMES**, *Louisiana State University, Baton Rouge, Louisiana*, Ambulatory Monitoring
- FREISLEBEN DE BLASIO, BIRGITTE**, *University of Oslo, Oslo, Norway*, Impedance Spectroscopy
- FRESTA, MASSIMO**, *University of Catanzaro Magna Græcia, Germaneto (CZ), Italy*, Drug Delivery Systems
- FREYTES, DONALD O.**, *McGowan Institute for Regenerative Medicine, Pittsburgh Pennsylvania*, Sterilization of Biologic Scaffold Materials
- FROELICHER, VICTOR**, *VA Medical Center, Palo Alto, California*, Exercise Stress Testing
- FUNG, EDWARD K.**, *Columbia University, New York, New York*, Photography, Medical
- GAGE, ANDREW A.**, *State University of New York at Buffalo, Buffalo, New York*, Cryosurgery
- GAGLIO, PAUL J.**, *Columbia University College of Physicians and Surgeons*, Liver Transplantation
- GARDNER, REED M.**, *LDS Hospital and Utah University, Salt Lake City, Utah*, Monitoring, Hemodynamic
- GEJERMAN, GLEN**, *Hackensack University Medical, Hackensack, New Jersey*, Radiation Therapy, Quality Assurance in
- GEORGE, MARK S.**, *Medical University of South Carolina Psychiatry and Behavioral Sciences, Charleston, South Carolina*, Electroconvulsive Therapy
- GHARIEB, R.R.**, *Infinite Biomedical Technologies, Baltimore, Maryland*, Neurological Monitors
- GLASGOW, GLENN P.**, *Loyola University of Chicago, Maywood, Illinois*, Radiation Protection Instrumentation
- GLASGOW, GLENN**, *University of Wisconsin-Madison, Madison, Wisconsin*, Codes and Regulations: Radiation
- GOEL, VIJAY K.**, *University of Toledo, and Medical College of Ohio, Toledo, Ohio*, Human Spine, Biomechanics of
- GOETSCH, STEVEN J.**, *San Diego Gamma Knife Center, La Jolla, California*, Gamma Knife
- GOLDBERG, JAY R.**, *Marquette University Milwaukee, Wisconsin*, Minimally Invasive Surgery
- GOLDBERG, ZELENNA**, *Department of Radiation Oncology, Davis, California*, Ionizing Radiation, Biological Effects of
- GOPALAKRISHNAKONE, P.**, *National University of Singapore, Singapore*, Immunologically Sensitive Field-Effect Transistors
- GOPAS, JACOB**, *Ben Gurion University of the Negev, Beer Sheva, Israel*, Monoclonal Antibodies
- GORGULHO, ALESSANDRA**, *UCLA Medical School, Los Angeles, California*, Stereotactic Surgery
- GOUGH, DAVID A.**, *University of California, La Jolla, California*, Glucose Sensors
- GOUSTOURIDIS, D.**, *NTUA, Athens, Attiki, Greece*, Capacitive Microsensors for Biomedical Applications
- GRABER, HARRY L.**, *State University of New York Downstate Medical Center, Brooklyn, New York*, Peripheral Vascular Noninvasive Measurements
- GRAÇA, M.**, *Louisiana State University, Baton Rouge, Louisiana*, Boron Neutron Capture Therapy
- GRANT, WALTER III**, *Baylor College of Medicine, Houston, Texas*, Radiation Therapy, Intensity Modulated
- GRAYDEN, EDWARD**, *Mayo Health Center, Albertlea, Minnesota*, Cardiopulmonary Resuscitation
- GREEN, JORDAN R.**, *University of Nebraska, Lincoln, Nebraska*, Communicative Disorders, Computer Applications for
- HAEMMERICH, DIETER**, *Medical University of South Carolina, Charleston, South Carolina*, Tissue Ablation
- HAMAM, HABIB**, *Université de Moncton, Moncton New Brunswick, Canada*, Lenses, Intraocular
- HAMMOND, PAUL A.**, *University of Glasgow, Glasgow, United Kingdom*, Ion-Sensitive Field-Effect Transistors
- HANLEY, JOSEPH**, *Hackensack University Medical, Hackensack, New Jersey*, Radiation Therapy, Quality Assurance in
- HARLEY, BRENDAN A.**, *Massachusetts Institute of Technology*, Skin Tissue Engineering for Regeneration
- HARPER, JASON C.**, *Sandia National Laboratories, Albuquerque, New Mexico*, Microbial Detection Systems
- HASMAN, ARIE**, *Maastricht, The Netherlands*, Medical Education, Computers in
- HASSOUNA, MAGDY**, *Toronto Western Hospital, Toronto, Canada*, Bladder Dysfunction, Neurostimulation of
- HAYASHI, KOZABURO**, *Okayama University of Science, Okayama, Japan*, Arteries, Elastic Properties of
- HENCH, LARRY L.**, *Imperial College London, London, United Kingdom*, Biomaterials: Bioceramics
- HETRICK, DOUGLAS A., Sr.** *Principal Scientist Medtronic, Inc., Minneapolis, Minnesota*, Bioimpedance in Cardiovascular Medicine
- HIRSCH-KUCHMA, MELISSA**, *University of Central Florida NanoScience Technology Center, Orlando, Florida*, Biosurface Engineering

- HOLDER, GRAHAM E.**, *Moorfields Eye Hospital, London, United Kingdom*, Electroretinography
- HOLMES, TIMOTHY**, *St. Agnes Cancer Center, Baltimore, Maryland*, Tomotherapy
- HONEYMAN-BUCK, JANICE C.**, *University of Florida, Gainesville, Florida*, Radiology Information Systems
- HOOPER, BRETT A.**, *Areté Associates, Arlington, Virginia*, Endoscopes
- HORN, BRUCE**, *Kaiser Permanente, Los Angeles, California*, X-Rays Production of
- HORNER, PATRICIA I.**, *Biomedical Engineering Society Landover, Maryland*, Medical Engineering Societies and Organizations
- HOWITZ, PAUL M.**, *University of Texas, San Antonio, Texas*, Fluorescence Measurements
- HOU, XIAOLIN**, *Risø National Laboratory, Roskilde, Denmark*, Neutron Activation Analysis
- HOVORKA, ROMAN**, *University of Cambridge, Cambridge, United Kingdom*, Pancreas, Artificial
- HUANG, H.K.**, *University of Southern California*, Teleradiology
- HUNT, ALAN J.**, *University of Michigan, Ann Arbor, Michigan*, Optical Tweezers
- HUTTEN, HELMUT**, *University of Technology, Graz, Australia*, Impedance Plethysmography
- LAIZZO, P.A.**, *University of Minnesota, Minneapolis, Minnesota*, Rehabilitation and Muscle Testing
- IBBOTT, GEOFFREY S.**, *Anderson Cancer Center, Houston, Texas*, Radiation Dosimetry, Three-Dimensional
- INGHAM, E.**, *University of Leeds, Leeds, United Kingdom*, Hip Joints, Artificial
- ISIK, CAN**, *Syracuse University, Syracuse, New York*, Blood Pressure Measurement
- JAMES, SUSAN P.**, *Colorado State University, Fort Collins, Colorado*, Biomaterials: Polymers
- JENSEN, WINNIE**, *Aalborg University, Aalborg, Denmark*, Electroneurography
- JIN, CHUNMING**, *North Carolina State University, Raleigh, North Carolina*, Biomaterials, Corrosion and Wear of
- JIN, Z.M.**, *University of Leeds, Leeds, United Kingdom*, Hip Joints, Artificial
- JOHNSON, ARTHUR T.**, *University of Maryland College Park, Maryland*, Medical Engineering Societies and Organizations
- JONES, JULIAN R.**, *Imperial College London, London, United Kingdom*, Biomaterials: Bioceramics
- JOSHI, ABHIJEET**, *Abbott Spine, Austin, Texas*, Spinal Implants
- JUNG, RANU**, *Arizona State University, Tempe, Arizona*, Functional Electrical Stimulation
- JURISSON, SILVIA S.**, *University of Missouri Columbia, Missouri*, Radionuclide Production and Radioactive Decay
- KAEDING, PATRICIA J.**, *Godfrey & Kahn S.C., Madison, Wisconsin*, Codes and Regulations: Medical Devices
- KAMATH, CELIA C.**, *Mayo Clinic, Rochester, Minnesota*, Quality-of-Life Measures, Clinical Significance of
- KANE, MOLLIE**, *Madison, Wisconsin*, Contraceptive Devices
- KATHERINE, ANDRIOLE P.**, *Harvard Medical School, Boston, Massachusetts*, Picture Archiving and Communication Systems
- KATSAGGELOS, AGGELOS K.**, *Northwestern University, Evanston, Illinois*, DNA Sequencing
- KATZ, J. LAWRENCE**, *University of Missouri-Kansas City, Kansas City, Missouri*, Bone and Teeth, Properties of
- KESAVAN, SUNIL**, *Akebono Corporation, Farmington Hills, Michigan*, Linear Variable Differential Transformers
- KHANG, GILSON**, *Chonbuk National University*, Biomaterials: Tissue Engineering and Scaffolds
- KHAODHIAR, LALITA**, *Harvard Medical School, Boston, Massachusetts*, Cutaneous Blood Flow, Doppler Measurement of
- KIM, MOON SUK**, *Korea Research Institutes of Chemical Technology*, Biomaterials: Tissue Engineering and Scaffolds
- KIM, YOUNG KON**, *Inje University, Kimhae City, Korea*, Alloys, Shape Memory
- KINDWALL, ERIC P.**, *St. Luke's Medical Center, Milwaukee, Wisconsin*, Hyperbaric Oxygenation
- KING, MICHAEL A.**, *University of Massachusetts, North Worcester, Massachusetts*, Nuclear Medicine, Computers in
- KLEBE, ROBERT J.**, *University of Texas, San Antonio, Texas*, Fluorescence Measurements
- KLEIN, BURTON**, *Burton Klein Associates, Newton, Massachusetts*, Gas and Vacuum Systems, Centrally Piped Medical
- KNOPER, STEVEN R.**, *University of Arizona College of Medicine*, Ventilatory Monitoring
- KONTAXAKIS, GEORGE**, *Universidad Politécnica de Madrid, Madrid, Spain*, Positron Emission Tomography
- KOTTKE-MARCHANT, KANDICE**, *The Cleveland Clinic Foundation, Cleveland, Ohio*, Vascular Graft Prosthesis
- KRIPFGANS, OLIVER**, *University of Michigan, Ann Arbor, Michigan*, Ultrasonic Imaging
- KULKARNI, AMOL D.**, *University of Wisconsin-Madison, Madison, Wisconsin*, Ocular Fundus Reflectometry, Visual Field Testing
- KUMARADAS, J. CARL**, *Ryerson University, Toronto, Ontario, Canada*, Hyperthermia, Interstitial
- KUNICKA, JOLANTA**, *Bayer HealthCare LLC, Tarrytown, New York*, Differential Counts, Automated
- KWAK, KWANJ JOO**, *University of Miami Miller School of Medicine, Miami, Florida*, Microscopy, Scanning Force
- LAKES, RODERIC**, *University of Wisconsin-Madison*, Bone and Teeth, Properties of
- LAKKIREDDY, DHANUNJAYA**, *The Cleveland Clinic Foundation, Cleveland, Ohio*, Hyperthermia, Ultrasonic
- LARSEN, COBY**, *Case Western Reserve University, Cleveland, Ohio*, Vascular Graft Prosthesis
- LASTER, BRENDA H.**, *Ben Gurion University of the Negev, Beer Sheva, Israel*, Monoclonal Antibodies
- LATTA, LOREN**, *University of Miami, Coral Gables, Florida*, Rehabilitation, Orthotics in
- LEDER, RON S.**, *Universidad Nacional Autonoma de Mexico Mexico, Distrito Federal*, Continuous Positive Airway Pressure
- LEE, CHIN**, *Harvard Medical School, Boston, Massachusetts*, Radiotherapy Treatment Planning, Optimization of; Hyperthermia, Interstitial
- LEE, HAI BANG**, *Korea Research Institutes of Chemical Technology*, Biomaterials: Tissue Engineering and Scaffolds
- LEE, SANG JIN**, *Korea Research Institutes of Chemical Technology*, Biomaterials: Tissue Engineering and Scaffolds
- LEI, LIU**, *Department of General Engineering, Urbana, Illinois*, Bioinformatics

- LEI, XING**, *Stanford University, Stanford, California*, Radiation Dose Planning, Computer-Aided
- LEWIS, MATTHEW C.**, *Medical College of Wisconsin, Milwaukee, Wisconsin*, Hyperbaric Oxygenation
- LI, CHAODI**, *University of Notre Dame, Notre Dame, Indiana*, Bone Cement, Acrylic
- LI, JONATHAN G.**, *University of Florida, Gainesville, Florida*, Radiation Dose Planning, Computer-Aided
- LI, QIAO**, *University of Michigan, Ann Arbor, Michigan*, Immunotherapy
- LI, YANBIN**, *University of Arkansas, Fayetteville, Arkansas*, Piezoelectric Sensors
- LIBOFF, A.R.**, *Oakland University, Rochester, Michigan*, Bone Ununited Fracture and Spinal Fusion, Electrical Treatment of
- LIGAS, JAMES**, *University of Connecticut, Farmington, Connecticut*, Respiratory Mechanics and Gas Exchange
- LIMOGE, AIME**, *The René Descartes University of Paris, Paris, France*, Electroanalgesia, Systemic
- LIN, PEI-JAN PAUL**, *Beth Israel Deaconess Medical Center, Boston, Massachusetts*, Mammography
- LIN, ZHIYUE**, *University of Kansas Medical Center, Kansas City, Kansas*, Electrogastrogram
- LINEAWEAVER, WILLIAM C.**, *University of Mississippi Medical Center, Jackson, Mississippi*, Hyperbaric Medicine
- LIPPING, TARMO**, *Tampere University of Technology, Pori, Finland*, Monitoring in Anesthesia
- LIU, XIAOHUA**, *The University of Michigan, Ann Arbor, Michigan*, Polymeric Materials
- LLOYD, J.J.**, *Regional Medical Physics Department, Newcastle-upon-Tyne, United Kingdom*, Ultraviolet Radiation in Medicine
- LOEB, ROBERT**, *University of Arizona, Tucson, Arizona*, Anesthesia Machines
- LOPES DE MELO, PEDRO**, *State University of Rio de Janeiro, Terreo Salas, Maracaná, Thermistors*
- LOUDON, ROBERT G.**, Lung Sounds
- LOW, DANIEL A.**, *Washington University School of Medicine, St. Louis, Missouri*, Radiation Therapy Simulator
- LU, LICHUN**, *Mayo Clinic, College of Medicine, Rochester, Minnesota*, Microscopy, Electron
- LU, ZHENG FENG**, *Columbia University, New York, New York*, Screen-Film Systems
- LYON, ANDREW W.**, *University of Calgary, Calgary, Canada*, Flame Atomic Emission Spectrometry and Atomic Absorption Spectrometry
- LYON, MARTHA E.**, *University of Calgary, Calgary, Canada*, Flame Atomic Emission Spectrometry and Atomic Absorption Spectrometry
- MA, C-M CHARLIE**, *Fox Chase Cancer Center, Philadelphia, Pennsylvania*, X-Ray Therapy Equipment, Low and Medium Energy
- MACIA, NARCISO F.**, *Arizona State University at the Polytechnic Campus, Mesa, Arizona*, Pneumotachometers
- MACKENZIE, COLIN F.**, *University of Maryland, School of Medicine, Shock, Treatment of*
- MACKIE, THOMAS R.**, *University of Wisconsin, Madison, Wisconsin*, Tomotherapy
- MADNANI, ANJU**, *LSU Medical Centre, Shreveport, Louisiana*, Transcutaneous Electrical Nerve Stimulation (TENS)
- MADNANI, SANJAY**, *LSU Medical Centre, Shreveport, Louisiana*, Transcutaneous Electrical Nerve Stimulation (TENS)
- MADSEN, MARK T.**, *University of Iowa, Iowa City, Iowa*, Anger Camera
- MAGNANO, MAURO**, *ENT Division Mauriziano Hospital, Torino, Italy*, Drug Delivery Systems
- MANDEL, RICHARD**, *Boston University School of Medicine, Boston, Massachusetts*, Colorimetry
- MANNING, KEEFE B.**, *Pennsylvania State University, University Park, Pennsylvania*, Flowmeters
- MAO, JEREMY J.**, *University of Illinois at Chicago, Chicago, Illinois*, Cartilage and Meniscus, Properties of
- MARCOLONGO, MICHELE**, *Drexel University, Philadelphia, Pennsylvania*, Spinal Implants
- MAREK, MIROSLAV**, *Georgia Institute of Technology, Atlanta, Georgia*, Biomaterials, Corrosion and Wear of
- MARION, NICHOLAS W.**, *University of Illinois at Chicago, Chicago, Illinois*, Cartilage and Meniscus, Properties of
- MASTERS, KRISTYN S.**, *University of Wisconsin, Madison, Wisconsin*, Tissue Engineering
- MAUGHAN, RICHARD L.**, *Hospital of the University of Pennsylvania*, Neutron Beam Therapy
- MCADAMS, ERIC**, *University of Ulster at Jordanstown, Newtownabbey, Ireland*, Bioelectrodes
- MCCARTHUR, SALLY L.**, *University of Sheffield, Sheffield, United Kingdom*, Biomaterials, Surface Properties of
- MC EWEN, MALCOM**, *National Research Council of Canada, Ontario, Canada*, Radiation Dosimetry for Oncology
- MCGOWAN, EDWARD J.**, *E.J. McGowan & Associates*, Biofeedback
- MCGRATH, SUSAN**, *Dartmouth College, Hanover, New Hampshire*, Oxygen Analyzers
- MEEKS, SANFORD L.**, *University of Florida, Gainesville, Florida*, Radiosurgery, Stereotactic
- MELISSA, PETER**, *University of Central Florida NanoScience Technology Center, Orlando, Florida*, Biosurface Engineering
- MENDELSON, YITZHAK**, *Worcester Polytechnic Institute*, Optical Sensors
- METZKER, MICHAEL L.**, *Baylor College of Medicine, Houston, Texas*, Polymerase Chain Reaction
- MEYEREND, M.E.**, *University of Wisconsin-Madison, Madison, Wisconsin*, Magnetic Resonance Imaging
- MICHLER, ROBERT**, *Montefiore Medical Center, Bronx, New York*, Heart-Lung Machines
- MICIC, MIODRAG**, *MP Biomedicals LLC, Irvine, California*, Microscopy and Spectroscopy, Near-Field
- MILLER, WILLIAM**, *University of Missouri Columbia, Missouri*, Radionuclide Production and Radioactive Decay
- MITTRA, ERIK**, *Stony Brook University, New York*, Bone Density Measurement
- MODELL, MARK**, *Harvard Medical School, Boston, Massachusetts*, Fiber Optics in Medicine
- MORE, ROBERT B.**, *RBMore Associates, Austin, Texas* Biomaterials Carbon
- MORE, ROBERT**, *Austin, Texas*, Heart Valves, Prosthetic
- MORROW, DARREN**, *Royal Adelaide Hospital, Adelaide, Australia*, Intraaortic Balloon Pump
- MOURTADA, FIRAS**, *MD Anderson Cancer Center, Houston, Texas*, Brachytherapy, Intravascular
- MOY, VINCENT T.**, *University of Miami, Miller School of Medicine, Miami, Florida*, Microscopy, Scanning Force
- MÜFTÜ, SINAN**, *Northeastern University, Boston, Massachusetts*, Tooth and Jaw, Biomechanics of
- MURPHY, RAYMOND L.H.**, Lung Sounds

- MURPHY, WILLIAM L.**, *University of Wisconsin, Madison, Wisconsin*, Tissue Engineering
- MURRAY, ALAN**, *Newcastle University Medical Physics, Newcastle upon Tyne, United Kingdom*, Pace makers
- MUTIC, SASA**, *Washington University School of Medicine, St. Louis, Missouri*, Radiation Therapy Simulator
- NARAYAN, ROGER J.**, *University of North Carolina, Chapel Hill, North Carolina*, Biomaterials, Corrosion and Wear of
- NATALE, ANDREA**, *The Cleveland Clinic Foundation, Cleveland, Ohio*, Hyperthermia, Ultrasonic
- NAZERAN, HOMER**, *The University of Texas, El Paso, Texas*, Electrocardiography, Computers in
- NEUMAN, MICHAEL R.**, *Michigan Technological University, Houghton, Houghton, Michigan*, Fetal Monitoring, Neonatal Monitoring
- NEUZIL, PAVEL**, *Institute of Bioengineering and Nanotechnology, Singapore*, Immunologically Sensitive Field-Effect Transistors
- NICKOLOFF, EDWARD L.**, *Columbia University, New York, New York*, X-Ray Quality Control Program
- NIEZGODA, JEFFREY A.**, *Medical College of Wisconsin, Milwaukee, Wisconsin*, Hyperbaric Oxygenation
- NISHIKAWA, ROBERT M.**, *The University of Chicago, Chicago, Illinois*, Computer-Assisted Detection and Diagnosis
- NUTTER, BRIAN**, *Texas Tech University, Lubbock, Texas*, Medical Records, Computers in
- O'DONOHUE, WILLIAM**, *University of Nevada, Reno, Nevada*, Sexual Instrumentation
- ORTON, COLIN**, *Harper Hospital and Wayne State University, Detroit, Michigan*, Medical Physics Literature
- OZCELIK, SELAHATTIN**, *Texas A&M University, Kingsville, Texas*, Drug Infusion Systems
- PANITCH, ALYSSA**, *Arizona State University, Tempe, Arizona*, Biomaterials: An Overview
- PAOLINO, DONATELLA**, *University of Catanzaro Magna Graecia, Germaneto (CZ), Italy*, Drug Delivery Systems
- PAPAIIOANNOU, GEORGE**, *University of Wisconsin, Milwaukee, Wisconsin*, Joints, Biomechanics of
- PARK, GRACE E.**, *Purdue University, West Lafayette, Indiana*, Porous Materials for Biological Applications
- PARMENTER, BRETT A.**, *State University of New York at Buffalo, Buffalo, New York*, Sleep Studies, Computer Analysis of
- PATEL, DIMPI**, *The Cleveland Clinic Foundation, Cleveland, Ohio*, Hyperthermia, Ultrasonic
- PEARCE, JOHN**, *The University of Texas, Austin, Texas*, Electrosurgical Unit (ESU)
- PELET, SERGE**, *Massachusetts Institute of Technology, Cambridge, Massachusetts*, Microscopy, Fluorescence
- PERIASAMY, AMMASI**, *University of Virginia, Charlottesville, Virginia*, Cellular Imaging
- PERSONS, BARBARA L.**, *University of Mississippi Medical Center, Jackson, Mississippi*, Hyperbaric Medicine
- PIPER, IAN**, *The University of Memphis, Memphis, Tennessee*, Monitoring, Intracranial Pressure
- POLETTI, CHRISTOPHER J.**, *National Institutes of Health*, Tactile Stimulation
- PREMINGER, GLENN M.**, *Duke University Medical Center, Durham, North Carolina*, Lithotripsy
- PRENDERGAST, PATRICK J.**, *Trinity College, Dublin, Ireland*, Orthopedics, Prosthesis Fixation for
- PREVITE, MICHAEL**, *Massachusetts Institute of Technology, Cambridge, Massachusetts*, Microscopy, Fluorescence
- PURDY, JAMES A.**, *UC Davis Medical Center, Sacramento, California*, Radiotherapy Accessories
- QI, HAIRONG**, *Advanced Concepts Analysis, Inc., Falls Church, Virginia*, Thermography
- QIN, YIXIAN**, *Stony Brook University, New York*, Bone Density Measurement
- QUAN, STUART F.**, *University of Arizona, Tucson, Arizona*, Ventilatory Monitoring
- QUIROGA, RODRIGO QUIAN**, *University of Leicester, Leicester, United Kingdom*, Evoked Potentials
- RAHAGHI, FARBOD N.**, *University of California, La Jolla, California*, Glucose Sensors
- RAHKO, PETER S.**, *University of Wisconsin Medical School*, Echocardiography and Doppler Echocardiography
- RALPH, LIETO**, *University of Wisconsin-Madison, Madison, Wisconsin*, Codes and Regulations: Radiation
- RAMANATHAN, LAKSHMI**, *Mount Sinai Medical Center*, Analytical Methods, Automated
- RAO, SATISH S.C.**, *University of Iowa College of Medicine, Iowa City, Iowa*, Anorectal Manometry
- RAPOPORT, DAVID M.**, *NYU School of Medicine, New York, New York*, Continuous Positive Airway Pressure
- REBELLO, KEITH J.**, *The Johns Hopkins University Applied Physics Lab, Laurel, Maryland*, Micro surgery
- REDDY, NARENDER**, *The University of Akron, Akron, Ohio*, Linear Variable Differential Transformers
- REN-DIH, SHEU**, *Memorial Sloan-Kettering Cancer Center, New York, New York*, Radiation Therapy Treatment Planning, Monte Carlo Calculations in
- RENGACHARY, SETTI S.**, *Detroit, Michigan*, Human Spine, Biomechanics of
- REPPERGER, DANIEL W.**, *Wright-Patterson Air Force Base, Dayton, Ohio*, Human Factors in Medical Devices
- RITCHEY, ERIC R.**, *The Ohio State University, Columbus, Ohio*, Contact Lenses
- RIVARD, MARK J.**, *Tufts New England Medical Center, Boston, Massachusetts*, Imaging Devices
- ROBERTSON, J. DAVID**, *University of Missouri, Columbia, Missouri*, Radionuclide Production and Radioactive Decay
- ROTH, BRADLEY J.**, *Oakland University, Rochester, Michigan*, Defibrillators
- ROWE-HORWEGE, R. WANDA**, *University of Texas Medical School, Houston, Texas*, Hyperthermia, Systemic
- RUMSEY, JOHN W.**, *University of Central Florida, Orlando, Florida*, Biosurface Engineering
- RUTKOWSKI, GREGORY E.**, *University of Minnesota, Duluth, Minnesota*, Engineered Tissue
- SALATA, O.V.**, *University of Oxford, Oxford, United Kingdom*, Nanoparticles
- SAMARAS, THEODOROS**, *Aristotle University of Thessaloniki Department of Physics, Thessaloniki, Greece*, Thermometry
- SANGOLE, ARCHANA P.**, *Transitional Learning Center at Galveston, Galveston, Texas*, Rehabilitation, Computers in Cognitive
- SARKOZI, LASZLO**, *Mount Sinai School of Medicine*, Analytical Methods, Automated
- SCHEK, HENRY III**, *University of Michigan, Ann Arbor, Michigan*, Optical Tweezers
- SCHMITZ, CHRISTOPH H.**, *State University of New York Downstate Medical Center, Brooklyn, New York*, Peripheral Vascular Noninvasive Measurements
- SCHUCKERS, STEPHANIE A.C.**, *Clarkson University, Potsdam, New York*, Arrhythmia Analysis, Automated

- SCOPE, KENNETH**, *Northwestern University, Chicago, Illinois*, Ventilators, Acute Medical Care
- SCOTT, ADZICK N.**, *University of Pennsylvania, Philadelphia, Pennsylvania*, Intrauterine Surgical Techniques
- SEAL, BRANDON L.**, *Arizona State University, Tempe, Arizona*, Biomaterials: An Overview
- SEALE, GARY**, *Transitional Learning Center at Galveston, Galveston, Texas*, Rehabilitation, Computers in Cognitive
- SEGERS, PATRICK**, *Ghent University, Belgium*, Hemodynamics
- SELIM, MOSTAFA A.**, *Cleveland Metropolitan General Hospital, Palm Coast, Florida*, Colposcopy
- SETHI, ANIL**, *Loyola University Medical Center, Maywood, Illinois*, X-Rays: Interaction with Matter
- SEVERINGHAUS, JOHN W.**, *University of California in San Francisco, CO₂ Electrodes*
- SHALODI, ABDELWAHAB D.**, *Cleveland Metropolitan General Hospital, Palm Coast, Florida*, Colposcopy
- SHANMUGASUNDARAM, SHOBANA**, *New Jersey Institute of Technology, Newark, New Jersey*, Polymeric Materials
- SHARD, ALEXANDER G.**, *University of Sheffield, Sheffield United Kingdom*, Biomaterials, Surface Properties of
- SHEN, LI-JIUAN**, *National Taiwan University School of Pharmacy, Taipei, Taiwan*, Colorimetry
- SHEN, WEI-CHIANG**, *University of Southern California School of Pharmacy, Los Angeles, California*, Colorimetry
- SHERAR, MICHAEL D.**, *London Health Sciences Centre and University of Western Ontario, London, Ontario, Canada*, Hyperthermia, Interstitial
- SHERMAN, DAVID**, *The Johns Hopkins University, Baltimore, Maryland*, Electroencephalography
- SHI, DONGLU**, *University of Cincinnati, Cincinnati, Ohio*, Biomaterials, Testing and Structural Properties of
- SHUCARD, DAVID W.M.**, *State University of New York at Buffalo, Buffalo, New York*, Sleep Studies, Computer Analysis of
- SIEDBAND, MELVIN P.**, *University of Wisconsin, Madison, Wisconsin*, Image Intensifiers and Fluoroscopy
- SILBERMAN, HOWARD**, *University of Southern California, Los Angeles, California*, Nutrition, Parenteral
- SILVERMAN, GORDON**, *Manhattan College, Computers in the Biomedical Laboratory*
- SILVERN, DAVID A.**, *Medical Physics Unit, Rabin Medical Center, Petah Tikva, Israel*, Prostate Seed Implants
- SINHA, PIYUSH**, *The Ohio State University, Columbus, Ohio*, Drug Delivery Systems
- SINHA, ABHIJIT ROY**, *University of Cincinnati, Cincinnati, Ohio*, Coronary Angioplasty and Guidewire Diagnostics
- SINKJÆR, THOMAS**, *Aalborg University, Aalborg, Denmark*, Electroneurography
- SLOAN, JEFFREY A.**, *Mayo Clinic, Rochester, Minnesota*, Quality-of-Life Measures, Clinical Significance of
- SO, PETER T.C.**, *Massachusetts Institute of Technology, Cambridge, Massachusetts*, Microscopy, Fluorescence
- SOBOL, WLAD T.**, *University of Alabama at Birmingham Health System, Birmingham, Alabama*, Nuclear Magnetic Resonance Spectroscopy
- SOOD, SANDEEP**, *University of Illinois at Chicago, Chicago, Illinois*, Hydrocephalus, Tools for Diagnosis and Treatment of
- SPECTOR, MYRON**, *Brigham and Women's Hospital, Boston, Massachusetts*, Biocompatibility of Materials
- SPELMAN, FRANCIS A.**, *University of Washington, Cochlear Protheses*
- SRINIVASAN, YESHWANTH**, *Texas Tech University, Lubbock, Texas*, Medical Records, Computers in
- SRIRAM, NEELAMEGHAM**, *University of Buffalo, Buffalo, New York*, Cell Counters, Blood
- STARKO, KENTON R.**, *Point Roberts, Washington*, Physiological Systems Modeling
- STARCSCHALL, GEORGE**, *The University of Texas*, Radiotherapy, Three-Dimensional Conformal
- STAVREV, PAVEL**, *Cross Cancer Institute, Edmonton, Alberta, Canada*, Radiotherapy Treatment Planning, Optimization of
- STENKEN, JULIE A.**, *Rensselaer Polytechnic Institute, Troy, New York*, Microdialysis Sampling
- STIEFEL, ROBERT**, *University of Maryland Medical Center, Baltimore, Maryland*, Equipment Acquisition
- STOKES, I.A.F.**, *Polytechnique Montreal, Montreal Quebec, Canada*, Scoliosis, Biomechanics of
- STONE, M.H.**, *University of Leeds, Leeds, United Kingdom*, Hip Joints, Artificial
- SU, XIAO-LI**, *BioDetection Instruments LLC, Fayetteville, Arkansas*, Piezoelectric Sensors
- SUBHAN, ARIF**, *Masterplan Technology Management, Chatsworth, California*, Equipment Maintenance, Biomedical
- SWEENEY, JAMES D.**, *Arizona State University, Tempe, Arizona*, Functional Electrical Stimulation
- SZETO, ANDREW Y.J.**, *San Diego State University, San Diego, California*, Blind and Visually Impaired, Assistive Technology for
- TAKAYAMA, SHUICHI**, *University of Michigan, Ann Arbor, Michigan*, Microbioreactors
- TAMUL, PAUL C.**, *Northwestern University, Chicago, Illinois*, Ventilators, Acute Medical Care
- TAMURA, TOSHIYO**, *Chiba University School of Engineering, Chiba, Japan*, Home Health Care Devices
- TANG, XIANGYANG**, *GE Healthcare Technologies, Waukesha, Wisconsin*, Computed Tomography Simulators
- TAYLOR, B.C.**, *The University of Akron, Akron, Ohio*, Cardiac Output, Indicator Dilution Measurement of
- TEMPLE, RICHARD O.**, *Transitional Learning Center at Galveston, Galveston, Texas*, Rehabilitation, Computers in Cognitive
- TEN, STANLEY**, *Salt Lake City, Utah*, Electroanalgesia, Systemic
- TERRY, TERESA M.**, *Walter Reed Army Institute of Research, Rockville, Maryland*, Blood Collection and Processing
- THAKOR, N.V.**, *Johns Hopkins University, Baltimore, Maryland*, Neurological Monitors
- THIERENS, HUBERT M.A.**, *University of Ghent, Ghent, Belgium*, Radiopharmaceutical Dosimetry
- THOMADSEN, BRUCE**, *University of Wisconsin-Madison, Madison, Wisconsin*, Codes and Regulations: Radiation
- TIPPER, J.L.**, *University of Leeds, Leeds, United Kingdom*, Hip Joints, Artificial
- TOGAWA, TATSUO**, *Waseda University, Saitama, Japan*, Integrated Circuit Temperature Sensor
- TORNAL, MARTIN**, *Duke University, Durham, North Carolina*, X-Ray Equipment Design
- TRAN-SON-TAY, ROGER**, *University of Florida, Gainesville, Florida*, Blood Rheology

- TRAUTMAN, EDWIN D.**, *RMF Strategies, Cambridge, Massachusetts*, Cardiac Output, Thermodilution Measurement of
- TREENA, LIVINGSTON ARINZEH**, *New Jersey Institute of Technology, Newark, New Jersey*, Polymeric Materials
- TRENTMAN, TERRENCE L.**, *Mayo Clinic Scottsdale*, Spinal Cord Stimulation
- TROKEN, ALEXANDER J.**, *University of Illinois at Chicago, Chicago, Illinois*, Cartilage and Meniscus, Properties of
- TSIFTARIS, SOTIRIOS A.**, *Northwestern University, Evanston, Illinois*, DNA Sequence
- TSOUKALAS, D.**, *NTUA, Athens, Attiki, Greece*, Capacitive Microsensors for Biomedical Applications
- TULIPAN, NOEL**, *Vanderbilt University Medical Center, Nashville, Tennessee*, Intrauterine Surgical Techniques
- TUTEJA, ASHOK K.**, *University of Utah, Salt Lake City, Utah*, Anorectal Manometry
- TY, SMITH N.**, *University of California, San Diego, California*, Physiological Systems Modeling
- TYRER, HARRY W.**, *University of Missouri-Columbia, Columbia, Missouri*, Cytology, Automated
- VALVANO, JONATHAN W.**, *The University of Texas, Austin, Texas*, Bioheat Transfer
- VAN DEN HEUVAL, FRANK**, *Wayne State University, Detroit, Michigan*, Imaging Devices
- VEIT, SCHNABEL**, *Aalborg University, Aalborg, Denmark*, Electroneurography
- VELANOVICH, VIC**, *Henry Ford Hospital, Detroit, Michigan*, Esophageal Manometry
- VENKATASUBRAMANIAN, GANAPRIYA**, *Arizona State University, Tempe, Arizona*, Functional Electrical Stimulation
- VERAART, CLAUDE**, *Catholique University of Louvain, Brussels, Belgium*, Visual Prostheses
- VERDONCK, PASCAL**, *Ghent University, Belgium*, Hemodynamics
- VERMARIEN, HERMAN**, *Vrije Universiteit Brussel, Brussels, Belgium*, Phonocardiography, Recorders, Graphic
- VEVES, ARISTIDIS**, *Harvard Medical School, Boston, Massachusetts*, Cutaneous Blood Flow, Doppler Measurement of
- VICINI, PAOLO**, *University of Washington, Seattle, Washington*, Pharmacokinetics and Pharmacodynamics
- VILLE, JÄNTTI**, *Tampere University of Technology, Pori, Finland*, Monitoring in Anesthesia
- VRBA, JINI**, *VSM MedTech Ltd.*, Biomagnetism
- WAGNER, THOMAS, H.**, *M. D. Anderson Cancer Center Orlando, Orlando, Florida*, Radiosurgery, Stereotactic
- WAHLEN, GEORGE E.**, *Veterans Affairs Medical Center and the University of Utah, Salt Lake City, Utah*, Anorectal Manometry
- WALKER, GLENN M.**, *North Carolina State University, Raleigh, North Carolina*, Microfluidics
- WALTERSPACHER, DIRK**, *The Johns Hopkins University, Baltimore, Maryland*, Electroencephalography
- WAN, LEO Q.**, *Liu Ping, Columbia University, New York, New York*, Cartilage and Meniscus, Properties of
- WANG, GE**, *University of Iowa, Iowa City, Iowa*, Computed Tomography Simulators
- WANG, HAIBO**, *Louisiana State University Health Center Shreveport, Louisiana*, Monitoring, Umbilical Artery and Vein, Ambulatory Monitoring
- WANG, HONG**, *Wayne State University, Detroit, Michigan*, Anesthesia, Computers in
- WANG, LE YI**, *Wayne State University, Detroit, Michigan*, Anesthesia, Computers in
- WANG, QIAN, A & M**, *University Health Science Center, Dallas, Texas*, Strain Gages
- WARWICK, WARREN J.**, *University of Minnesota Medical School, Minneapolis, Minnesota*, Cystic Fibrosis Sweat Test
- WATANABE, YOICHI**, *Columbia University Radiation Oncology, New York, New York*, Phantom Materials in Radiology
- WAXLER, MORRIS**, *Godfrey & Kahn S.C., Madison, Wisconsin*, Codes and Regulations: Medical Devices
- WEBSTER, THOMAS J.**, *Purdue University, West Lafayette, Indiana*, Porous Materials for Biological Applications
- WEGENER, JOACHIM**, *University of Oslo, Oslo, Norway*, Impedance Spectroscopy
- WEI, SHYY**, *University of Michigan, Ann Arbor, Michigan*, Blood Rheology
- WEINMEISTER, KENT P.**, *Mayo Clinic Scottsdale*, Spinal Cord Stimulation
- WEIZER, ALON Z.**, *Duke University Medical Center, Durham, North Carolina*, Lithotripsy
- WELLER, PETER**, *City University, London, United Kingdom*, Intraaortic Balloon Pump
- WELLS, JASON**, *LSU Medical Centre, Shreveport, Louisiana*, Transcutaneous Electrical Nerve Stimulation (TENS)
- WENDELKEN, SUZANNE**, *Dartmouth College, Hanover, New Hampshire*, Oxygen Analyzers
- WHELAN, HARRY T.**, *Medical College of Wisconsin, Milwaukee, Wisconsin*, Hyperbaric Oxygenation
- WHITE, ROBERT**, *Memorial Hospital, Regional Newborn Program, South Bend, Indiana*, Incubators, Infant
- WILLIAMS, LAWRENCE E.**, *City of Hope, Duarte, California*, Nuclear Medicine Instrumentation
- WILSON, KERRY**, *University of Central Florida, Orlando, Florida*, Biosurface Engineering
- WINEGARDEN, NEIL**, *University Health Network Microarray Centre, Toronto, Ontario, Canada*, Microarrays
- WOJCIKIEWICZ, EWA P.**, *University of Miami Miller School of Medicine, Miami, Florida*, Microscopy, Scanning Force
- WOLBARST, ANTHONY B.**, *Georgetown Medical School, Washington, DC*, Radiotherapy Treatment Planning, Optimization of
- WOLF, ERIK**, *University of Pittsburgh, Pittsburgh, Pennsylvania*, Mobility Aids
- WOOD, ANDREW**, *Swinburne University of Technology, Melbourne, Australia*, Nonionizing Radiation, Biological Effects of
- WOODCOCK, BRIAN**, *University of Michigan, Ann Arbor, Michigan*, Blood, Artificial
- WREN, JOAKIM**, *Linköping University, Linköping, Sweden*, Thermocouples
- XIANG, ZHOU**, *Brigham and Women's Hospital, Boston, Massachusetts*, Biocompatibility of Materials
- XUEJUN, WEN**, *Clemson University, Clemson, South Carolina*, Biomaterials, Testing and Structural Properties of
- YAN, ZHOU**, *University of Notre Dame, Notre Dame, Indiana*, Bone Cement, Acrylic
- YANNAS, IOANNIS V.**, *Massachusetts Institute of Technology*, Skin Tissue Engineering for Regeneration
- YASZEMSKI, MICHAEL J.**, *Mayo Clinic, College of Medicine, Rochester, Minnesota*, Microscopy, Electron

- YENI, YENER N.**, *Henry Ford Hospital, Detroit, Michigan*, Joints, Biomechanics of
- YLI-HANKALA, ARVI**, *Tampere University of Technology, Pori, Finland*, Monitoring in Anesthesia
- YOKO, KAMOTANI**, *University of Michigan, Ann Arbor, Michigan*, Microbioreactors
- YOON, KANG JI**, *Korea Institute of Science and Technology, Seoul, Korea*, Micropower for Medical Applications
- YORKE, ELLEN**, *Memorial Sloan-Kettering Cancer Center, New York, New York*, Radiation Therapy Treatment Planning, Monte Carlo Calculations in
- YOSHIDA, KEN**, *Aalborg University, Aalborg, Denmark*, Electroneurography
- YOUNGSTEDT, SHAWN D.**, *University of South Carolina, Columbia, South Carolina*, Sleep Laboratory
- YU, YIH-CHOUNG**, *Lafayette College, Easton, Pennsylvania*, Blood Pressure, Automatic Control of
- ZACHARIAH, EMMANUEL S.**, *University of Medicine and Dentistry of New Jersey, New Brunswick, New Jersey*, Immunologically Sensitive Field-Effect Transistors
- ZAIDER, MARCO**, *Memorial Sloan Kettering Cancer Center, New York, New York*, Prostate Seed Implants
- ZAPANTA, CONRAD M.**, *Penn State College of Medicine, Hershey, Pennsylvania*, Heart, Artificial
- ZARDENETA, GUSTAVO**, *University of Texas, San Antonio, Texas*, Fluorescence Measurements
- ZELMANOVIC, DAVID**, *Bayer HealthCare LLC, Tarrytown, New York*, Differential Counts, Automated
- ZHANG, MIN**, *University of Washington, Seattle, Washington*, Biomaterials: Polymers
- ZHANG, YI**, *University of Buffalo, Buffalo, New York*, Cell Counters, Blood
- ZHU, XIAOYUE**, *University of Michigan, Ann Arbor, Michigan*, Microbioreactors
- ZIAIE, BABAK**, *Purdue University, W. Lafayette, Indiana*, Biotelemetry
- ZIELINSKI, TODD M.**, *Medtronic, Inc., Minneapolis, Minnesota*, Bioimpedance in Cardiovascular Medicine
- ZIESSMAN, HARVEY A.**, *Johns Hopkins University*, Computed Tomography, Single Photon Emission

PREFACE

This six-volume work is an alphabetically organized compilation of almost 300 articles that describe critical aspects of medical devices and instrumentation.

It is comprehensive. The articles emphasize the contributions of engineering, physics, and computers to each of the general areas of anesthesiology, biomaterials, burns, cardiology, clinical chemistry, clinical engineering, communicative disorders, computers in medicine, critical care medicine, dermatology, dentistry, ear, nose, and throat, emergency medicine, endocrinology, gastroenterology, genetics, geriatrics, gynecology, hematology, hepatology, internal medicine, medical physics, microbiology, nephrology, neurology, nutrition, obstetrics, oncology, ophthalmology, orthopedics, pain, pediatrics, peripheral vascular disease, pharmacology, physical therapy, psychiatry, pulmonary medicine, radiology, rehabilitation, surgery, tissue engineering, transducers, and urology.

The discipline is defined through the synthesis of the core knowledge from all the fields encompassed by the application of engineering, physics, and computers to problems in medicine. The articles focus not only on what is now useful but also on what is likely to be useful in future medical applications.

These volumes answer the question, "What are the branches of medicine and how does technology assist each of them?" rather than "What are the branches of technology and how could each be used in medicine?" To keep this work to a manageable length, the practice of medicine that is unassisted by devices, such as the use of drugs to treat disease, has been excluded.

The articles are accessible to the user; each benefits from brevity of condensation instead of what could easily have been a book-length work. The articles are designed not for peers, but rather for workers from related fields who wish to take a first look at what is important in the subject.

The articles are readable. They do not presume a detailed background in the subject, but are designed for any person with a scientific background and an interest in technology. Rather than attempting to teach the basics of physiology or Ohm's law, the articles build on such basic concepts to show how the worlds of life science and physical science meld to produce improved systems. While the ideal reader might be a person with a Master's degree in biomedical engineering or medical physics or an M.D. with a physical science undergraduate degree, much of the material will be of value to others with an interest in this growing field. High school students and hospital patients can skip over more technical areas and still gain much from the descriptive presentations.

The *Encyclopedia of Medical Devices and Instrumentation* is excellent for browsing and searching for those new divergent associations that may advance work in a peripheral field. While it can be used as a reference for facts, the articles are long enough that they can serve as an educational instrument and provide genuine understanding of a subject.

One can use this work just as one would use a dictionary, since the articles are arranged alphabetically by topic. Cross references assist the reader looking for subjects listed under slightly different names. The index at the end leads the reader to all articles containing pertinent information on any subject. Listed on pages xxi to xxx are all the abbreviations and acronyms used in the *Encyclopedia*. Because of the increasing use of SI units in all branches of science, these units are provided throughout the *Encyclopedia* articles as well as on pages xxxi to xxxv in the section on conversion factors and unit symbols.

I owe a great debt to the many people who have contributed to the creation of this work. At John Wiley & Sons, Encyclopedia Editor George Telecki provided the idea and guiding influence to launch the project. Sean Pidgeon was Editorial Director of the project. Assistant Editors Roseann Zappia, Sarah Harrington, and Surlan Murrell handled the myriad details of communication between publisher, editor, authors, and reviewers and stimulated authors and reviewers to meet necessary deadlines.

My own background has been in the electrical aspects of biomedical engineering. I was delighted to have the assistance of the editorial board to develop a comprehensive encyclopedia. David J. Beebe suggested cellular topics such as microfluidics. Jerry M. Calkins assisted in defining the chemically related subjects, such as anesthesiology. Michael R. Neuman suggested subjects related to sensors, such as in his own work—neonatology. Joon B. Park has written extensively on biomaterials and suggested related subjects. Edward S. Sternick provided many suggestions from medical physics. The Editorial Board was instrumental both in defining the list of subjects and in suggesting authors.

This second edition brings the field up to date. It is available on the web at <http://www.mrw.interscience.wiley.com/emdi>, where articles can be searched simultaneously to provide rapid and comprehensive information on all aspects of medical devices and instrumentation.

JOHN G. WEBSTER
University of Wisconsin, Madison

LIST OF ARTICLES

ALLOYS, SHAPE MEMORY
AMBULATORY MONITORING
ANALYTICAL METHODS, AUTOMATED
ANESTHESIA MACHINES
ANESTHESIA, COMPUTERS IN
ANGER CAMERA
ANORECTAL MANOMETRY
ARRHYTHMIA ANALYSIS, AUTOMATED
ARTERIES, ELASTIC PROPERTIES OF
AUDIOMETRY
BIOCOMPATIBILITY OF MATERIALS
BIOELECTRODES
BIOFEEDBACK
BIOHEAT TRANSFER
BIOIMPEDANCE IN CARDIOVASCULAR MEDICINE
BIOINFORMATICS
BIOMAGNETISM
BIOMATERIALS, ABSORBABLE
BIOMATERIALS: AN OVERVIEW
BIOMATERIALS: BIOCERAMICS
BIOMATERIALS: CARBON
BIOMATERIALS, CORROSION AND WEAR OF
BIOMATERIALS FOR DENTISTRY
BIOMATERIALS: POLYMERS
BIOMATERIALS, SURFACE PROPERTIES OF
BIOMATERIALS, TESTING AND STRUCTURAL
PROPERTIES OF
BIOMATERIALS: TISSUE ENGINEERING AND
SCAFFOLDS
BIOMECHANICS OF EXERCISE FITNESS
BIOMEDICAL ENGINEERING EDUCATION
BIOSURFACE ENGINEERING
BIOTELEMETRY
BLADDER DYSFUNCTION, NEUROSTIMULATION
OF
BLIND AND VISUALLY IMPAIRED, ASSISTIVE
TECHNOLOGY FOR
BLOOD COLLECTION AND PROCESSING
BLOOD GAS MEASUREMENTS
BLOOD PRESSURE MEASUREMENT
BLOOD PRESSURE, AUTOMATIC CONTROL OF
BLOOD RHEOLOGY
BLOOD, ARTIFICIAL
BONE AND TEETH, PROPERTIES OF
BONE CEMENT, ACRYLIC
BONE DENSITY MEASUREMENT
BONE UNUNITED FRACTURE AND SPINAL FUSION,
ELECTRICAL TREATMENT OF
BORON NEUTRON CAPTURE THERAPY
BRACHYTHERAPY, HIGH DOSAGE RATE
BRACHYTHERAPY, INTRAVASCULAR
CAPACITIVE MICROSENSORS FOR BIOMEDICAL
APPLICATIONS
CARDIAC OUTPUT, FICK TECHNIQUE FOR
CARDIAC OUTPUT, INDICATOR DILUTION
MEASUREMENT OF
CARDIAC OUTPUT, THERMODILUTION
MEASUREMENT OF
CARDIOPULMONARY RESUSCITATION
CARTILAGE AND MENISCUS, PROPERTIES OF
CELL COUNTERS, BLOOD
CELLULAR IMAGING
CHROMATOGRAPHY
CO₂ ELECTRODES
COBALT 60 UNITS FOR RADIOTHERAPY
COCHLEAR PROSTHESES
CODES AND REGULATIONS: MEDICAL DEVICES
CODES AND REGULATIONS: RADIATION
COLORIMETRY
COLPOSCOPY
COMMUNICATION DEVICES
COMMUNICATIVE DISORDERS, COMPUTER
APPLICATIONS FOR
COMPUTED TOMOGRAPHY
COMPUTED TOMOGRAPHY SCREENING
COMPUTED TOMOGRAPHY SIMULATORS
COMPUTED TOMOGRAPHY, SINGLE PHOTON
EMISSION
COMPUTER-ASSISTED DETECTION AND DIAGNOSIS
COMPUTERS IN THE BIOMEDICAL LABORATORY
CONTACT LENSES
CONTINUOUS POSITIVE AIRWAY PRESSURE
CONTRACEPTIVE DEVICES
CORONARY ANGIOPLASTY AND GUIDEWIRE
DIAGNOSTICS
CRYOSURGERY
CUTANEOUS BLOOD FLOW, DOPPLER
MEASUREMENT OF
CYSTIC FIBROSIS SWEAT TEST
CYTOLOGY, AUTOMATED
DEFIBRILLATORS
DIFFERENTIAL COUNTS, AUTOMATED
DIGITAL ANGIOGRAPHY
DNA SEQUENCE
DRUG DELIVERY SYSTEMS
DRUG INFUSION SYSTEMS
ECHOCARDIOGRAPHY AND DOPPLER
ECHOCARDIOGRAPHY
ELECTROANALGESIA, SYSTEMIC
ELECTROCARDIOGRAPHY, COMPUTERS IN
ELECTROCONVULSIVE THERAPY
ELECTROENCEPHALOGRAPHY
ELECTROGASTROGRAM
ELECTROMYOGRAPHY
ELECTRONEUROGRAPHY
ELECTROPHORESIS

- ELECTROPHYSIOLOGY
 ELECTRORETINOGRAPHY
 ELECTROSURGICAL UNIT (ESU)
 ENDOSCOPES
 ENGINEERED TISSUE
 ENVIRONMENTAL CONTROL
 EQUIPMENT ACQUISITION
 EQUIPMENT MAINTENANCE, BIOMEDICAL
 ESOPHAGEAL MANOMETRY
 EVOKED POTENTIALS
 EXERCISE STRESS TESTING
 EYE MOVEMENT, MEASUREMENT TECHNIQUES FOR
 FETAL MONITORING
 FIBER OPTICS IN MEDICINE
 FLAME ATOMIC EMISSION SPECTROMETRY AND
 ATOMIC ABSORPTION SPECTROMETRY
 FLOWMETERS
 FLUORESCENCE MEASUREMENTS
 FUNCTIONAL ELECTRICAL STIMULATION
 GAMMA KNIFE
 GAS AND VACUUM SYSTEMS, CENTRALLY PIPED
 MEDICAL
 GASTROINTESTINAL HEMORRHAGE
 GLUCOSE SENSORS
 HEART VALVE PROSTHESES
 HEART VALVE PROSTHESES, IN VITRO FLOW
 DYNAMICS OF
 HEART VALVES, PROSTHETIC
 HEART, ARTIFICIAL
 HEART-LUNG MACHINES
 HEAT AND COLD, THERAPEUTIC
 HEMODYNAMICS
 HIGH FREQUENCY VENTILATION
 HIP JOINTS, ARTIFICIAL
 HOME HEALTH CARE DEVICES
 HUMAN FACTORS IN MEDICAL DEVICES
 HUMAN SPINE, BIOMECHANICS OF
 HYDROCEPHALUS, TOOLS FOR DIAGNOSIS
 AND TREATMENT OF
 HYPERBARIC MEDICINE
 HYPERBARIC OXYGENATION
 HYPERTHERMIA, INTERSTITIAL
 HYPERTHERMIA, SYSTEMIC
 HYPERTHERMIA, ULTRASONIC
 IMAGE INTENSIFIERS AND FLUOROSCOPY
 IMAGING DEVICES
 IMMUNOLOGICALLY SENSITIVE FIELD-EFFECT
 TRANSISTORS
 IMMUNOTHERAPY
 IMPEDANCE PLETHYSMOGRAPHY
 IMPEDANCE SPECTROSCOPY
 INCUBATORS, INFANT
 INTEGRATED CIRCUIT TEMPERATURE SENSOR
 INTRAAORTIC BALLOON PUMP
 INTRAUTERINE SURGICAL TECHNIQUES
 IONIZING RADIATION, BIOLOGICAL EFFECTS OF
 ION-SENSITIVE FIELD-EFFECT TRANSISTORS
 JOINTS, BIOMECHANICS OF
 LARYNGEAL PROSTHETIC DEVICES
 LENSES, INTRAOCULAR
 LIGAMENT AND TENDON, PROPERTIES OF
 LINEAR VARIABLE DIFFERENTIAL TRANSFORMERS
 LITHOTRIPSY
 LIVER TRANSPLANTATION
 LUNG SOUNDS
 MAGNETIC RESONANCE IMAGING
 MAMMOGRAPHY
 MEDICAL EDUCATION, COMPUTERS IN
 MEDICAL ENGINEERING SOCIETIES
 AND ORGANIZATIONS
 MEDICAL GAS ANALYZERS
 MEDICAL PHYSICS LITERATURE
 MEDICAL RECORDS, COMPUTERS IN
 MICROARRAYS
 MICROBIAL DETECTION SYSTEMS
 MICROBIOREACTORS
 MICRODIALYSIS SAMPLING
 MICROFLUIDICS
 MICROPOWER FOR MEDICAL APPLICATIONS
 MICROSCOPY AND SPECTROSCOPY, NEAR-FIELD
 MICROSCOPY, CONFOCAL
 MICROSCOPY, ELECTRON
 MICROSCOPY, FLUORESCENCE
 MICROSCOPY, SCANNING FORCE
 MICROSCOPY, SCANNING TUNNELING
 MICROSURGERY
 MINIMALLY INVASIVE SURGERY
 MOBILITY AIDS
 MONITORING IN ANESTHESIA
 MONITORING, HEMODYNAMIC
 MONITORING, INTRACRANIAL PRESSURE
 MONITORING, UMBILICAL ARTERY AND VEIN
 MONOCLONAL ANTIBODIES
 NANOPARTICLES
 NEONATAL MONITORING
 NEUROLOGICAL MONITORS
 NEUTRON ACTIVATION ANALYSIS
 NEUTRON BEAM THERAPY
 NONIONIZING RADIATION, BIOLOGICAL EFFECTS OF
 NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY
 NUCLEAR MEDICINE INSTRUMENTATION
 NUCLEAR MEDICINE, COMPUTERS IN
 NUTRITION, PARENTERAL
 OCULAR FUNDUS REFLECTOMETRY
 OCULAR MOTILITY RECORDING AND NYSTAGMUS
 OFFICE AUTOMATION SYSTEMS
 OPTICAL SENSORS
 OPTICAL TWEEZERS
 ORTHOPEDIC DEVICES, MATERIALS AND
 DESIGN FOR
 ORTHOPEDICS, PROSTHESIS FIXATION FOR
 OXYGEN ANALYZERS
 OXYGEN MONITORING
 PACEMAKERS
 PANCREAS, ARTIFICIAL
 PERIPHERAL VASCULAR NONINVASIVE
 MEASUREMENTS
 PHANTOM MATERIALS IN RADIOLOGY
 PHARMACOKINETICS AND PHARMACODYNAMICS
 PHONOCARDIOGRAPHY
 PHOTOGRAPHY, MEDICAL
 PHYSIOLOGICAL SYSTEMS MODELING

PICTURE ARCHIVING AND COMMUNICATION SYSTEMS
PIEZOELECTRIC SENSORS
PNEUMOTACHOMETERS
POLYMERASE CHAIN REACTION
POLYMERIC MATERIALS
POROUS MATERIALS FOR BIOLOGICAL APPLICATIONS
POSITRON EMISSION TOMOGRAPHY
PROSTATE SEED IMPLANTS
PULMONARY PHYSIOLOGY
QUALITY-OF-LIFE MEASURES, CLINICAL SIGNIFICANCE OF
RADIATION DOSE PLANNING, COMPUTER-AIDED
RADIATION DOSIMETRY FOR ONCOLOGY
RADIATION DOSIMETRY, THREE-DIMENSIONAL
RADIATION PROTECTION INSTRUMENTATION
RADIATION THERAPY, INTENSITY MODULATED
RADIATION THERAPY SIMULATOR
RADIATION THERAPY TREATMENT PLANNING, MONTE CARLO CALCULATIONS IN
RADIATION THERAPY, QUALITY ASSURANCE IN RADIOLOGY INFORMATION SYSTEMS
RADIONUCLIDE PRODUCTION AND RADIOACTIVE DECAY
RADIOPHARMACEUTICAL DOSIMETRY
RADIOSURGERY, STEREOTACTIC
RADIOTHERAPY ACCESSORIES
RADIOTHERAPY, HEAVY ION
RADIOTHERAPY, INTRAOPERATIVE
RADIOTHERAPY, THREE-DIMENSIONAL CONFORMAL
RADIOTHERAPY TREATMENT PLANNING, OPTIMIZATION OF
RECORDERS, GRAPHIC
REHABILITATION AND MUSCLE TESTING
REHABILITATION, COMPUTERS IN COGNITIVE
REHABILITATION, ORTHOTICS IN
RESIN-BASED COMPOSITES
RESPIRATORY MECHANICS AND GAS EXCHANGE
SAFETY PROGRAM, HOSPITAL
SCOLIOSIS, BIOMECHANICS OF
SCREEN-FILM SYSTEMS
SEXUAL INSTRUMENTATION
SHOCK, TREATMENT OF
SKIN SUBSTITUTE FOR BURNS, BIOACTIVE
SKIN TISSUE ENGINEERING FOR REGENERATION
SKIN, BIOMECHANICS OF
SLEEP LABORATORY
SLEEP STUDIES, COMPUTER ANALYSIS OF
SPINAL CORD STIMULATION
SPINAL IMPLANTS
STATISTICAL METHODS
STEREOTACTIC SURGERY
STERILIZATION OF BIOLOGIC SCAFFOLD MATERIALS
STRAIN GAGES
TACTILE STIMULATION
TELERADIOLOGY
TEMPERATURE MONITORING
THERMISTORS
THERMOCOUPLES
THERMOGRAPHY
THERMOMETRY
TISSUE ABLATION
TISSUE ENGINEERING
TOMOTHERAPY
TONOMETRY, ARTERIAL
TOOTH AND JAW, BIOMECHANICS OF
TRACER KINETICS
TRANSCUTANEOUS ELECTRICAL NERVE STIMULATION (TENS)
ULTRASONIC IMAGING
ULTRAVIOLET RADIATION IN MEDICINE
VASCULAR GRAFT PROSTHESIS
VENTILATORS, ACUTE MEDICAL CARE
VENTILATORY MONITORING
VISUAL FIELD TESTING
VISUAL PROSTHESES
X-RAY EQUIPMENT DESIGN
X-RAY QUALITY CONTROL PROGRAM
X-RAY THERAPY EQUIPMENT, LOW AND MEDIUM ENERGY
X-RAYS: INTERACTION WITH MATTER
X-RAYS, PRODUCTION OF

ABBREVIATIONS AND ACRONYMS

AAMI	Association for the Advancement of Medical Instrumentation	ALS	Advanced life support; Amyotropic lateral sclerosis
AAPM	American Association of Physicists in Medicine	ALT	Alanine aminotransferase
ABC	Automatic brightness control	ALU	Arithmetic and logic unit
ABET	Accreditation board for engineering training	AM	Amplitude modulation
ABG	Arterial blood gases	AMA	American Medical Association
ABLB	Alternative binaural loudness balance	amu	Atomic mass units
ABS	Acrylonitrile–butadiene–styrene	ANOVA	Analysis of variance
ac	Alternating current	ANSI	American National Standards Institute
AC	Abdominal circumference; Affinity chromatography	AP	Action potential; Alternative pathway; Anteroposterior
ACA	Automated clinical analyzer	APD	Anterioposterior diameter
ACES	Augmentative communication evaluation system	APL	Adjustable pressure limiting valve; Applied Physics Laboratory
ACL	Anterior chamber lens	APR	Anatomically programmed radiography
ACLS	Advanced cardiac life support	AR	Amplitude reduction; Aortic regurgitation; Autoregressive
ACOG	American College of Obstetrics and Gynecology	Ara-C	Arabinosylcytosine
ACR	American College of Radiology	ARD	Absorption rate density
ACS	American Cancer Society; American College of Surgeons	ARDS	Adult respiratory distress syndrome
A/D	Analog-to-digital	ARGUS	Arrhythmia guard system
ADC	Agar diffusion chambers; Analog-to-digital converter	ARMA	Autoregressive-moving-average model
ADCC	Antibody-dependent cellular cytotoxicity	ARMAX	Autoregressive-moving-average model with external inputs
ADCL	Accredited Dosimetry Calibration Laboratories	AS	Aortic stenosis
ADP	Adenosine diphosphate	ASA	American Standards Association
A-D-T	Admission, discharge, and transfer	ASCII	American standard code for information interchange
AE	Anion exchange; Auxiliary electrode	ASD	Antisiphon device
AEA	Articulation error analysis	ASHE	American Society for Hospital Engineering
AEB	Activation energy barrier	ASTM	American Society for Testing and Materials
AEC	Automatic exposure control	AT	Adenosine-thiamide; Anaerobic threshold; Antithrombin
AED	Automatic external defibrillator	ATA	Atmosphere absolute
AEMB	Alliance for Engineering in Medicine and Biology	ATLS	Advanced trauma life support
AES	Auger electron spectroscopy	ATN	Acute tubular necrosis
AESC	American Engineering Standards Committee	ATP	Adenosine triphosphate
AET	Automatic exposure termination	ATPD	Ambient temperature pressure dry
AFO	Ankle-foot orthosis	ATPS	Ambient temperature pressure saturated
AGC	Automatic gain control	ATR	Attenuated total reflection
AHA	American Heart Association	AUC	Area under curve
AI	Arterial insufficiency	AUMC	Area under moment curve
AICD	Automatic implantable cardiac defibrillator	AV	Atrioventricular
AID	Agency for International Development	AZT	Azido thymidine
AIDS	Acquired immune deficiency syndrome	BA	Biliary atresia
AL	Anterior leaflet	BAEP	Brainstem auditory evoked potential
ALG	Antilymphocyte globulin	BAPN	Beta-amino-propionitril
		BAS	Boston anesthesia system
		BASO	Basophil
		BB	Buffer base
		BBT	Basal body temperature

BCC	Body-centered cubic	CCTV	Closed circuit television system
BCD	Binary-coded decimal	CCU	Coronary care unit; Critical care unit
BCG	Ballistocardiogram	CD	Current density
BCLS	Basic cardiac life support	CDR	Complimentary determining region
BCRU	British Committee on Radiation Units and Measurements	CDRH	Center for Devices and Radiological Health
BDI	Beck depression inventory	CEA	Carcinoembryonic antigen
BE	Base excess; Binding energy	CF	Conversion factor; Cystic fibrosis
BET	Brunauer, Emmett, and Teller methods	CFC	Continuous flow cytometer
BH	His bundle	CFR	Code of Federal Regulations
BI	Biological indicators	CFU	Colony forming units
BIH	Beth Israel Hospital	CGA	Compressed Gas Association
BIPM	International Bureau of Weights and Measurements	CGPM	General Conference on Weights and Measures
BJT	Bipolar junction transistor	CHO	Carbohydrate
BMDP	Biomedical Programs	CHO	Chinese hamster ovary
BME	Biomedical engineering	CI	Combination index
BMET	Biomedical equipment technician	CICU	Cardiac intensive care unit
BMO	Biomechanically optimized	CIF	Contrast improvement factor
BMR	Basal metabolic rate	CIN	Cervical intraepithelial neoplasia
BOL	Beginning of life	CK	Creatine kinase
BP	Bereitschafts potential; Break point	CLAV	Clavicle
BR	Polybutadiene	CLSA	Computerized language sample analysis
BRM	Biological response modifier	CM	Cardiomyopathy; Code modulation
BRS	Bibliographic retrieval services	CMAD	Computer managed articulation diagnosis
BSS	Balanced salt solution	CMI	Computer-managed instruction
BTG	Beta thromboglobulin	CMRR	Common mode rejection ratio
BTPS	Body temperature pressure saturated	CMV	Conventional mechanical ventilation; Cytomegalovirus
BUN	Blood urea nitrogen	CNS	Central nervous system
BW	Body weight	CNV	Contingent negative variation
CA	Conductive adhesives	CO	Carbon monoxide; Cardiac output
CABG	Coronary artery by-pass grafting	COBAS	Comprehensive Bio-Analysis System
CAD/CAM	Computer-aided design/computer-aided manufacturing	COPD	Chronic obstructive pulmonary disease
CAD/D	Computer-aided drafting and design	COR	Center of rotation
CADD	Central axis depth dose	CP	Cerebral palsy; Closing pressure; Creatine phosphate
CAI	Computer assisted instruction; Computer-aided instruction	CPB	Cardiopulmonary bypass
CAM	Computer-assisted management	CPET	Cardiac pacemaker electrode tips
cAMP	Cyclic AMP	CPM	Computerized probe measurements
CAPD	Continuous ambulatory peritoneal dialysis	CPP	Cerebral perfusion pressure; Cryoprecipitated plasma
CAPP	Child amputee prosthetic project	CPR	Cardiopulmonary resuscitation
CAT	Computerized axial tomography	cps	Cycles per second
CATS	Computer-assisted teaching system; Computerized aphasia treatment system	CPU	Central Processing unit
CAVH	Continuous arteriovenous hemofiltration	CR	Center of resistance; Conditioned response; Conductive rubber; Creatinine
CB	Conjugated bilirubin; Coulomb barrier	CRBB	Complete right bundle branch block
CBC	Complete blood count	CRD	Completely randomized design
CBF	Cerebral blood flow	CRL	Crown rump length
CBM	Computer-based management	CRT	Cathode ray tube
CBV	Cerebral blood volume	CS	Conditioned stimulus; Contrast scale; Crown seat
CC	Closing capacity	CSA	Compressed spectral array
CCC	Computer Curriculum Company	CSF	Cerebrospinal fluid
CCD	Charge-coupled device	CSI	Chemical shift imaging
CCE	Capacitance contact electrode	CSM	Chemically sensitive membrane
CCF	Cross-correlation function	CT	Computed tomography; Computerized tomography
CCL	Cardiac catheterization laboratory	CTI	Cumulative toxicity response index
CCM	Critical care medical services	CV	Closing volume
CCPD	Continuous cycling peritoneal dialysis		

C.V.	Coefficient of variation	EBS	Early burn scar
CVA	Cerebral vascular accident	EBV	Epstein–Barr Virus
CVP	Central venous pressure	EC	Ethyl cellulose
CVR	Cardiovascular resistance	ECC	Emergency cardiac care; Extracorporeal circulation
CW	Continuous wave	ECCE	Extracapsular cataract extinction
CWE	Coated wire electrodes	ECD	Electron capture detector
CWRU	Case Western Reserve University	ECG	Electrocardiogram
DAC	Digital-to-analog converter	ECM	Electrochemical machining
DAS	Data acquisition system	ECMO	Extracorporeal membrane oxygenation
dB	Decibel	ECOD	Extracranial cerebrovascular occlusive disease
DB	Direct body	ECRI	Emergency Care Research Institute
DBMS	Data base management system	ECS	Exner's Comprehensive System
DBS	Deep brain stimulation	ECT	Electroconvulsive shock therapy; Electroconvulsive therapy; Emission computed tomography
dc	Direct current	EDD	Estimated date of delivery
DCCT	Diabetes control and complications trial	EDP	Aortic end diastolic pressure
DCP	Distal cavity pressure	EDTA	Ethylenediaminetetraacetic acid
DCS	Dorsal column stimulation	EDX	Energy dispersive X-ray analysis
DDC	Deck decompression chamber	EEG	Electroencephalogram
DDS	Deep diving system	EEI	Electrode electrolyte interface
DE	Dispersive electrode	EELV	End-expiratory lung volume
DEN	Device experience network	EER	Electrically evoked response
DERS	Drug exception ordering system	EF	Ejection fraction
DES	Diffuse esophageal spasm	EF	Electric field; Evoked magnetic fields
d.f.	Distribution function	EFA	Estimated fetal age
DHCP	Distributed Hospital Computer Program	EGF	Epidermal growth factor
DHE	Dihematoporphyrin ether	EGG	Electrogastrogram
DHEW	Department of Health Education and Welfare	EIA	Enzyme immunoassay
DHHS	Department of Health and Human Services	EIU	Electrode impedance unbalance
DHT	Duration of hypothermia	ELF	Extra low frequency
DI	Deionized water	ELGON	Electrical goniometer
DIC	Displacement current	ELISA	Enzyme-linked immunosorbent assay
DIS	Diagnostic interview schedule	ELS	Energy loss spectroscopy
DL	Double layer	ELV	Equivalent lung volume
DLI	Difference lumen for intensity	EM	Electromagnetic
DM	Delta modulation	EMBS	Engineering in Medicine and Biology Society
DME	Dropping mercury electrode	emf	Electromotive force
DN	Donation number	EMG	Electromyogram
DNA	Deoxyribonucleic acid	EMGE	Integrated electromyogram
DOF	Degree of freedom	EMI	Electromagnetic interference
DOS	Drug ordering system	EMS	Emergency medical services
DOT-NHTSA	Department of Transportation Highway Traffic Safety Administration	EMT	Emergency medical technician
DPB	Differential pencil beam	ENT	Ear, nose, and throat
DPG	Diphosphoglycerate	EO	Elbow orthosis
DQE	Detection quantum efficiency	EOG	Electrooculography
DRESS	Depth-resolved surface coil spectroscopy	EOL	End of life
DRG	Diagnosis-related group	EOS	Eosinophil
DSA	Digital subtraction angiography	EP	Elastoplastic; Evoked potentiate
DSAR	Differential scatter-air ratio	EPA	Environmental protection agency
DSB	Double strand breaks	ER	Evoked response
DSC	Differential scanning calorimetry	ERCP	Endoscopic retrograde cholangiopancreatography
D-T	Deuterium-on-tritium	ERG	Electron radiography; Electroretinogram
DTA	Differential thermal analysis	ERMF	Event-related magnetic field
d.u.	Density unit	ERP	Event-related potential
DUR	Duration	ERV	Expiratory reserve volume
DVT	Deep venous thrombosis		
EA	Esophageal accelerometer		
EB	Electron beam		
EBCDIC	Extended binary code decimal interchange code		

ESCA	Electron spectroscopy for chemical analysis	GC	Gas chromatography; Guanine-cytosine
ESI	Electrode skin impedance	GDT	Gas discharge tube
ESRD	End-stage renal disease	GFR	Glomerular filtration rate
esu	Electrostatic unit	GHb	Glycosylated hemoglobin
ESU	Electrosurgical unit	GI	Gastrointestinal
ESWL	Extracorporeal shock wave lithotripsy	GLC	Gas-liquid chromatography
ETO, Eto	Ethylene oxide	GMV	General minimum variance
ETT	Exercise tolerance testing	GNP	Gross national product
EVA	Ethylene vinyl acetate	GPC	Giant papillary conjunctivitis
EVR	Endocardial viability ratio	GPH	Gas-permeable hard
EW	Extended wear	GPH-EW	Gas-permeable hard lens extended wear
FAD	Flavin adenine dinucleotide	GPO	Government Printing Office
FARA	Flexible automation random analysis	GSC	Gas-solid chromatography
FBD	Fetal biparietal diameter	GSR	Galvanic skin response
FBS	Fetal bovine serum	GSWD	Generalized spike-wave discharge
fcc	Face centered cubic	HA	Hydroxyapatite
FCC	Federal Communications Commission	HAM	Helical axis of motion
Fct	Fluorocrit	Hb	Hemoglobin
FDA	Food and Drug Administration	HBE	His bundle electrogram
FDCA	Food, Drug, and Cosmetic Act	HBO	Hyperbaric oxygenation
FE	Finite element	HC	Head circumference
FECG	Fetal electrocardiogram	HCA	Hypothermic circulatory arrest
FEF	Forced expiratory flow	HCFA	Health care financing administration
FEL	Free electron lasers	HCL	Harvard Cyclotron Laboratory
FEM	Finite element method	hcp	Hexagonal close-packed
FEP	Fluorinated ethylene propylene	HCP	Half cell potential
FES	Functional electrical stimulation	HDPE	High density polyethylene
FET	Field-effect transistor	HECS	Hospital Equipment Control System
FEV	Forced expiratory volume	HEMS	Hospital Engineering Management System
FFD	Focal spot to film distance	HEPA	High efficiency particulate air filter
FFT	Fast Fourier transform	HES	Hydroxyethylstarch
FGF	Fresh gas flow	HETP	Height equivalent to a theoretical plate
FHR	Fetal heart rate	HF	High-frequency; Heating factor
FIC	Forced inspiratory capacity	HFCWO	High-frequency chest wall oscillation
FID	Flame ionization detector; Free-induction decay	HFER	High-frequency electromagnetic radiation
FIFO	First-in-first-out	HFJV	High-frequency jet ventilation
FITC	Fluorescent indicator tagged polymer	HFO	High-frequency oscillator
FL	Femur length	HFOV	High-frequency oscillatory ventilation
FM	Frequency modulation	HFPPV	High-frequency positive pressure ventilation
FNS	Functional neuromuscular stimulation	HFV	High-frequency ventilation
FO	Foramen ovale	HHS	Department of Health and Human Services
FO-CRT	Fiber optics cathode ray tube	HIBC	Health industry bar code
FP	Fluorescence polarization	HIMA	Health Industry Manufacturers Association
FPA	Fibrinopeptide A	HIP	Hydrostatic indifference point
FR	Federal Register	HIS	Hospital information system
FRC	Federal Radiation Council; Functional residual capacity	HK	Hexokinase
FSD	Focus-to-surface distance	HL	Hearing level
FTD	Focal spot to tissue-plane distance	HMBA	Hexamethylene bisacetamide
FTIR	Fourier transform infrared	HMO	Health maintenance organization
FTMS	Fourier transform mass spectrometer	HMWPE	High-molecular-weight polyethylene
FU	Fluorouracil	HOL	Higher-order languages
FUDR	Floxuridine	HP	Heating factor; His-Purkinje
FVC	Forced vital capacity	HpD	Hematoporphyrin derivative
FWHM	Full width at half maximum	HPLC	High-performance liquid chromatography
FWTM	Full width at tenth maximum	HPNS	High-pressure neurological syndrome
GABA	Gamma amino buteric acid	HPS	His-Purkinje system
GAG	Glycosaminoglycan	HPX	High peroxidase activity
GBE	Gas-bearing electrodyamometer		

HR	Heart rate; High-resolution	IMIA	International Medical Informatics Association
HRNB	Halstead-Reitan Neuropsychological Battery	IMS	Information management system
H/S	Hard/soft	IMV	Intermittent mandatory ventilation
HSA	Human serum albumin	INF	Interferon
HSG	Hysterosalpingogram	IOL	Intraocular lens
HTCA	Human tumor cloning assay	IPC	Ion-pair chromatography
HTLV	Human T cell lymphotropic virus	IPD	Intermittent peritoneal dialysis
HU	Heat unit; Houndsfield units; Hydroxyurea	IPG	Impedance plethysmography
HVL	Half value layer	IPI	Interpulse interval
HVR	Hypoxic ventilatory response	IPPB	Intermittent positive pressure breathing
HVT	Half-value thickness	IPTS	International practical temperature scale
IA	Image intensifier assembly; Inominate artery	IR	Polyisoprene rubber
IABP	Intraaortic balloon pumping	IRB	Institutional Review Board
IAEA	International Atomic Energy Agency	IRBBB	Incomplete right bundle branch block
IAIMS	Integrated Academic Information Management System	IRPA	International Radiation Protection Association
IASP	International Association for the Study of Pain	IRRAS	Infrared reflection-absorption spectroscopy
IC	Inspiratory capacity; Integrated circuit	IRRS	Infrared reflection spectroscopy
ICCE	Intracapsular cataract extraction	IRS	Internal reflection spectroscopy
ICD	Intracervical device	IRV	Inspiratory reserve capacity
ICDA	International classification of diagnoses	IS	Image size; Ion-selective
ICL	Ms-clip lens	ISC	Infant skin servo control
ICP	Inductively coupled plasma; Intracranial pressure	ISDA	Instantaneous screw displacement axis
ICPA	Intracranial pressure amplitude	ISE	Ion-selective electrode
ICRP	International Commission on Radiological Protection	ISFET	Ion-sensitive field effect transistor
ICRU	International Commission on Radiological Units and Measurements	ISIT	Intensified silicon-intensified target tube
ICU	Intensive care unit	ISO	International Organization for Standardization
ID	Inside diameter	ISS	Ion scattering spectroscopy
IDDM	Insulin dependent diabetes mellitus	IT	Intrathecal
IDE	Investigational device exemption	ITEP	Institute of Theoretical and Experimental Physics
IDI	Index of inspired gas distribution	ITEPI	Instantaneous trailing edge pulse impedance
I:E	Inspiratory: expiratory	ITLC	Instant thin-layer chromatography
IEC	International Electrotechnical Commission; Ion-exchange chromatography	IUD	Intrauterine device
IEEE	Institute of Electrical and Electronics Engineers	IV	Intravenous
IEP	Individual educational program	IVC	Inferior vena cava
BETS	Inelastic electron tunneling spectroscopy	IVP	Intraventricular pressure
IF	Immunofluorescent	JCAH	Joint Commission on the Accreditation of Hospitals
IFIP	International Federation for Information Processing	JND	Just noticeable difference
IFMBE	International Federation for Medical and Biological Engineering	JRP	Joint replacement prosthesis
IGFET	Insulated-gate field-effect transistor	KB	Kent bundle
IgG	Immunoglobulin G	Kerma	Kinetic energy released in unit mass
IgM	Immunoglobulin M	KO	Knee orthosis
IHP	Inner Helmholtz plane	KPM	Kilopond meter
IHSS	Idiopathic hypertrophic subaortic stenosis	KRPB	Krebs-Ringer physiological buffer
II	Image intensifier	LA	Left arm; Left atrium
IIIES	Image intensifier input-exposure sensitivity	LAD	Left anterior descending; Left axis deviation
IM	Intramuscular	LAE	Left atrial enlargement
IMFET	Immunologically sensitive field-effect transistor	LAK	Lymphokine activated killer
		LAL	Limulus amoebocyte lysate
		LAN	Local area network
		LAP	Left atrial pressure
		LAT	Left anterior temporalis
		LBBB	Left bundle branch block
		LC	Left carotid; Liquid chromatography

LCC	Left coronary cusp	MDP	Mean diastolic aortic pressure
LCD	Liquid crystal display	MDR	Medical device reporting
LDA	Laser Doppler anemometry	MDS	Multidimensional scaling
LDF	Laser Doppler flowmetry	ME	Myoelectric
LDH	Lactate dehydrogenase	MED	Minimum erythema dose
LDPE	Low density polyethylene	MEDPAR	Medicare provider analysis and review
LEBS	Low-energy brief stimulus	MEFV	Maximal expiratory flow volume
LED	Light-emitting diode	MEG	Magnetoencephalography
LEED	Low energy electron diffraction	MeSH	Medline subject heading
LES	Lower esophageal sphincter	METS	Metabolic equivalents
LESP	Lower esophageal sphincter pressure	MF	Melamine-formaldehyde
LET	Linear energy transfer	MFP	Magnetic field potential
LF	Low frequency	MGH	Massachusetts General Hospital
LH	Luteinizing hormone	MHV	Magnetic heart vector
LHT	Local hyperthermia	MI	Myocardial infarction
LL	Left leg	MIC	Minimum inhibitory concentration
LLDPE	Linear low density polyethylene	MIFR	Maximum inspiratory flow rate
LLPC	Liquid-liquid partition chromatography	MINET	Medical Information Network
LLW	Low-level waste	MIR	Mercury-in-rubber
LM	Left masseter	MIS	Medical information system; Metal-insulator-semiconductor
LNNB	Luria-Nebraska Neuropsychological Battery	MIT	Massachusetts Institute of Technology
LOS	Length of stay	MIT/BIH	Massachusetts Institute of Technology/ Beth Israel Hospital
LP	Late potential; Lumboperitoneal	MMA	Manual metal arc welding
LPA	Left pulmonary artery	MMA	Methyl methacrylate
LPC	Linear predictive coding	MMECT	Multiple-monitored ECT
LPT	Left posterior temporalis	MMFR	Maximum midexpiratory flow rate
LPV	Left pulmonary veins	mm Hg	Millimeters of mercury
LRP	Late receptor potential	MMPI	Minnesota Multiphasic Personality Inventory
LS	Left subclavian	MMSE	Minimum mean square error
LSC	Liquid-solid adsorption chromatography	MO	Membrane oxygenation
LSI	Large scale integrated	MONO	Monocyte
LSV	Low-amplitude shear-wave viscoelastometry	MOSFET	Metal oxide silicon field-effect transistor
LTI	Low temperature isotropic	MP	Mercaptopurine; Metacarpal-phalangeal
LUC	Large unstained cells	MPD	Maximal permissible dose
LV	Left ventricle	MR	Magnetic resonance
LVAD	Left ventricular assist device	MRG	Magnetoretinogram
LVDT	Linear variable differential transformer	MRI	Magnetic resonance imaging
LVEP	Left ventricular ejection period	MRS	Magnetic resonance spectroscopy
LVET	Left ventricular ejection time	MRT	Mean residence time
LVH	Left ventricular hypertrophy	MS	Mild steel; Multiple sclerosis
LYMPH	Lymphocyte	MSR	Magnetically shielded room
MAA	Macroaggregated albumin	MTBF	Mean time between failure
MAC	Minimal auditory capabilities	MTF	Modulation transfer function
MAN	Manubrium	MTTR	Mean time to repair
MAP	Mean airway pressure; Mean arterial pressure	MTX	Methotroxate
MAST	Military assistance to safety and traffic	MUA	Motor unit activity
MBA	Monoclonal antibody	MUAP	Motor unit action potential
MBV	Maximum breathing ventilation	MUAPT	Motor unit action potential train
MBX	Monitoring branch exchange	MUMPI	Missouri University Multi-Plane Imager
MCA	Methyl cryanoacrylate	MUMPS	Massachusetts General Hospital utility multiuser programming system
MCG	Magnetocardiogram	MV	Mitral valve
MCI	Motion Control Incorporated	MVO ₂	Maximal oxygen uptake
MCM1	Millon Clinical Multiaxial Inventory	MVTR	Moisture vapor transmission rate
MCT	Microcatheter transducer	MVV	Maximum voluntary ventilation
MCV	Mean corpuscular volume	MW	Molecular weight
MDC	Medical diagnostic categories		
MDI	Diphenylmethane diisocyanate; Medical Database Informatics		

NAA	Neutron activation analysis	OPG	Ocular pneumoplethysmography
NAD	Nicotinamide adenine dinucleotide	OR	Operating room
NADH	Nicotinamide adenine dinucleotide, reduced form	OS	Object of known size; Operating system
NADP	Nicotinamide adenine dinucleotide phosphate	OTC	Over the counter
NAF	Neutrophil activating factor	OV	Offset voltage
NARM	Naturally occurring and accelerator- produced radioactive materials	PA	Posteroanterior; Pulmonary artery; Pulse amplitude
NBB	Normal buffer base	PACS	Picture archiving and communications systems
NBD	Neuromuscular blocking drugs	PAD	Primary afferent depolarization
N-BPC	Normal bonded phase chromatography	PAM	Pulse amplitude modulation
NBS	National Bureau of Standards	PAN	Polyacrylonitrile
NCC	Noncoronary cusp	PAP	Pulmonary artery pressure
NCCLS	National Committee for Clinical Laboratory Standards; National Committee on Clinical Laboratory Standards	PAR	Photoactivation ratio
NCRP	National Council on Radiation Protection	PARFR	Program for Applied Research on Fertility Regulation
NCT	Neutron capture theory	PARR	Poetanesesthesia recovery room
NEEP	Negative end-expiratory pressure	PAS	Photoacoustic spectroscopy
NEMA	National Electrical Manufacturers Association	PASG	Pneumatic antishock garment
NEMR	Nonionizing electromagnetic radiation	PBI	Penile brachial index
NEQ	Noise equivalent quanta	PBL	Positive beam limitation
NET	Norethisterone	PBT	Polybutylene terephthalate
NEUT	Neutrophil	PC	Paper chromatography; Personal computer; Polycarbonate
NFPA	National Fire Protection Association	PCA	Patient controlled analgesia; Principal components factor analysis
NH	Neonatal hepatitis	PCG	Phonocardiogram
NHE	Normal hydrogen electrode	PCI	Physiological cost index
NHLBI	National Heart, Lung, and Blood Institute	PCL	Polycaprolactone; Posterior chamber lens
NIR	Nonionizing radiation	PCR	Percent regurgitation
NIRS	National Institute for Radiologic Science	PCRC	Perinatal Clinical Research Center
NK	Natural killer	PCS	Patient care system
NMJ	Neuromuscular junction	PCT	Porphyria cutanea tarda
NMOS	N-type metal oxide silicon	PCWP	Pulmonary capillary wedge pressure
NMR	Nuclear magnetic resonance	PD	Peritoneal dialysis; Poly-p-dioxanone; Potential difference; Proportional and derivative
NMS	Neuromuscular stimulation	PDD	Percent depth dose; Perinatal Data Directory
NPH	Normal pressure hydrocephalus	PDE	Pregelged disposable electrodes
NPL	National Physical Laboratory	p.d.f.	Probability density function
NR	Natural rubber	PDL	Periodontal ligament
NRC	Nuclear Regulatory Commission	PDM	Pulse duration modulation
NRZ	Non-return-to-zero	PDMSX	Polydimethyl siloxane
NTC	Negative temperature coefficient	PDS	Polydioxanone
NTIS	National Technical Information Service	PE	Polyethylene
NVT	Neutrons versus time	PEEP	Positive end-expiratory pressure
NYHA	New York Heart Association	PEFR	Peak expiratory now rate
ob/gyn	Obstetrics and gynecology	PEN	Parenteral and enteral nutrition
OCR	Off-center ratio; Optical character recognition	PEP	Preejection period
OCV	Open circuit voltage	PEPPER	Programs examine phonetic find phonological evaluation records
OD	Optical density; Outside diameter	PET	Polyethylene terephthalate; Positron-emission tomography
ODC	Oxyhemoglobin dissociation curve	PEU	Polyetherurethane
ODT	Oxygen delivery truck	PF	Platelet factor
ODU	Optical density unit	PFA	Phosphonoformic add
OER	Oxygen enhancement ratio	PFC	Petrofluorochemical
OFD	Object to film distance; Occiputo-frontal diameter	PFT	Pulmonary function testing
OHL	Outer Helmholtz layer	PG	Polyglycolide; Propylene glycol
OHP	Outer Helmholtz plane		
OIH	Orthoiodohippurate		

PGA	Polyglycolic add	PURA	Prolonged ultraviolet-A radiation
PHA	Phytohemagglutinin; Pulse-height analyzer	PUVA	Psoralens and longwave ultraviolet light photochemotherapy
PHEMA	Poly-2-hydroxyethyl methacrylate	P/V	Pressure/volume
PI	Propidium iodide	PVC	Polyvinyl chloride; Premature ventricular contraction
PID	Pelvic inflammatory disease; Proportional/integral/derivative	PVI	Pressure-volume index
PIP	Peak inspiratory pressure	PW	Pulse wave; Pulse width
PL	Posterior leaflet	PWM	Pulse width modulation
PLA	Polylactic acid	PXE	Pseudo-xanthoma elasticum
PLATO	Program Logic for Automated Teaching Operations	QA	Quality assurance
PLD	Potentially lethal damage	QC	Quality control
PLED	Periodic lateralized epileptiform discharge	R-BPC	Reverse bonded phase chromatography
PLT	Platelet	R/S	Radiopaque-spherical
PM	Papillary muscles; Preventive maintenance	RA	Respiratory amplitude; Right arm
PMA	Polymethyl acrylate	RAD	Right axis deviation
p.m.f.	Probability mass function	RAE	Right atrial enlargement
PMMA	Polymethyl methacrylate	RAM	Random access memory
PMOS	P-type metal oxide silicon	RAP	Right atrial pressure
PMP	Patient management problem; Poly(4-methylpentane)	RAT	Right anterior temporalis
PMT	Photomultiplier tube	RB	Right bundle
PO	Per os	RBBB	Right bundle branch block
P_{O_2}	Partial pressure of oxygen	RBC	Red blood cell
POBT	Polyoxybutylene terephthalate	RBE	Relative biologic effectiveness
POM	Polyoxymethylene	RBF	Rose bengal fecal excretion
POMC	Patient order management and communication system	RBI	Resting baseline impedance
POPRAS	Problem Oriented Perinatal Risk Assessment System	RCBD	Randomized complete block diagram
PP	Perfusion pressure; Polypropylene; Postprandial (after meals)	rCBF	Regional cerebral blood flow
PPA	Phonemic process analysis	RCC	Right coronary cusp
PPF	Plasma protein fraction	RCE	Resistive contact electrode
PPM	Pulse position modulation	R&D	Research and development
PPSFH	Polymerized phyridoxalated stroma-free hemoglobin	r.e.	Random experiment
PR	Pattern recognition; Pulse rate	RE	Reference electrode
PRBS	Pseudo-random binary signals	REM	Rapid eye movement; Return electrode monitor
PRP	Pulse repetition frequency	REMATE	Remote access and telecommunication system
PRO	Professional review organization	RES	Reticuloendothelial system
PROM	Programmable read only memory	RESNA	Rehabilitation Engineering Society of North America
PS	Polystyrene	RF	Radio frequency; Radiographic-nuoroscopic
PSA	Pressure-sensitive adhesive	RFI	Radio-frequency interference
PSF	Point spread function	RFP	Request for proposal
PSI	Primary skin irritation	RFQ	Request for quotation
PSP	Postsynaptic potential	RH	Relative humidity
PSR	Proton spin resonance	RHE	Reversible hydrogen electrode
PSS	Progressive systemic sclerosis	RIA	Radioimmunoassay
PT	Plasma thromboplastin	RM	Repetition maximum; Right masseter
PTB	Patellar tendon bearing orthosis	RMR	Resting metabolic rate
PTC	Plasma thromboplastin component; Positive temperature coefficient; Pressurized personal transfer capsule	RMS	Root mean square
PTCA	Percutaneous transluminal coronary angioplasty	RN	Radionuclide
PTFE	Polytetrafluoroethylene	RNCA	Radionuclide cineangiogram
PTT	Partial thromboplastin time	ROI	Regions of interest
PUL	Percutaneous ultrasonic lithotripsy	ROM	Range of motion; Read only memory
		RP	Retinitis pigmentosa
		RPA	Right pulmonary artery
		RPP	Rate pressure product
		RPT	Rapid pull-through technique
		RPV	Right pulmonary veins
		RQ	Respiratory quotient

RR	Recovery room	SEBS	Surgical isolation barrier system
RRT	Recovery room time; Right posterior temporalis	SID	Source to image reception distance
RT	Reaction time	SIMFU	Scanned intensity modulated focused ultrasound
RTD	Resistance temperature device	SIMS	Secondary ion mass spectroscopy; System for isometric muscle strength
RTT	Revised token test	SISI	Short increment sensitivity index
r.v.	Random variable	SL	Surgical lithotomy
RV	Residual volume; Right ventricle	SLD	Sublethal damage
RVH	Right ventricular hypertrophy	SLE	Systemic lupus erythemotodes
RVOT	Right ventricular outflow tract	SMA	Sequential multiple analyzer
RZ	Return-to-zero	SMAC	Sequential multiple analyzer with computer
SA	Sinoatrial; Specific absorption	SMR	Sensorimotor
SACH	Solid-ankle-cushion-heel	S/N	Signal-to-noise
SAD	Source-axis distance; Statistical Analysis System	S:N/D	Signal-to-noise ratio per unit dose
SAINT	System analysis of integrated network of tasks	SNP	Sodium nitroprusside
SAL	Sterility assurance level; Surface averaged lead	SNR	Signal-to-noise ratio
SALT	Systematic analysis of language transcripts	SOA	Sources of artifact
SAMI	Socially acceptable monitoring instrument	SOAP	Subjective, objective, assessment, plan
SAP	Systemic arterial pressure	SOBP	Spread-out Bragg peak
SAR	Scatter-air ratio; Specific absorption rate	SP	Skin potential
SARA	System for anesthetic and respiratory gas analysis	SPECT	Single photon emission computed tomography
SBE	Subbacterial endocarditis	SPL	Sound pressure level
SBR	Styrene-butadiene rubbers	SPRINT	Single photon ring tomograph
SC	Stratum corneum; Subcommittees	SPRT	Standard platinum resistance thermometer
SCAP	Right scapula	SPSS	Statistical Package for the Social Sciences
SCE	Saturated calomel electrode; Sister chromatid exchange	SQUID	Superconducting quantum interference device
SCI	Spinal cord injury	SQV	Square wave voltammetry
SCRAD	Sub-Committee on Radiation Dosimetry	SR	Polysulfide rubbers
SCS	Spinal cord stimulation	SRT	Speech reception threshold
SCUBA	Self-contained underwater breathing apparatus	SS	Stainless steel
SD	Standard deviation	SSB	Single strand breaks
SDA	Stepwise discriminant analysis	SSD	Source-to-skin distance; Source-to-surface distance
SDS	Sodium dodecyl sulfate	SSE	Stainless steel electrode
S&E	Safety and effectiveness	SSEP	Somatosensory evoked potential
SE	Standard error	SSG	Solid state generator
SEC	Size exclusion chromatography	SSP	Skin stretch potential
SEM	Scanning electron microscope; Standard error of the mean	SSS	Sick sinus syndrome
SEP	Somatosensory evoked potential	STD	Source-tray distance
SEXAFS	Surface extended X-ray absorption fine structure	STI	Systolic time intervals
SF	Surviving fraction	STP	Standard temperature and pressure
SFD	Source-film distance	STPD	Standard temperature pressure dry
SFH	Stroma-free hemoglobin	SV	Stroke volume
SFTR	Sagittal frontal transverse rotational	SVC	Superior vena cava
SG	Silica gel	SW	Standing wave
SGF	Silica gel fraction	TAA	Tumor-associated antigens
SGG	Spark gap generator	TAC	Time-averaged concentration
SGOT	Serum glutamic oxaloacetic transaminase	TAD	Transverse abdominal diameter
SGP	Strain gage plethysmography; Stress-generated potential	TAG	Technical Advisory Group
SHE	Standard hydrogen electrode	TAH	Total artificial heart
SI	Le Système International d'Unités	TAR	Tissue-air ratio
		TC	Technical Committees
		TCA	Tricarboxylic acid cycle
		TCD	Thermal conductivity detector
		TCES	Transcutaneous cranial electrical stimulation

TCP	Tricalcium phosphate	UHMWPE	Ultra high molecular weight polyethylene
TDD	Telecommunication devices for the deaf	UL	Underwriters Laboratory
TDM	Therapeutic drug monitoring	ULF	Ultralow frequency
TE	Test electrode; Thermoplastic elastomers	ULTI	Ultralow temperature isotropic
TEAM	Technology evaluation and acquisition methods	UMN	Upper motor neuron
TEM	Transmission electron microscope; Transverse electric and magnetic mode; Transverse electromagnetic mode	UO	Urinary output
TENS	Transcutaneous electrical nerve stimulation	UPTD	Unit pulmonary oxygen toxicity doses
TEP	Tracheoesophageal puncture	UR	Unconditioned response
TEPA	Triethylenephosphoramidate	US	Ultrasound; Unconditioned stimulus
TF	Transmission factor	USNC	United States National Committee
TFE	Tetrafluorethylene	USP	United States Pharmacopeia
TI	Totally implantable	UTS	Ultimate tensile strength
TICCIT	Time-shared Interaction Computer-Controlled Information Television	UV	Ultraviolet; Umbilical vessel
TLC	Thin-layer chromatography; Total lung capacity	UVR	Ultraviolet radiation
TLD	Thermoluminescent dosimetry	V/F	Voltage-to-frequency
TMJ	Temporomandibular joint	VA	Veterans Administration
TMR	Tissue maximum ratio; Topical magnetic resonance	VAS	Visual analog scale
TNF	Tumor necrosis factor	VBA	Vaginal blood volume in arousal
TOF	Train-of-four	VC	Vital capacity
TP	Thermal performance	VCO	Voltage-controlled oscillator
TPC	Temperature pressure correction	VDT	Video display terminal
TPD	Triphasic dissociation	VECG	Vectorelectrocardiography
TPG	Transvalvular pressure gradient	VEP	Visually evoked potential
TPN	Total parenteral nutrition	VF	Ventricular fibrillation
TR	Temperature rise	VOP	Venous occlusion plethysmography
tRNA	Transfer RNA	VP	Ventriculoperitoneal
TSH	Thyroid stimulating hormone	VPA	Vaginal pressure pulse in arousal
TSS	Toxic shock syndrome	VPB	Ventricular premature beat
TTD	Telephone devices for the deaf	VPR	Volume pressure response
TTI	Tension time index	VSD	Ventricular septal defect
TTR	Transition temperature range	VSWR	Voltage standing wave ratio
TTV	Trimming tip version	VT	Ventricular tachycardia
TTY	Teletypewriter	VTG	Vacuum tube generator
TUR	Transurethral resection	VTS	Viewscan text system
TURP	Transurethral resections of the prostate	VV	Variable version
TV	Television; Tidal volume; Tricuspid valve	WAIS-R	Weschler Adult Intelligence Scale-Revised
TVER	Transscleral visual evoked response	WAK	Wearable artificial kidney
TW	Traveling wave	WAML	Wide-angle mobility light
TxB ₂	Thromboxane B ²	WBAR	Whole-body autoradiography
TZ	Transformation zone	WBC	White blood cell
UES	Upper esophageal sphincter	WG	Working Groups
UP	Urea-formaldehyde	WHO	World Health Organization; Wrist hand orthosis
UffIS	University Hospital Information System	WLF	Williams-Landel-Ferry
UHMW	Ultra high molecular weight	WMR	Work metabolic rate
		w/o	Weight percent
		WORM	Write once, read many
		WPW	Wolff-Parkinson-White
		XPS	X-ray photon spectroscopy
		XR	Xeroradiograph
		YAG	Yttrium aluminum garnet
		ZPL	Zero pressure level

CONVERSION FACTORS AND UNIT SYMBOLS

SI UNITS (ADOPTED 1960)

A new system of metric measurement, the International System of Units (abbreviated SI), is being implemented throughout the world. This system is a modernized version of the MKSA (meter, kilogram, second, ampere) system, and its details are published and controlled by an international treaty organization (The International Bureau of Weights and Measures).

SI units are divided into three classes:

Base Units	
length	meter [†] (m)
mass [‡]	kilogram (kg)
time	second (s)
electric current	ampere (A)
thermodynamic temperature§	kelvin (K)
amount of substance	mole (mol)
luminous intensity	candela (cd)
Supplementary Units	
plane angle	radian (rad)
solid angle	steradian (sr)

Derived Units and Other Acceptable Units

These units are formed by combining base units, supplementary units, and other derived units. Those derived units having special names and symbols are marked with an asterisk (*) in the list below:

<i>Quantity</i>	<i>Unit</i>	<i>Symbol</i>	<i>Acceptable equivalent</i>
*absorbed dose	gray	Gy	J/kg
acceleration	meter per second squared	m/s ²	
*activity (of ionizing radiation source)	becquerel	Bq	1/s
area	square kilometer	km ²	
	square hectometer	hm ²	ha (hectare)
	square meter	m ²	

[†]The spellings “metre” and “litre” are preferred by American Society for Testing and Materials (ASTM); however, “-er” will be used in the Encyclopedia.

[‡]“Weight” is the commonly used term for “mass.”

§Wide use is made of “Celsius temperature” (*t*) defined $t = T - T_0$ where *T* is the thermodynamic temperature, expressed in kelvins, and $T_0 = 273.15$ K by definition. A temperature interval may be expressed in degrees Celsius as well as in kelvins.

<i>Quantity equivalent</i>	<i>Unit</i>	<i>Symbol</i>	<i>Acceptable</i>
* capacitance	farad	F	C/V
concentration (of amount of substance)	mole per cubic meter	mol/m ³	
* conductance	siemens	S	A/V
current density	ampere per square meter	A/m ²	
density, mass density	kilogram per cubic meter	kg/m ³	g/L; mg/cm ³
dipole moment (quantity)	coulomb meter	C·m	
* electric charge, quantity of electricity	coulomb	C	A·s
electric charge density	coulomb per cubic meter	C/m ³	
electric field strength	volt per meter	V/m	
electric flux density	coulomb per square meter	C/m ²	
* electric potential, potential difference, electromotive force	volt	V	W/A
* electric resistance	ohm	Ω	V/A
* energy, work, quantity of heat	megajoule	MJ	
	kilojoule	kJ	
	joule	J	N·m
	electron volt [†]	eV [†]	
	kilowatt hour [†]	kW·h [†]	
energy density	joule per cubic meter	J/m ³	
* force	kilonewton	kN	
	newton	N	kg·m/s ²
* frequency	megahertz	MHz	
	hertz	Hz	1/s
heat capacity, entropy	joule per kelvin	J/K	
heat capacity (specific), specific entropy	joule per kilogram kelvin	J/(kg·K)	
heat transfer coefficient	watt per square meter kelvin	W/(m ² ·K)	
* illuminance	lux	lx	lm/m ²
* inductance	henry	H	Wb/A
linear density	kilogram per meter	kg/m	
luminance	candela per square meter	cd/m ²	
* luminous flux	lumen	lm	cd·sr
magnetic field strength	ampere per meter	A/m	
* magnetic flux	weber	Wb	V·s
* magnetic flux density	tesla	T	Wb/m ²
molar energy	joule per mole	J/mol	
molar entropy, molar heat capacity	joule per mole kelvin	J/(mol·K)	
moment of force, torque	newton meter	N·m	
momentum	kilogram meter per second	kg·m/s	
permeability	henry per meter	H/m	
permittivity	farad per meter	F/m	
* power, heat flow rate, radiant flux	kilowatt	kW	
	watt	W	J/s
power density, heat flux density, irradiance	watt per square meter	W/m ²	
* pressure, stress	megapascal	MPa	
	kilopascal	kPa	
	pascal	Pa	N/m ²
sound level	decibel	dB	
specific energy	joule per kilogram	J/kg	
specific volume	cubic meter per kilogram	m ³ /kg	
surface tension	newton per meter	N/m	
thermal conductivity	watt per meter kelvin	W/(m·K)	
velocity	meter per second	m/s	
	kilometer per hour	km/h	
viscosity, dynamic	pascal second	Pa·s	
	millipascal second	mPa·s	

[†]This non-SI unit is recognized as having to be retained because of practical importance or use in specialized fields.

<i>Quantity</i>	<i>Unit</i>	<i>Symbol</i>	<i>Acceptable equivalent</i>
viscosity, kinematic	square meter per second	m ² /s	
	square millimeter per second	mm ² /s	
	cubic meter	m ³	
	cubic decimeter	dm ³	L(liter)
	cubic centimeter	cm ³	mL
wave number	1 per meter	m ⁻¹	
	1 per centimeter	cm ⁻¹	

In addition, there are 16 prefixes used to indicate order of magnitude, as follows:

<i>Multiplication factor</i>	<i>Prefix</i>	<i>Symbol</i>	<i>Note</i>
10 ¹⁸	exa	E	
10 ¹⁵	peta	P	
10 ¹²	tera	T	
10 ⁹	giga	G	
10 ⁸	mega	M	
10 ³	kilo	k	
10 ²	hecto	h ^a	^a Although hecto, deka, deci, and centi are SI prefixes, their use should be avoided except for SI unit-multiples for area and volume and nontechnical use of centimeter, as for body and clothing measurement.
10	deka	da ^a	
10 ⁻¹	deci	d ^a	
10 ⁻²	centi	c ^a	
10 ⁻³	milli	m	
10 ⁻⁶	micro	μ	
10 ⁻⁹	nano	n	
10 ⁻¹²	pico	p	
10 ⁻¹⁵	femto	f	
10 ⁻¹⁸	atto	a	

For a complete description of SI and its use the reader is referred to ASTM E 380.

CONVERSION FACTORS TO SI UNITS

A representative list of conversion factors from non-SI to SI units is presented herewith. Factors are given to four significant figures. Exact relationships are followed by a dagger (†). A more complete list is given in ASTM E 380-76 and ANSI Z210.1-1976.

<i>To convert from</i>	<i>To</i>	<i>Multiply by</i>
acre	square meter (m ²)	4.047 × 10 ³
angstrom	meter (m)	1.0 × 10 ^{-10†}
are	square meter (m ²)	1.0 × 10 ^{2†}
astronomical unit	meter (m)	1.496 × 10 ¹¹
atmosphere	pascal (Pa)	1.013 × 10 ⁵
bar	pascal (Pa)	1.0 × 10 ^{5†}
barrel (42 U.S. liquid gallons)	cubic meter (m ³)	0.1590
Btu (International Table)	joule (J)	1.055 × 10 ³
Btu (mean)	joule (J)	1.056 × 10 ³
Bt (thermochemical)	joule (J)	1.054 × 10 ³
bushel	cubic meter (m ³)	3.524 × 10 ⁻²
calorie (International Table)	joule (J)	4.187
calorie (mean)	joule (J)	4.190
calorie (thermochemical)	joule (J)	4.184 [†]
centimeters of water (39.2 °F)	pascal (Pa)	98.07
centipoise	pascal second (Pa·s)	1.0 × 10 ^{-3†}
centistokes	square millimeter per second (mm ² /s)	1.0 [†]

<i>To convert from</i>	<i>To</i>	<i>Multiply by</i>
cfm (cubic foot per minute)	cubic meter per second (m ³ /s)	4.72 × 10 ⁻⁴
cubic inch	cubic meter (m ³)	1.639 × 10 ⁻⁴
cubic foot	cubic meter (m ³)	2.832 × 10 ⁻²
cubic yard	cubic meter (m ³)	0.7646
curie	becquerel (Bq)	3.70 × 10 ^{10†}
debye	coulomb-meter (C·m)	3.336 × 10 ⁻³⁰
degree (angle)	radian (rad)	1.745 × 10 ⁻²
denier (international)	kilogram per meter (kg/m)	1.111 × 10 ⁻⁷
	tex	0.1111
dram (apothecaries')	kilogram (kg)	3.888 × 10 ⁻³
dram (avoirdupois)	kilogram (kg)	1.772 × 10 ⁻³
dram (U.S. fluid)	cubic meter (m ³)	3.697 × 10 ⁻⁶
dyne	newton(N)	1.0 × 10 ^{-6†}
dyne/cm	newton per meter (N/m)	1.00 × 10 ^{-3†}
electron volt	joule (J)	1.602 × 10 ⁻¹⁹
erg	joule (J)	1.0 × 10 ^{-7†}
fathom	meter (m)	1.829
fluid ounce (U.S.)	cubic meter (m ³)	2.957 × 10 ⁻⁵
foot	meter (m)	0.3048†
foot-pound force	joule (J)	1.356
foot-pound force	newton meter (N·m)	1.356
foot-pound force per second	watt(W)	1.356
footcandle	lux (lx)	10.76
furlong	meter (m)	2.012 × 10 ²
gal	meter per second squared (m/s ²)	1.0 × 10 ^{-2†}
gallon (U.S. dry)	cubic meter (m ³)	4.405 × 10 ⁻³
gallon (U.S. liquid)	cubic meter (m ³)	3.785 × 10 ⁻³
gilbert	ampere (A)	0.7958
gill (U.S.)	cubic meter (m ³)	1.183 × 10 ⁻⁴
grad	radian	1.571 × 10 ⁻²
grain	kilogram (kg)	6.480 × 10 ⁻⁵
gram force per denier	newton per tex (N/tex)	8.826 × 10 ⁻²
hectare	square meter (m ²)	1.0 × 10 ^{4†}
horsepower (550 ft·lbf/s)	watt(W)	7.457 × 10 ²
horsepower (boiler)	watt(W)	9.810 × 10 ³
horsepower (electric)	watt(W)	7.46 × 10 ^{2†}
hundredweight (long)	kilogram (kg)	50.80
hundredweight (short)	kilogram (kg)	45.36
inch	meter (m)	2.54 × 10 ^{-2†}
inch of mercury (32 °F)	pascal (Pa)	3.386 × 10 ³
inch of water (39.2 °F)	pascal (Pa)	2.491 × 10 ²
kilogram force	newton (N)	9.807
kilopond	newton (N)	9.807
kilopond-meter	newton-meter (N·m)	9.807
kilopond-meter per second	watt (W)	9.807
kilopond-meter per min	watt(W)	0.1635
kilowatt hour	megajoule (MJ)	3.6†
kip	newton (N)	4.448 × 10 ²
knot international	meter per second (m/s)	0.5144
lambert	candela per square meter (cd/m ²)	3.183 × 10 ³
league (British nautical)	meter (m)	5.559 × 10 ²
league (statute)	meter (m)	4.828 × 10 ³
light year	meter (m)	9.461 × 10 ¹⁵
liter (for fluids only)	cubic meter (m ³)	1.0 × 10 ^{-3†}
maxwell	weber (Wb)	1.0 × 10 ^{-8†}
micron	meter (m)	1.0 × 10 ^{-6†}
mil	meter (m)	2.54 × 10 ^{-5†}
mile (U.S. nautical)	meter (m)	1.852 × 10 ^{3†}
mile (statute)	meter (m)	1.609 × 10 ³
mile per hour	meter per second (m/s)	0.4470

<i>To convert from</i>	<i>To</i>	<i>Multiply by</i>
millibar	pascal (Pa)	1.0×10^2
millimeter of mercury (0 °C)	pascal (Pa)	$1.333 \times 10^{2\dagger}$
millimeter of water (39.2 °F)	pascal (Pa)	9.807
minute (angular)	radian	2.909×10^{-4}
myriagram	kilogram (kg)	10
myriameter	kilometer (km)	10
oersted	ampere per meter (A/m)	79.58
ounce (avoirdupois)	kilogram (kg)	2.835×10^{-2}
ounce (troy)	kilogram (kg)	3.110×10^{-2}
ounce (U.S. fluid)	cubic meter (m ³)	2.957×10^{-5}
ounce-force	newton (N)	0.2780
peck (U.S.)	cubic meter (m ³)	8.810×10^{-3}
pennyweight	kilogram (kg)	1.555×10^{-3}
pint (U.S. dry)	cubic meter (m ³)	5.506×10^{-4}
pint (U.S. liquid)	cubic meter (m ³)	4.732×10^{-4}
poise (absolute viscosity)	pascal second (Pa·s)	0.10 [†]
pound (avoirdupois)	kilogram (kg)	0.4536
pound (troy)	kilogram (kg)	0.3732
poundal	newton (N)	0.1383
pound-force	newton (N)	4.448
pound per square inch (psi)	pascal (Pa)	6.895×10^3
quart (U.S. dry)	cubic meter (m ³)	1.101×10^{-3}
quart (U.S. liquid)	cubic meter (m ³)	9.464×10^{-4}
quintal	kilogram (kg)	$1.0 \times 10^{2\dagger}$
rad	gray (Gy)	$1.0 \times 10^{-2\dagger}$
rod	meter (m)	5.029
roentgen	coulomb per kilogram (C/kg)	2.58×10^{-4}
second (angle)	radian (rad)	4.848×10^{-6}
section	square meter (m ²)	2.590×10^6
slug	kilogram (kg)	14.59
spherical candle power	lumen (lm)	12.57
square inch	square meter (m ²)	6.452×10^{-4}
square foot	square meter (m ²)	9.290×10^{-2}
square mile	square meter (m ²)	2.590×10^6
square yard	square meter (m ²)	0.8361
store	cubic meter (m ³)	1.0 [†]
stokes (kinematic viscosity)	square meter per second (m ² /s)	$1.0 \times 10^{-4\dagger}$
tex	kilogram per meter (kg/m)	$1.0 \times 10^{-6\dagger}$
ton (long, 2240 pounds)	kilogram (kg)	1.016×10^3
ton (metric)	kilogram (kg)	$1.0 \times 10^{3\dagger}$
ton (short, 2000 pounds)	kilogram (kg)	9.072×10^2
torr	pascal (Pa)	1.333×10^2
unit pole	weber (Wb)	1.257×10^{-7}
yard	meter (m)	0.9144 [†]

RADIOTHERAPY, COBALT 60 UNITS FOR.

See COBALT 60 UNITS FOR RADIOTHERAPY.

RADIOTHERAPY, HEAVY IONS AND ELECTRONS

F. D. BECCHETTI
University of Michigan
Ann Arbor, Michigan

INTRODUCTION

The use of subatomic particles to treat cancer and other medical conditions can be traced back (1) to the discovery of natural radioactivity by Bequerel in 1896. Marie Sklodowska Curie and her husband Pierre Curie quickly identified the primary radiation emitted by radioactive materials such as uranium and thorium as consisting of three principal types: α , β , and γ rays. Among these rays, only γ rays, being high energy photons, are a form of electromagnetic radiation. Therefore, they are similar to lower energy photons and, in particular, X rays with regard to their interaction in matter (2–4).

In contrast, the α and β rays were observed to be much more ionizing than γ rays, and unlike γ rays had a finite range in materials. The latter feature is characteristic of energetic subatomic particles and, indeed, the α ray was later identified as an energetic 4 He ion. It is emitted via nuclear α decay from heavy radioactive elements. Likewise, β rays were identified as energetic electrons emitted via nuclear β decay from light and heavy radioactive elements. Subsequently, the Curies were able to identify one particular highly radioactive element, radium. With considerable time and effort, they were able to extract small but usable pure samples of this from large amounts of uranium ore. As is also the case with X rays, workers using strong radioactive sources, including the Curies, often developed skin rashes and other symptoms related to exposure to natural radiation from radioactive materials. Medical doctors quickly realized that this radiation also could be used in medical applications.

Thus, following closely on the work by Roentgen et al. on the application of manmade radiation (i.e., X rays) to treat cancer, the Curies and others used radium and other natural radioactive sources to treat cancer tumors (Fig. 1) (1). Although the α , β , and γ rays emitted from nuclear decay were emitted with a much higher energy (MeV vs keV for X rays), the α and β particles had very short ranges in tissue (e.g., a mm or less for α particles and a few cm for β particles). Therefore, these sources, particularly including the associated MeV gamma rays, primarily were used to treat surface tumors (Fig. 1). An early form of brachytherapy using radioactive needles also was developed to treat deeper tumors (1).



Figure 1. An early form of radiation treatment (ca. 1920) using radioactive sources (1).

It would await the development of high energy particle accelerators after WWII to provide α and β particles (helium ions and electrons) with sufficient energy and intensity to be effective in treating deep tumors (5). However, belatedly, in 1932, Chadwick discovered another basic subatomic particle, the neutron. Being uncharged, neutrons unlike α and β particles behave more like X rays and γ rays in tissue. The discovery of the neutron coincided with the invention of the cyclotron by E. Lawrence, and an early use of the cyclotron was to produce energetic neutrons (5–7). Stone et al. proposed to use neutrons for the treatment of cancer and many treatments were performed in the late 1930s.

The modern era of particle-beam radiotherapy begins after WWII with the development of linear accelerators (LINACs), betatrons, the synchro-cyclotron, and the synchrotron (5–8) to provide electron, proton, α and heavy-ion beams at energies sufficient to penetrate many centimeters of tissue. Much of this work was pioneered at the University of California-Berkeley by E. Lawrence, his brother

Table 1. Worldwide Charged Particle Patient Totals

January 2005 WHO	WHERE	WHAT	DATE FIRST RX	DATE LAST RX	RECENT PATIENT TOTAL	DATE OF TOTAL
Berkeley 184	CA. USA	p	1954	-1957	30	
Berkeley	CA. USA	He	1957	-1992	2054	
Uppsala (1)	Sweden	p	1957	-1976	73	
Harvard	MA. USA	p	1961	-2002	9116	
Dubna (1)	Russia	p	1967	-1996	124	
ITEP, Moscow	Russia	p	1969		3785	Dec-04
Los Alamos	NM. USA	π^-	1974	-1982	230	
St. Petersburg	Russia	p	1975		1145	April-04
Berkeley	CA. USA	ion	1975	-1992	433	
Chiba	Japan	p	1979		145	Apr-02
TRIUMF	Canada	π^-	1979	-1994	367	
PSI (SIN)	Switzerland	π^-	1980	-1993	503	
PMRC (1), Tsukuba	Japan	p	1983	-2000	700	
PSI (72 MeV)	Switzerland	p	1984		4182	Dec-04
Uppsala (2)	Sweden	p	1989		418	Jan-04
Clatterbridge	England	p	1989		1372	Dec-04
Loma Linda	CA, USA	p	1990		9585	Nov-04
Louvain-la-Neuve	Belgium	p	1991	-1993	21	
Nice	France	p	1991		2555	April-04
Orsay	France	p	1991		2805	Dec-03
iThemba LABS	South Africa	p	1993		468	Nov-04
MPRI (1)	IN USA	p	1993	-1999	34	
UCSF - CNL	CA USA	p	1994		632	June-04
HIMAC, Chiba	Japan	C ion	1994		1796	Feb-04
TRIUMF	Canada	p	1995		89	Dec-03
PSI (200 MeV)	Switzerland	p	1996		209	Dec-04
G.S.I Darmstadt	Germany	C ion	1997		198	Dec-03
H.M.I, Berlin	Germany	p	1998		546	Dec-04
NCC, Kashiwa	Japan	p	1998		300	Oct-04
Dubna (2)	Russia	p	1999		296	Dec-04
HIBMC, Hyogo	Japan	p	2001		483	Dec-04
PMRC (2), Tsukuba	Japan	p	2001		492	July 04
NPTC, MGH	MA USA	p	2001		973	Dec-04
HIBMC, Hyogo	Japan	C ion	2002		30	Dec-02
INFN-LNS, Catania	Italy	p	2002		82	Oct-04
WERC	Japan	p	2002		19	Oct-04
Shizuoka	Japan	p	2003		100	Dec-04
MPRI (2)	IN USA	p	2004		21	July-04
Wanjie, Zibo	China	p	2004		1	Dec-04
					1100	pions
					4511	ions
					40801	protons
				TOTAL	46412	all particles

Adopted from PTOG 35 Newsletter (Jan. 2005) (12).

John, a medical doctor, and C. Tobias. Using the 184 inch synchro-cyclotron to accelerate high energy α particles, an active radiation therapy program, as an adjunct to the primary nuclear physics research program, was carried out from 1954 to 1986 (Table 1) (9–12).

The justification for using a heavy, charged subatomic particle such as a proton, α or heavier ion lies in the fact that the main kinetic energy loss per path length in tissue ($\Delta E/\Delta X$) occurs primarily via collisional losses i.e., via many collisions with atomic electrons (2). Each collision absorbs a small amount of energy with only minimal scattering of the incident particle, which produces a well-defined energy loss vs depth curve (linear-energy transfer or LET) and hence range of the particle in tissue or its near equivalent, water (Fig. 2) (13). In medical

physics, LET is usually specified in $\text{keV}/\mu\text{m}$ or, alternatively, $\text{keV}/\text{g}/\text{cm}^2$ for a specific medium. Specifically, in a given material, the LET for a charged, nonrelativistic heavy particle of mass m and atomic number z moving at a velocity v [hence, kinetic energy $E = (1/2)mv^2$] has the behavior (2–5)

$$\begin{aligned} \text{LET} &= k z^2/v^2 \\ &= K z^2/(E/m) \end{aligned} \quad (1)$$

where k and K are constants that depend on the atomic composition of the medium. Suitable integration of LET from $E = \text{incident } E (= E_0)$ to $E = 0$ then gives the range, R of the particle in the medium. Based on the above, we would expect R to then be proportional to E_0^n with $n = 2$, which is approximately true but, because of various

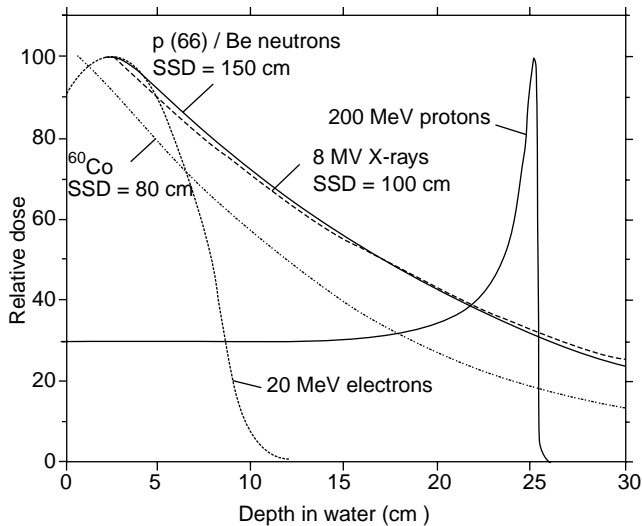


Figure 2. Depth-dose curves in water (which is similar to soft tissue) for various types of photons and particles commonly used in radiation therapy (arbitrary normalization; adapted from Ref. 13).

atomic effects (2–5), one typically has $1.4 < n < 2$ depending on z and E_0 .

Owing to the $1/E$ dependence of LET (Eq. 1, above), protons, alphas, and heavier particles such as ^{12}C ions exhibit a very high LET at the end of their range, resulting in a sharp “Bragg peak” (Fig. 2). In cancer treatment, this sharp LET peak is often then spread out using energy-loss absorbers or other means to produce a spread-out Bragg peak (SOBP) suitable for treating an extended tumor (Fig. 3).

In the case of heavy particles, the biological dose will depend on the LET as well as the relative biological effectiveness (RBE) of the incident particle (2). The RBE may

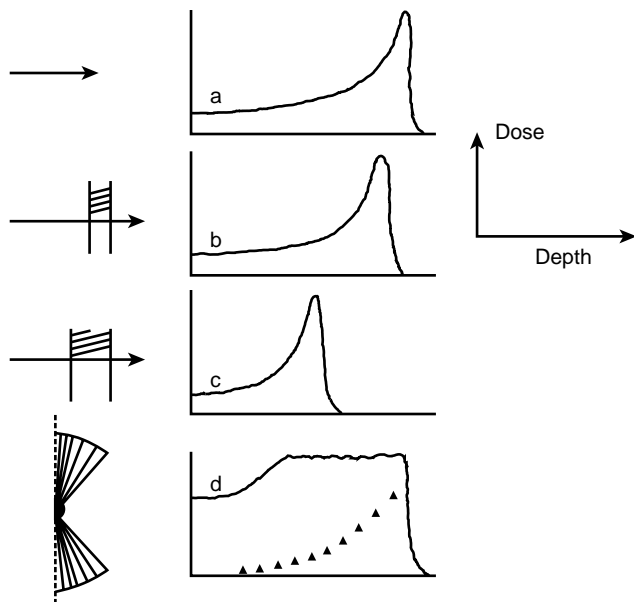


Figure 3. Illustration showing modification of a particle depth-dose curve by means of a spread-out Bragg peak (SOBP; adapted from Ref. 5).

depend on m , z , E , and several other factors including the LET. Although electrons, X rays, and γ rays generally have an RBE near unity, heavy particles (including neutrons) can have much larger values *viz.* $\text{RBE} = 4$ or greater (5–11,14,15). In addition, energetic heavy particles can lose energy and even transmute into a different particle via nuclear reactions and fragmentation (9–11,14–16), which can extend the net LET beyond the range of the incident particle due to atomic collisional losses alone and also introduce scattering of the incident beam. These features can become significant for heavier particles such as α particles and ^{12}C ions.

In contrast, photons such as X rays and γ rays lose energy through several different mechanisms (2), which normally only involve a few collisions for a particular photon. Each collision usually produces a scattered, secondary electron that can carry away (and hence absorb) significant energy from the photon as well as scatter the photon. The energy loss of all the secondary electrons and their RBE (near unity) yields the biological radiation dose attributed to the photon. One can sum this dose over many incident photons to generate a dose-distance curve for a beam of photons (Fig. 2). However, no single X-ray or γ -ray photon has a well-defined energy-loss curve or range per se. Instead, like visible-light photons, a beam of such radiation primarily will be attenuated or scattered as it passes through tissue and thus yield the dose–distance curves shown in Fig. 2.

In contrast to charged particles, most of the dose for a single photon beam thus occurs near the surface and continues through the treatment area including the exit region (Fig. 2). Treating deep-seated tumors while sparing adjacent healthy tissue from a lethal dose requires multiaxis beam treatment planning together with fractionation of the dose (15–20), which can present a problem for tumors near critical organs, such as the spinal cord, and optical nerve, and, for such tumors, certain particle beams can be advantageous.

Energetic electron beams exhibit characteristics of both particle beams as well as X rays and γ rays. Although the primary energy loss is via collisions with atomic electrons, owing to the incident mass also being that of an electron, this energy loss is accompanied with significant scattering of the primary electrons (2), which results in a complicated dose-distance profile (Figs. 2 and 4) (21). In addition, the electrons used in radiation therapy are at relativistic energies (i.e., $E_0 > 1$ MeV) and their range, unlike that of heavier particles, increases more slowly (2) with incident energy ($n \approx 1$), but, due to scattering, the range along the incident beam axis is not well defined. Hence, the large penumbra exhibited (Fig. 4) can limit the usefulness of electron beams in treating deep tumors. The application of a magnetic field to reduce the penumbra has been proposed and is under active study by several research groups (see below).

Subatomic antimatter particles such as antiprotons and positrons (antielectrons) have also been proposed for cancer therapy, which is based on past work (14–22) done at the Los Alamos Meson Factory (LAMPF) using another form of antimatter, the negative pi meson (negative pion). Antimatter particles like the negative pion undergo a

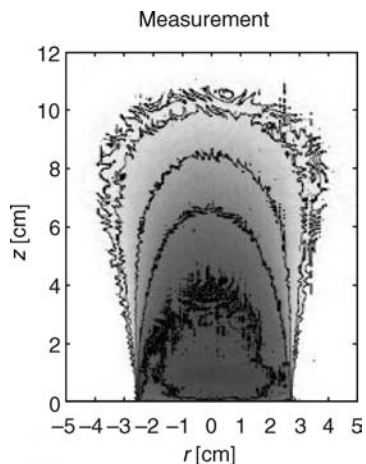


Figure 4. A 2D profile of a ca. 20 MeV electron beam incident on a tissue-equivalent plastic phantom (21).

nuclear annihilation reaction when they stop near the end of their range, which converts the mass of the antiparticle and a related nuclear constituent into photon energy, which produces additional radiation to enhance the dose near the end of the range (and enhanced dose peak, Fig. 5). The LAMPF, TRIUMF, and SIN nuclear research laboratories had active negative pion radiation oncology treatment programs from the early 1970s to the early 1990s (Table 1). However, owing to practical considerations associated with the large accelerator needed, this type of treatment is no longer in active use.

A characterization of the dose and LET of various particle beams used in radiation oncology is displayed in Fig. 6 (23). The features noted can help determine the type of beam and treatment plan that may be optimal to treat a specific type of tumor. Table 2 lists the various types of particle-beam modalities, their characteristics, and the type of cancers generally treated with each modality (24).

Until recently, excluding electrons, most particle-beam cancer treatment facilities used accelerators primarily designed and operated for nuclear physics research. Such accelerators were generally not optimized for radiation

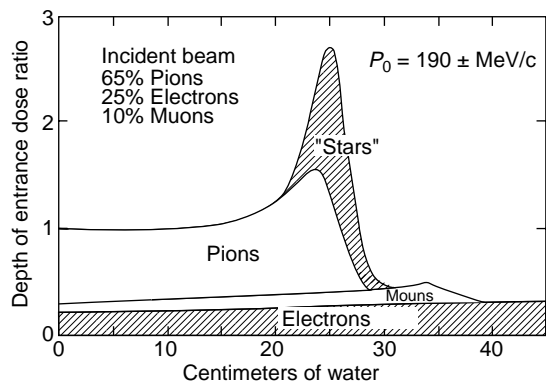


Figure 5. Dose-depth curve for negative pions (a form of antimatter) in water. Note the enhanced dose at the end of the range due to annihilation radiation (adopted from Ref. 5).

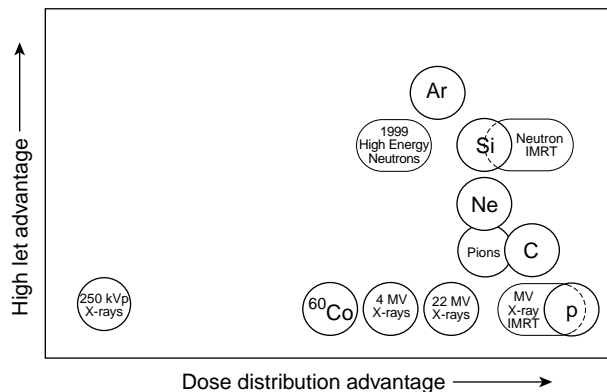


Figure 6. Characteristics of dose and LET for various forms of radiotherapy beams (23).

therapy and were costly to operate and maintain. As an example, nuclear research accelerators are often designed for much higher beam intensities and beam energy than needed for tumor treatment. Beam time also may be limited, which complicates treatment by requiring fractionated doses, and the patient-related aspects in a research laboratory are not ideal in most cases. Nonetheless, these facilities, many still in operation, demonstrated the effectiveness of particle-beam therapy for certain types of tumor treatment, which has led to the construction (or planning) of several dedicated particle-beam cancer treatment facilities throughout the world. A listing of these, together with some of the pioneering, older charged-particle facilities is given in Table 1.

In the following sections, we will describe in more detail specific types of particle-beam treatments and related apparatus. Additional information may be found in other sections of this volume as well as in several review articles (9–11,14).

In addition, several groups such as the Proton Therapy Co-operative Group (PTCOG) (25) and the TERA Foundation (26) issue newsletters and hold regular conferences with proceedings covering recent advances in particle-beam therapy.

Many of the techniques used in particle-beam radiation therapy are similar to those used in more conventional radiation therapy using photons, and extensive literature for the latter exists (17–20, 27–30). Likewise, several detailed references discuss the effect of radiation on cell survival and, hence, RBE for particular types of radiation including particle beams (2–5,7,15,31–35).

ELECTRON-BEAM RADIOTHERAPY

The most common type of particle-beam radiotherapy involves the use of energetic electrons, either directly from an electron accelerator (8,17,27,36) or as the ionizing radiation in radio-isotope brachytherapy. Most hospital-based radiation-oncology departments use an electron LINAC, typically with electron beam energies from 5 MeV to 25 MeV. This beam produces, via bremsstrahlung, on a tungsten or similar target, the megavolt X rays (i.e., high energy photons used for standard radiotherapy)

Table 2. Summary of Present External Beam Radiotherapy Options for Malignant Tumors^a

Particle	Tumor Characteristics	Energy deposition	Bragg peak	Radiation source	Accelerator cost ^b
Photons	Rapidly growing, oxygenated	Low LET	No	Cobalt 60; electron linac; microtron	1
Electrons	Superficial	Low LET	No	Electron Linac; microtron	1–2
Protons	Early stage, near-critical structures	Low LET	Yes	Synchrotron; cyclotron	10–15
Fast neutrons	Slow growing, hypoxic	High LET	No	Proton linac; cyclotron;	8–10
Heavy ions	Same as fast neutrons	High LET	Yes	Synchrotron	40
Pions	Same as fast neutrons	High LET	Yes	Proton linac; cyclotron	35–40
Slow neutrons	Glioblastoma; Some melanomas	Very high LET with BNCT	No	Low energy accelerator; nuclear reactor	1–2

^aAdopted from Table 3 of Ref. 5.

^bRelative cost not including building and clinical equipment costs. Assumes room-temperature magnets. Superconducting magnets can reduce accelerator cost in some cases (e.g., see Ref. 24).

(Fig. 7) (8). The electron LINAC usually is used in conjunction with an isocentric gantry (Fig. 8) (8) to permit stereotactic treatment with minimal movement of the patient.

In principle, with an electron beam of sufficient intensity and energy, the direct beam also can be used for radiotherapy. However, as noted, the large collisional scattering of the electrons in tissue results in a large angular spread (i.e., penumbra) (2) of the beam (Fig. 4), which limits the usefulness of the direct beam for therapy other than for treatment of skin cancer, other shallow tumors, or for noncancerous skin diseases (Table 2). However, owing to their compactness, a gantry-mounted electron-beam system (Fig. 8) can also be used in an operating room to directly irradiate an internal tumor (or surrounding area) made accessible via surgery. This technique, interoperative electron-beam radiation therapy (IOERT), may be particularly advantageous for treatment of certain types of cancers when combined with more conventional treatment (37).

It was once thought that higher energy electrons ($E > 25$ MeV) and the corresponding high energy X rays produced could have advantages in radiotherapy relative to the more typical energies used ($E < 25$ MeV). These higher energy electrons were typically produced using a special LINAC or a compact race-track microtron (36) accelerator (Fig. 9) with a separate gantry for the beam (Fig. 8). High energy electrons possibly suitable for radiation therapy also recently have been produced using high power compact pulsed lasers (38).

At the higher electron energies, nuclear reactions and significant bremsstrahlung can take place and contribute to the patient dose. Although the higher energy may have some benefit, it is generally offset by the larger dose imparted to surrounding healthy tissue because of the longer range of the secondary electrons involved. Likewise, improvements in radiotherapy techniques using lower energy photons, such as intensity-modulated radiotherapy (IMRT) and Monte-Carlo treatment planning (17,18), have offset most of the advantages of higher energies. Thus, few

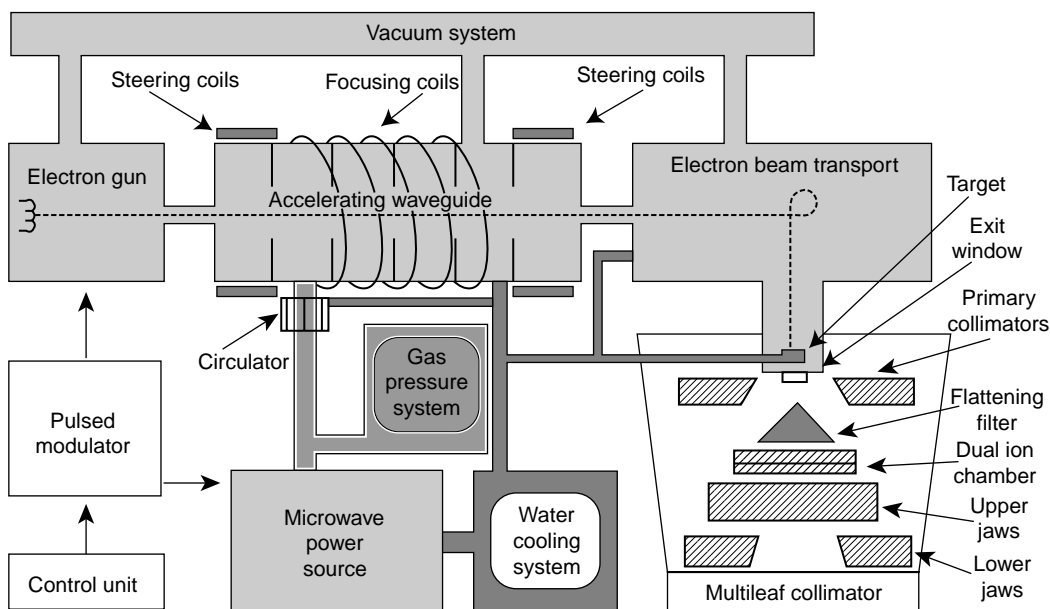


Figure 7. Schematic diagram of an electron LINAC of the type used for radiation therapy (8).

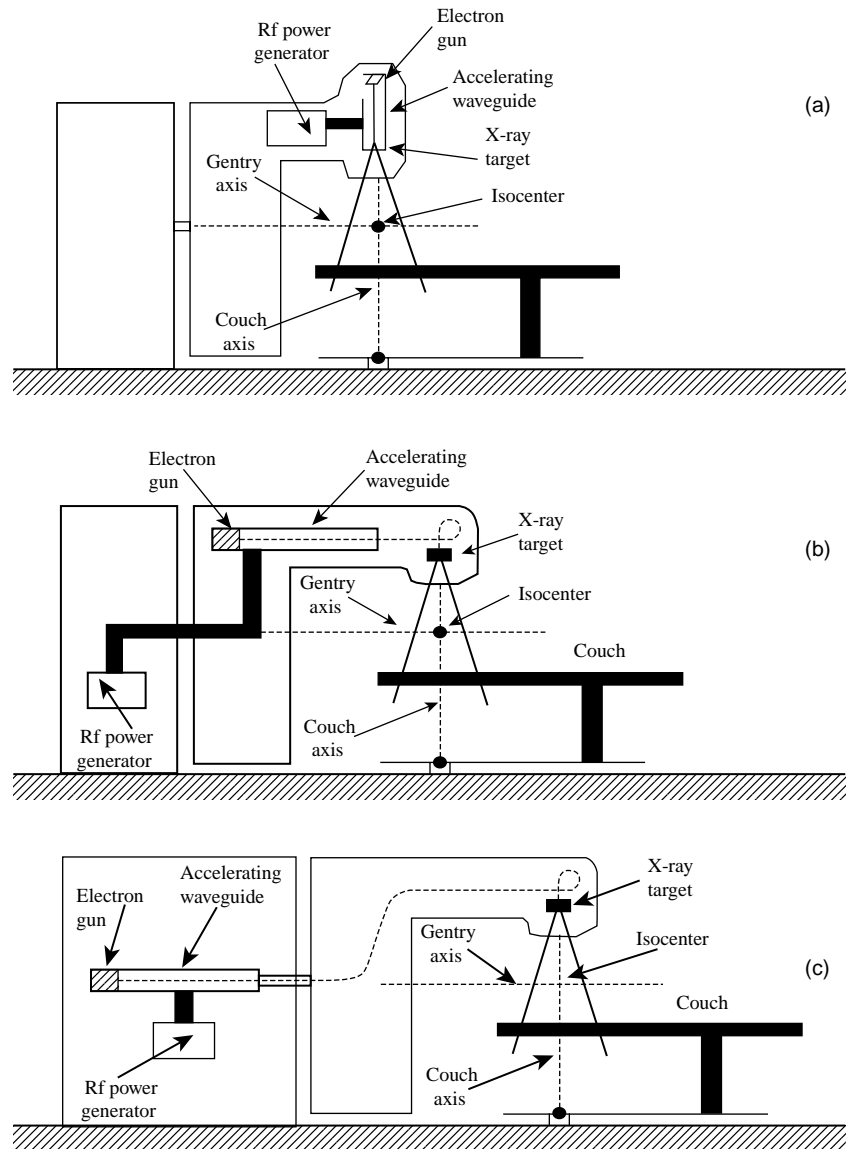


Figure 8. Various isocentric gentry arrangements used with electron accelerators for radiation treatment using secondary X rays or the direct electron beam (8).

facilities use electron beams or X rays with energies greater than 25 MeV. However, it has been indicated by theoretical calculations and by recent experiments with phantoms (21,39–44) that high energy direct electron beams confined by a high magnetic field potentially could be useful for some cancer treatments (see below).

FAST AND SLOW NEUTRON RADIOOTHERAPY

Fast Neutrons

As noted previously, it was realized shortly after their discovery in 1932 by Chadwick that fast neutrons could have advantages over X rays and γ rays in cancer therapy (5–7,14). Unlike protons and α particles, neutrons being massive, yet uncharged, can penetrate more easily into tissue even at modest energies (e.g., 20–40 MeV in kinetic energy). Such neutron energies are readily available via nuclear reactions such as $^9\text{Be}(d, n)$ using conventional cyclotrons, including many of those available during the

1930s (6). Hence, fast neutrons were used in cancer treatment well before high energy protons, α , and other heavy particles became available.

As a neutron therapy beam is produced as a secondary beam from a nuclear reaction on a production target with a primary charged-particle beam, the associated accelerator generally must have a high primary beam current (e.g., a proton or deuteron beam at the μA level). The accelerator facility also must then have the necessary massive shielding for fast neutrons. In some cases, the primary accelerator also can then be used to produce radioisotopes (such as ^{18}F and ^{11}C) that can be used for positron emission tomography (PET). Like other early cancer-therapy accelerators, most of the first-generation fast-neutron treatment facilities were adjuncts to nuclear research facilities. Recently, several new dedicated fast-neutron treatment facilities have become available (6,23,45).

Neutrons in tissue (and other material) behave quite differently than protons, α particles, or other heavy charged particles. Specifically, being uncharged, the dose

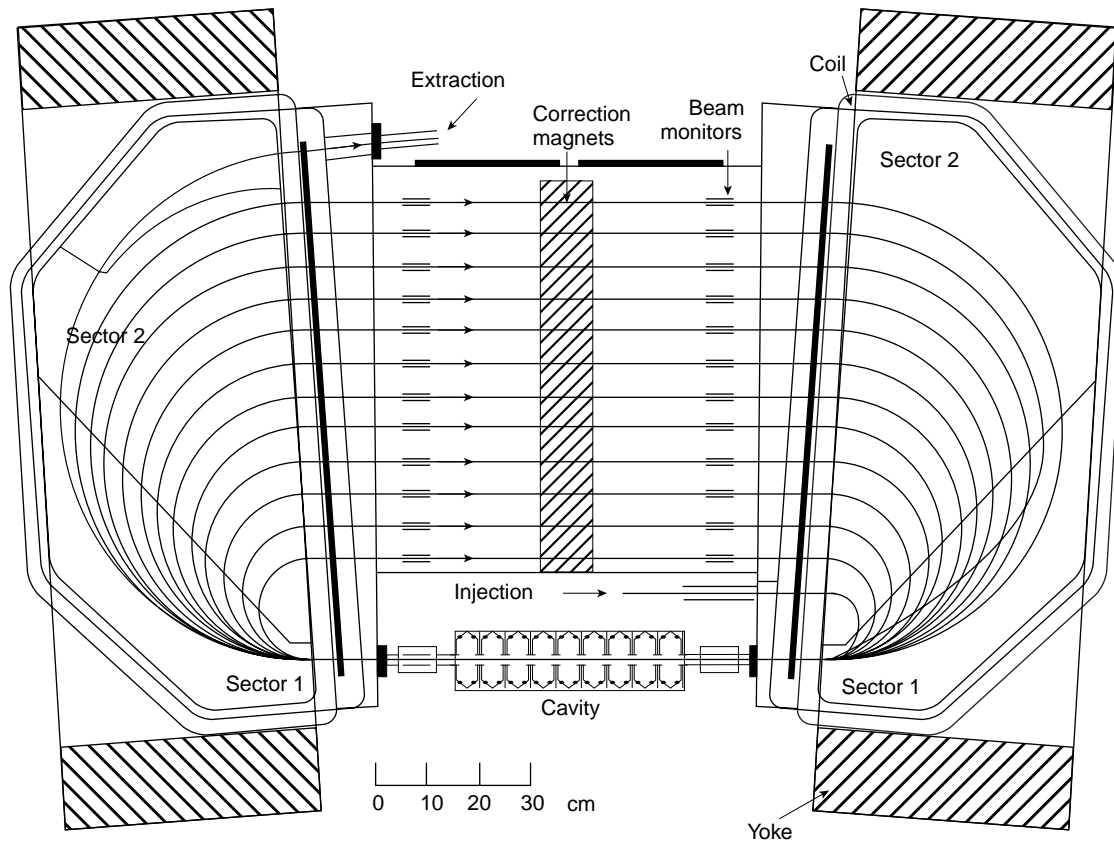


Figure 9. Compact race-track electron microtron used to produce high energy electrons; $E > 25$ MeV (36).

deposited along the path is *not* due to a large number of low LET collisions with atomic electrons. Instead, large-LET nuclear collisions play a dominant role with neutron-proton and neutron-nucleus collisions important in tissue (2,6,7). The latter produce secondary ionization and biological dose due to the recoil protons and other recoiling tissue nuclei. In this sense, recoil protons (and other recoil nuclei) play the role that the secondary electrons play in X-ray and γ -ray therapy. It, then, is not surprising that the deposited dose curve due to fast neutrons looks similar to that from X rays and γ rays (Fig. 2). However, the RBE for neutrons is generally greater than that of the latter e.g., $\times 4$ or larger (2,5–7,23,45).

Like X rays and γ rays, and unlike charged particles, neutrons are attenuated and scattered in tissue and do not have a well-defined range. Treatment of a localized tumor requires either highly fractionated doses or stereotactic treatment to spare healthy tissue, which, together with uncertainties in the RBE for neutrons, was a major problem in early studies of fast-neutron cancer therapy. Many patients were found to suffer long-term complications from the treatment (7), and many of the early neutron treatment facilities ceased operation until the latent effects of the therapy were better understood.

Today, fast-neutron therapy is usually restricted to treatment of special types of tumors (Table 2) where this type of therapy has been shown to be advantageous, yet without a high probability of long-term complications, or where such complications may be justified (e.g., for older

patients), which includes cancers of the salivary glands, prostate cancer, and several types of soft-tissue and inoperable cancers. Neutrons have been shown to be especially advantageous (23,45) in treating radiation-resistant cancer cells.

A recent state-of-the-art dedicated fast-neutron treatment facility is the one located at the Harper-Grace Hospital in Detroit, Michigan (23,45), which uses an innovative gantry-mounted, compact super-conducting cyclotron (45) to produce a high intensity 48 MeV deuteron beam (Fig. 10) (23). This beam then impinges on a beryllium target to produce a range of MeV-energy neutrons for treatment. Having the cyclotron mounted on the treatment gantry minimizes the size of the facility. A special multipin collimator is used to collimate the 2D profile of the treatment beam. Treatments are generally fractionated and done in a special prescribed sequence with X-ray therapy to enhance destruction of radiation-resistant tumor cells that could later metastasize. This type of treatment appears to significantly enhance the long-term survival rate for certain stages of prostate cancer (23,45).

Slow Neutrons and BNCT

As suggested by G. Locher in 1936, one method to increase the localized dose from fast and slow neutrons is to tag tumors with certain elements such as boron, which preferentially capture neutrons (46). Boron, specifically the isotope ^{10}B , which is about 20% of natural boron, has a

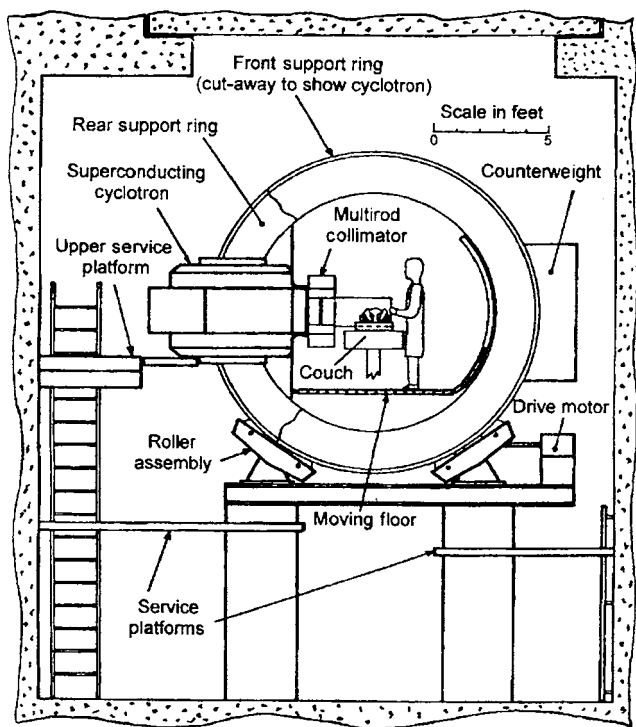


Figure 10. Schematic of the Harper-Grace Hospital (Detroit) gantry-mounted compact super-conducting cyclotron and treatment area used for fast-neutron radiotherapy (23).

high cross section for the reaction $^{10}\text{B}(n, \alpha)$ and produces (Fig. 11) (23) highly localized, low energy α particles (1.47 MeV) and ^7Li recoil ions (0.8 MeV), which quickly stop, yielding a highly localized LET that greatly enhances the local dose to a tumor. The boron is preferentially attached to the tumor site using a tumor-specific boron-loaded pharmaceutical (46,47). As this technique works particularly well with slow neutrons (keV to MeV), it can be used with slow or low energy neutrons produced from small accelerators or from nuclear reactors, which is the basis for boron-neutron capture therapy (BNCT), and a number of clinical trials using BNCT are underway, primarily outside of the United States.

Related to BNCT, the manmade isotope ^{252}Cf , which produces both energetic α particles ($E_\alpha = 6.1$ MeV) together with fission fragments and associated fission neutrons ($E_n \approx 2-3$ MeV), has been proposed (46,47), together with boron-loaded tumor-specific compounds, as a special form of BNCT brachytherapy.

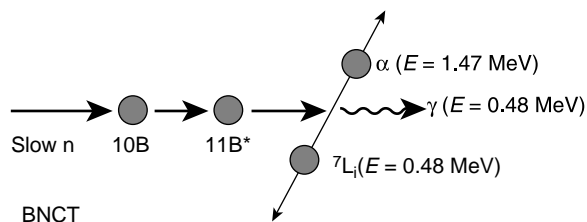


Figure 11. Illustration of the nuclear processes involved in BNCT (adopted from Ref. 23).

Proton-Beam Radiotherapy

Excluding electron beams, which, as noted, have limitations for treatment of deep tumors, protons are presently the primary beams used in particle-beam therapy. A number of cyclotrons, originally operated as nuclear research facilities, have since been converted for use as medical treatment facilities (Table 1). A conventional cyclotron i.e., one that uses a fixed radio frequency (RF) accelerating voltage and a single large conventional magnet (6) is typically limited in proton beam energy to less than 80 MeV, which limits the penetration depth to less than a few cm in tissue (Fig. 2). Hence, these facilities (about one-third of all proton-beam therapy facilities) are generally limited to treatment of shallow tumors, especially eye tumors, or for treatment of certain noncancerous conditions such as age-induced macular degeneration (AMD) (10,12,48,49). As many of these facilities are located in a nuclear research laboratory rather than in or near a hospital, certain treatments requiring fractionated doses or treatment done in combination with other modalities can be problematic.

Treatment of deep-seated tumors requires a proton beam energy of 200 MeV or more (Fig. 2). At this energy, the protons' relativistic increase of mass with increasing velocity requires the use of a large separated-sector cyclotron or high field cyclotron, such as those in use at Indiana University, Massachusetts General Hospital (Fig. 12) (49), and elsewhere, or a "race track" synchrotron adapted from high energy physics. An example of the latter is the synchrotron at the Loma Linda proton-beam treatment facility (49). Synchrotrons are also generally used to produce heavier particles [e.g., ^{12}C ions used for radiation treatment (see below)].

A cyclotron produces a beam with a 100% macroscopic duty cycle, although it still has a beam modulated by the accelerating voltage RF, typically tens of MHz (6). The synchrotron produces a beam modulated by the ramping time of the variable-field accelerator magnets. The latter can be on the order of seconds (6), which generally is not a problem in radiation therapy and can be used as an advantage in some treatments. A third type of accelerator, the synchro-cyclotron, developed after WWII at LBL and used at LBL for α -beam radiotherapy for many years, is presently only in limited use (Table 1).

As the direct proton beam in these accelerators is used for therapy, only a modest beam intensity is required relative to the beam intensities needed for a nuclear research accelerator or one used to produce neutrons or pions, *viz.* namps vs μ amps, which can simplify the accelerator design, the building, and shielding required. However, in some cases, a high intensity beam is desirable in order to produce radioisotopes used in PET and other imaging methods.

All types of proton facilities, owing to the high "magnetic-rigidity" of high energy protons, require a set of large (and costly) beam-switching magnets and patient-treatment gantries (Fig. 13) (49). Likewise, raster-scanning the beam and varying the dose-depth required to treat a particular tumor is not trivial, which can be done (Fig. 14) (49) with electronic elements (active scanning) or with shaped absorbers (passive modulation).

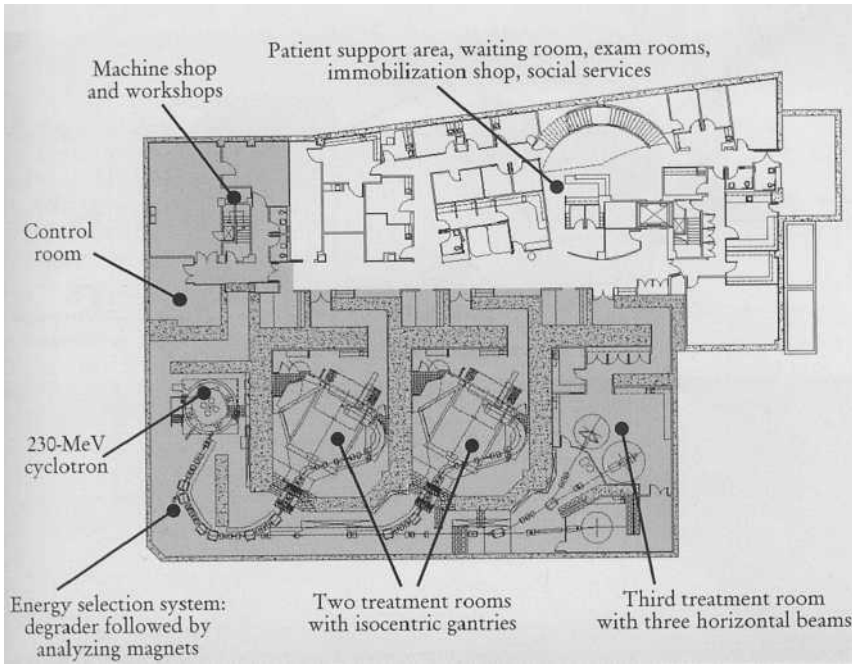


Figure 12. High field spiral-ridge 230 MeV proton cyclotron facility and treatment room layout of the Massachusetts General Hospital Northeast regional proton-beam radiotherapy facility. (Reprinted with permission from Ref. 49, copyright 2002 American Physical Society.)

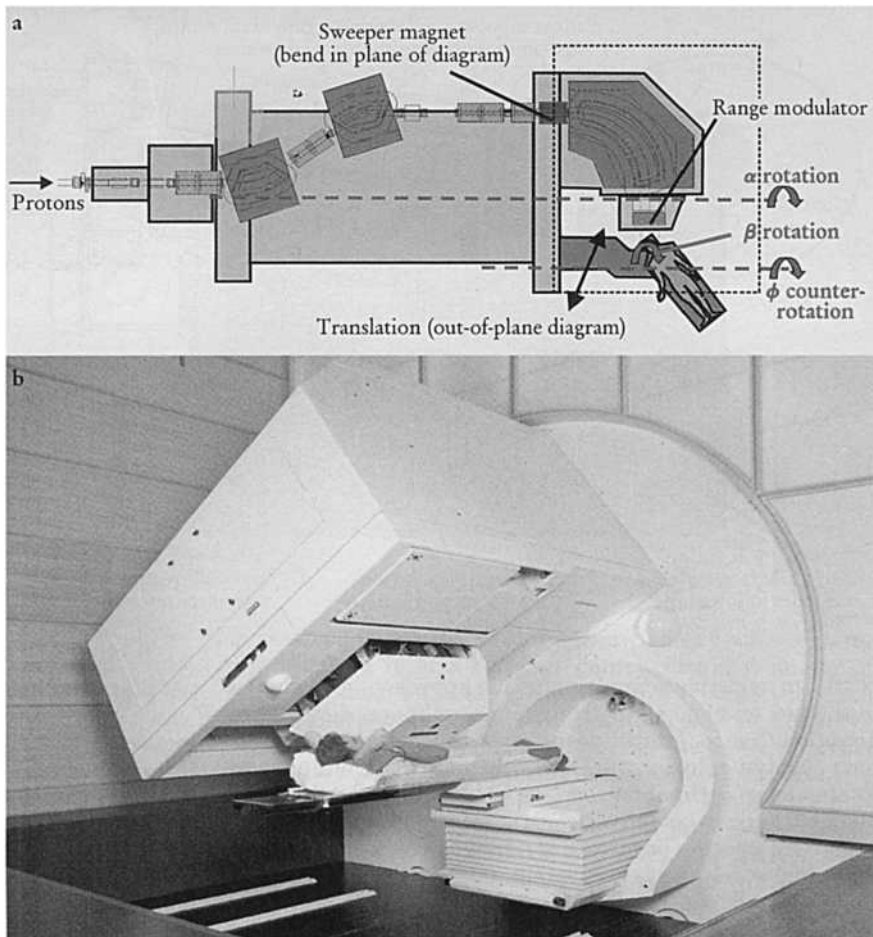


Figure 13. The isocentric proton-beam gantry used at the PSI (Switzerland) proton-beam cancer treatment facility. (Reprinted with permission from Ref. 49, copyright 2002 American Physical Society.)

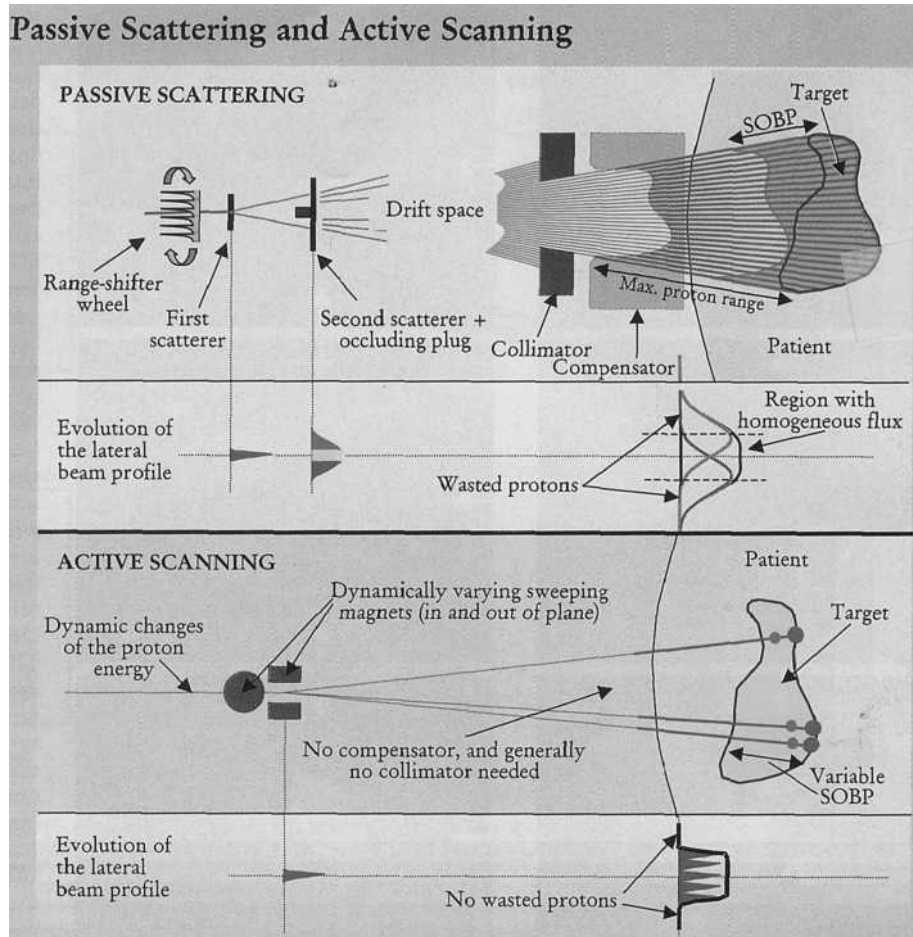


Figure 14. Schematics of typical passive and active ion-beam raster-scanning apparatus used to generate suitable dose profiles for treatment of specific tumors. (Reprinted with permission from Ref. 49, copyright 2002 American Physical Society.)

Although the primary energy loss mechanism and, hence, dose from high energy protons is still due to collisions with atomic electrons (2), a significant fraction of the incident beam can induce nuclear reactions in beam collimators and in the patient, which, in the patient, can lead to the production of neutrons or other reaction products complicating the calculation of the biological dose. Some of these reactions can be detected by observing the annihilation radiation (i.e., emission of back-to-back 511 keV γ rays) from proton-rich positron-emitting nuclear reaction products (9,10,50). This technique (based on PET) has been adapted to image the dose profile of heavy-ion beams used in radio therapy (see below) where the beam itself may fragment (10,11,15,16) into positron-emitting nuclei.

Related to this technique, collimating an energetic proton beam and stopping it (including in a patient) can produce copious amounts of fast neutrons. Although the direct internal dose to the patient due to the neutrons may be small, the facility itself may require extensive neutron shielding for protection of the workers and surrounding environment (Fig. 12), which results in a very large “foot-print” for such a facility and a building cost that often exceeds the cost of the accelerator and beam line themselves. As a result, only a few hospital-based proton-therapy facilities presently exist. These facilities are generally

dedicated to the treatment of certain tumors (Table 2) and for conditions that are otherwise inoperable or difficult to treat with conventional radiotherapy. However, it is estimated that, in the United States alone, over 10,000 patients/year could benefit from proton-beam radiotherapy treatment (12).

As the advantages of proton therapy becomes better documented for certain types of tumors, more medical insurance companies will likely approve treatment, which would justify the construction of more facilities. Nonetheless, the number of patients treated worldwide with protons has increased steadily, with over 40,000 treated through 2004 (Table 1).

Heavy-Ion Radiotherapy

Excluding antimatter particles, the particle beams with the highest LET and, hence, dose rate per unit path length (i.e., dose-depth profile) are heavy ions (HI) with $z \geq 2$. (Eq. 1), which leads to an extremely sharp Bragg peak and, hence, localization of the dose near the end of the HIs’ range (2,10,11,14). Also, the biological dose is enhanced over the HI LET alone owing to the large values of RBE determined for HIs. The latter can be on the order of 10 or more near the end of the HIs’ range (15).

As with proton therapy, in a typical treatment, the sharp Bragg peak in the dose curve is spread out using absorbers (or other means) to provide a suitable overlap with the tumor (SOBP) (9–11,14). Like protons, energetic HIs, like those needed for therapy (typically a few hundred MeV/nucleon), can also induce nuclear reactions. In the case of HI therapy beams, these reactions, which are primarily beam fragmentation, can transmute the incident beam into other nuclear species (9–11,14–16), which usually include lighter, lower z fragments, which can extend the radiation dose well beyond the range of the primary HI beam (16). Fortunately, many of the fragments and residual nuclei produced in HI (or, as noted, proton) beam-induced nuclear reactions are positron emitters and their intensity and location can be imaged via the back-to-back 511 keV γ rays emitted (PET) (9,10,50–52), which is now being actively exploited in treatment planning at GSI and elsewhere (see below).

It also has been suggested to specifically produce a short-lived positron-emitting secondary beam such as ^9C for radiation treatment and, thus, facilitate direct imaging of the treatment beam itself (51), however, while feasible, it is not yet a practical option. Among other problems, such beams are easily fragmented, and as a secondary beam, production requires a high intensity primary HI beam accelerator. Instead, like proton therapy, a direct, low intensity HI beam is more practical for therapy (53–55).

Excluding the use of natural α -emitting radioactive sources, the use of HIs in cancer therapy was pioneered at the University of California, Berkeley, laboratory now known as the Lawrence Berkeley Laboratory (LBL). The Bevatron at LBL, which originally was constructed to discover the antiproton, was converted to accelerate HIs such as ^{12}C , ^{16}O , and ^{20}Ne at energies up to several hundred MeV/nucleon (9–11,14). These energies are those required for cancer treatment of deep tumors. Over 400 patients were treated from 1975 to 1992 at LBL with HIs.

As part of this program, many techniques were developed for controlling and monitoring the HI dose-depth profiles, and providing a suitable SOBP when needed, which included imaging positron-emitting HI beam fragments (9–11,14) and reaction products, which, as noted, is a technique later adapted for proton and HI therapy at other facilities.

More recently, the HI nuclear research facility at Darmstadt, Germany (GSI) has run a HI prototype cancer therapy facility primarily using ^{12}C beams at energies of several hundred MeV/nucleon. They have implemented the online PET imaging technique to deduce dose-depth profiles (52) for clinical treatment planning and verification (Fig. 15) (52). They, likewise, have done extensive measurements and modeling to determine the RBE appropriate for HIs in tissue (15), which, as noted previously, can be relatively large ($\times 10$ or more) and thus must be included in treatment planning (51–56). (The latter also has implications for space travel and other activities involving radiation from heavy particles in cosmic rays, etc.).

Many of the tumors treated at the GSI facility cannot be optimally treated with conventional radiotherapy or surgery due to the close proximity of the tumor to a critical area (Fig. 15). As with other particle-beam therapies, most

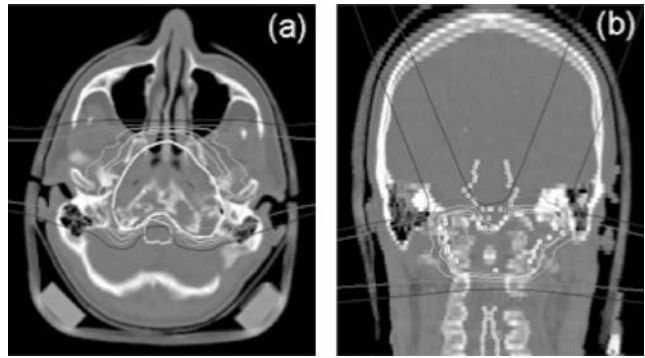


Figure 15. Example of a tumor treatment plan used for HI radiotherapy at GSI (52).

HI treatments involve fractionated doses, and the HI therapy is often used in conjunction with other treatment modalities such as chemotherapy or photon therapy. Like other pioneering particle-beam facilities, GSI is a nuclear research facility and not *a priori* part of a hospital facility, which can often limit patient throughput. Nonetheless, several hundred patients have been treated since 1997 (Table 1). Based on the results demonstrated, the GSI nuclear and biomedical research group together with a hospital facility in Heidelberg, Germany is finishing the constructing of a dedicated HI accelerator and HI treatment facility. This facility is expected to become fully operational in mid-2005 and will serve as a German national treatment facility for the 3000 or more patients identified as optimal for HI therapy each year in Germany. A similar facility is under construction in Italy (12).

At present (2005), the primary facilities built and dedicated to HI cancer therapy are the HIBMC and HIMAC facilities (53) in Japan (Fig. 16) (54). As at GSI, these facilities primarily use ^{12}C ions at energies of several hundred MeV/nucleon. (HIBMC also has the capability to use protons in radiotherapy). Several thousand patients have been treated at HIBMC and HIMAC (Table 1). Like proton therapy facilities, the large footprint (Fig. 16) and costs associated with an HI accelerator and the associated treatment facilities will generally limit their availability. In countries with national health services, one, two, or at most three HI facilities may accommodate those patients who might benefit from HI therapy. The situation becomes complicated in countries like the United States where private insurers must approve treatment.

FUTURE DEVELOPMENTS

Magnetically-Confined Electron Beams

It has been suggested, and recently demonstrated with experiments, that one might use high energy electrons ($E \leq 100$ MeV) for radiotherapy with suitable magnetic fields applied to reduce (Fig. 17)(21) the penumbra from scattering (21, 31–44). Again, if the direct electron beam is used, relatively low intensity beams can be used, which simplifies the accelerator, beam handling, and shielding. Most hospitals operate and maintain electron accelerators ($E = 10$ –25 MeV), mostly LINACs, for radiation therapy

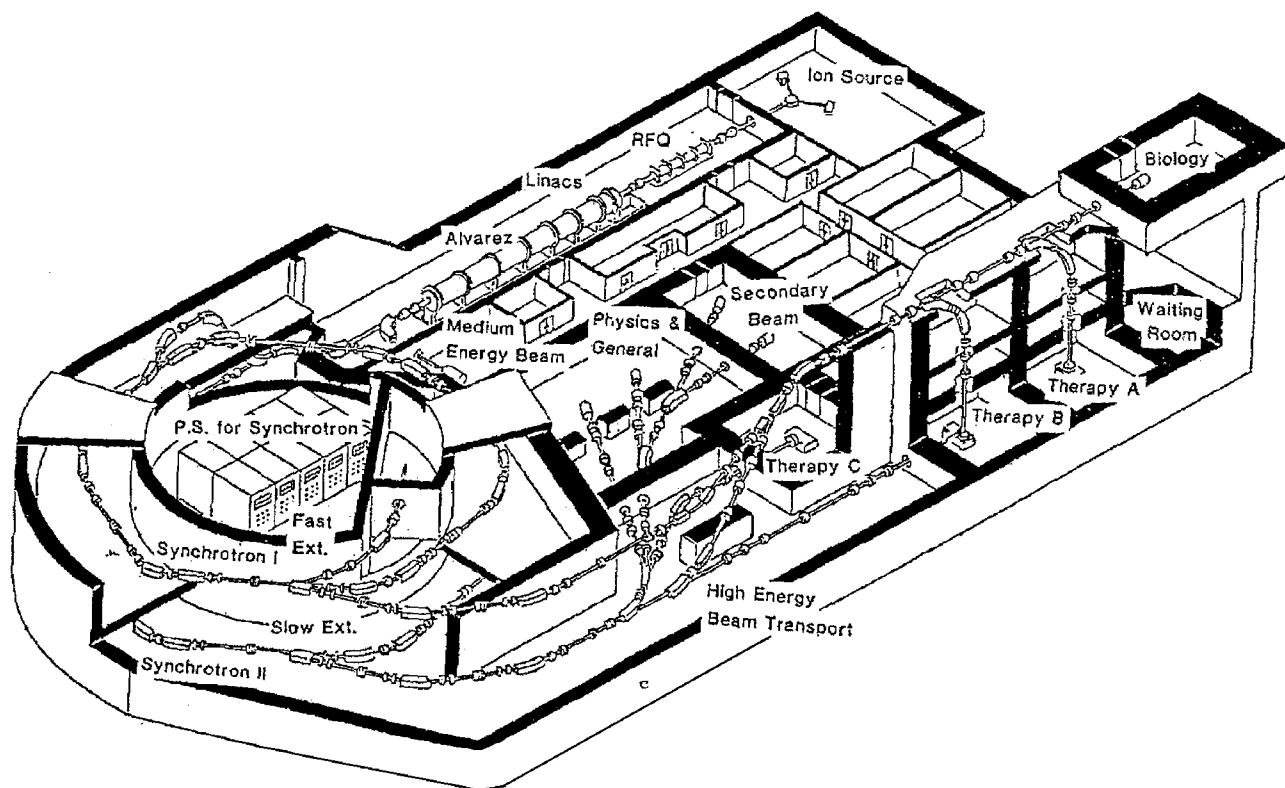


Figure 16. Layout of the HI-beam cancer treatment facility HIMAC in Chiba, Japan (adopted from Ref. 54).

and higher energy electrons (50–100 MeV) are readily produced using expanded versions of LINACS (6) or tabletop race-track microtrons (36). Both are quite feasible for operation at most hospital clinical oncology facilities.

Recent developments in super-conducting magnet technology including gantry-mounted systems and LHe-free systems make such magnets technically feasible. These magnets could be used in conjunction with a suitable electron accelerator for cancer therapy for certain types of tumors (soft tissue, etc.). A sample of a calculated multibeam dose profile in a skull-tissue phantom using

35 MeV electrons confined with an axial (solenoidal) $B = 6$ T magnetic field (44) is shown in Fig. 18.

Superconducting Accelerators and Beam Gantries

One method to possibly reduce the footprint (and cost) of proton and HI radiotherapy facilities is to use, more widely, superconducting magnet technology for both the accelerator and beam line components (24). However, extensive radiation shielding may still be required.

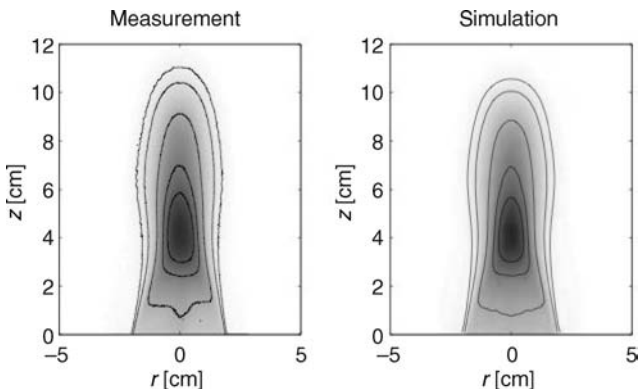


Figure 17. Measured and calculated 2D isodose profiles for a ca. 20 MeV electron beam incident on a tissue-equivalent phantom with an applied 2 T longitudinal magnetic field. Compare with Fig. 4 (adopted from Ref. 21).

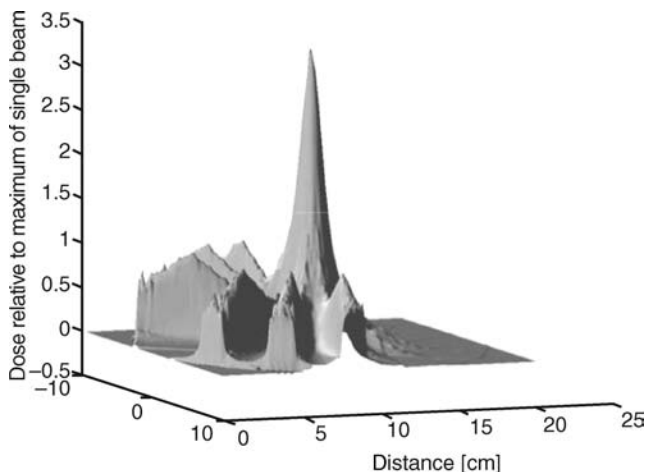


Figure 18. A simulated stereotatic dose profile in a head phantom for a set of 35 MeV electron beams confined by a $B = 6$ T longitudinal magnetic field (44).

Pulsed-Laser Accelerators

It is now possible to produce very high electric fields in plasmas using compact ultra-fast pulsed lasers, which can be used to accelerate electrons, protons, and other ions to MeV energies (38). Although the particle-beam intensities and, in some cases, energies demonstrated so far are less than those needed for radiotherapy, such “table-top” accelerators may prove viable in the future.

ACKNOWLEDGMENTS

The author thanks J. Sisterson, Ph.D., Yu Chen, Ph.D., D. Litzenberg, Ph.D., Prof. L. Jones, and Hao Jiang for their assistance.

References are on page 609.

See also IONIZING RADIATION, BIOLOGICAL EFFECTS OF; RADIATION DOSIMETRY FOR ONCOLOGY; RADIOTHERAPY TREATMENT PLANNING, OPTIMIZATION OF.

RADIOTHERAPY, INTRAOPERATIVE

PETER J. BIGGS
Harvard Medical School
Boston, Massachusetts

INTRODUCTION

What Is Intraoperative Radiotherapy?

Intraoperative radiotherapy (IORT) is a technique that combines radiation therapy with surgery to irradiate tumors *in situ*, without delivering a significant dose to surrounding normal, critical structures and is usually performed using electron beams from linear accelerators. This contrasts with external beam therapy using X-ray beams where the dose that can safely be delivered to most tumors is limited by the dose that is given consequentially to normal, critical structures.

The practice of IORT began soon after X rays were used therapeutically, almost 100 years ago. However, although several investigators used this technique over the ensuing years with various X-ray modalities, it was not until the mid-1960s that intraoperative radiotherapy made a serious mark on the field of radiotherapy with the work of Abe (1–4) in Japan using electron beams from linear accelerators. This was rapidly followed in the mid-1970s by investigations in the United States, first at Howard University (5), then at the Massachusetts General Hospital (MGH) (6), followed by the National Cancer Institute (7), and then the Mayo Clinic (8). The reason for this dramatic change was due to the introduction of linear accelerators capable of generating high energy electron beams.

In 1992, Coia and Hanks (9) reported on patterns of care study, which indicated that of 1293 radiation oncology facilities in the United States, 108 reported doing IORT, of which 29 have two or more residents. Since there were ~88 training programs in existence at that time, roughly one-third of hospitals or medical centers with residency training programs were performing IORT. They did not

indicate whether or not this list included only electron beam IORT or other modalities.

Initially, IORT flourished in both the academic and community hospital setting, but it is clear from informal surveys and anecdotal evidence that fewer centers are now performing IORT compared with 1992. The reasons for this decline in interest are twofold. First, establishing the usefulness of IORT as a beneficial adjunctive therapy has been difficult. Second, IORT as practiced by the majority of centers, those that do not have the luxury of a dedicated or conventional mobile linear accelerator in the operating room (OR), is technically difficult and demands time on the part of a clinical staff that is under great time constraints, brought on by the present reimbursement climate. Thus, this method taxes the interests of all the parties after a number of years. The uphill battle faced by proponents of intraoperative radiation therapy is the high cost of a dedicated facility in the operating room. A dedicated linear accelerator in an operating room is no longer a cost-effective option for any hospital (10), due to the cost of the machine as well as the radiation shielding. The entry into the field of IORT of mobile linear accelerators that can be used in existing OR rooms without requiring additional shielding makes the cost and logistics of setting up an IORT program much easier and therefore provides a stimulus to the field. There are now three manufacturers of such equipment and >30 units have been installed in the United States and Europe.

In addition to using electron beams from linear accelerators, many other radiotherapy modalities can be classified as IORT. They all share the same principle that the dose is delivered only locally, so that dose to the skin uninvolved adjacent tissues and organs is minimized. These include high dose rate brachytherapy that is delivered in an operative setting and stereotactic radiosurgery using a 50 kV X-ray device. The Papillon technique (11) was a technique whereby 50 kVp X rays were used to irradiate lesions on the rectal wall by inserting the X-ray tube into the dilated rectum. Orthovoltage X rays are still used in one or two places (12), based on the issue of lower cost, but there is a clinical price to pay since the dose to any bone in the field is much greater than the prescribed dose to tissue, and this can result in osteoradionecrosis. However, this can be obviated to some extent by heavy filtration of the beam at the cost of a lower dose rate.

General Description of the Treatment Apparatus

Historically, adapters were made to fit existing linear accelerators so that the electron beam could be directed onto the tumor while at the same time protecting normal tissue. This was first achieved by having an applicator, ~30 cm long and generally circular in cross-section that mates with another cylinder attached to the head of the linear accelerator. This mating piece is centered with the radiation beam and only slightly larger than the treatment applicator. Thus, by adjusting couch movements very carefully, the treatment applicator can be slid into the mating piece, even when the gantry is angled far from the vertical. The greater the gap between the two pieces, the easier it is to achieve docking, but the greater the possible degree of

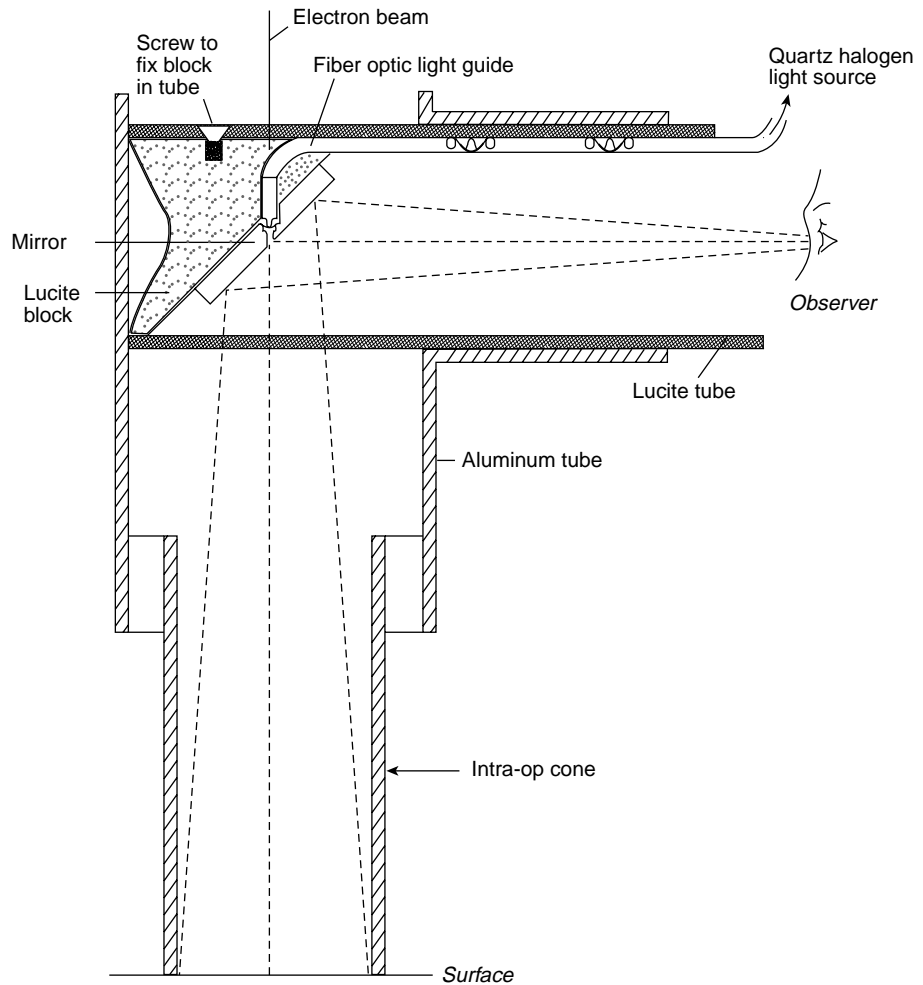


Figure 1. Diagram of optical system used for a hard dock system. Note that the unit slides in and out of the field and that the reflector is metallic.

misalignment, meaning the beam is not aligned with the applicator in the patient, so the beam, as delivered, will have different dose characteristics from a perfectly aligned system. Typically, the gap between the two cylindrical pieces is $\sim 0.2\text{--}0.3$ mm. This method is known as the hard dock technique (13). The advantage of this method is that the system is self-aligning, within limits defined by the gap between the two mating pieces, but the disadvantages are that (1) the docking process can be lengthy, depending on the gantry angle, (2) there is a potential safety issue if either the couch or the gantry moves during the procedure, and (3) the treatment field is no longer visible once the docking process is complete. Hence, the hard dock procedure requires a lock-out system for the drive motors during the treatment when there is no one in the treatment room and an optical system to view the field enclosed by the applicator after the applicator has been docked. A cut-away view of such an optical system is shown in Fig. 1. A photograph of a completed hard dock process is shown in Fig. 2. This photo shows that the docking or alignment process can be a complex and time-consuming task in some situations.

The alternative method that avoids the problem of the hard dock procedure is the soft dock procedure. In this method, the treatment applicator is separated from the head of the machine and thus the potential for patient injury is greatly lessened and the treatment field can be



Figure 2. Example of electron beam alignment using the hard docking process. The linear accelerator is a conventional machine in the therapy department and the applicator was fabricated in-house.

viewed even after the docking is complete. The disadvantage of this method is that an alternative to the simple mechanical alignment is needed. This need has been answered by a variety of optical systems as explained below. A description of two commercial alignment systems is given below. A description of several noncommercial soft-docking systems has been published by several authors (14–17).

The Clinical Rationale for IORT

To understand the rationale for intraoperative radiotherapy, it is necessary to understand some of the basic principles of radiotherapy. Patients are routinely treated for cancer using fractionated radiation. This means that the radiation is delivered in increments on a daily basis, 5 days week⁻¹ for up to 8 weeks, depending on the lesion under treatment. The reason for doing this is that if one were to deliver a tumoricidal amount of radiation to the tumor in one or a few fractions, serious long-term side effects to normal tissue and organs would result. By fractionating the radiation, an equivalent tumoricidal dose can be given to the tumor without serious long-term side effects. Fractionation schemes currently in use have developed empirically over the history of radiotherapy going back to the use of early X-ray tubes. The daily fraction dose is limited by acute radiation effects. These effects, such as reddening of the skin in the era of low energy X rays, bowel problems, and so on, appear during the course of treatment, but will generally resolve themselves without long-term consequences after the radiation treatment has been completed. This maximum radiation dose may produce acute effects in a few patients, due to biological variation between individuals and, hence, their response to radiation. The prescription dose or the maximum overall dose, which is the number of daily fractions times the daily dose, is also limited. This limit is due to normal tissue and organ tolerance. Thus, for example, abdominal radiation for rectal or colon cancer is limited to ~50 Gy because of small bowel complications (Gy is the unit of absorbed dose and is expressed in J·kg⁻¹). The kidneys can tolerate a dose of no more than ~15 Gy and treatment for lung cancer requires keeping the dose to the spinal cord below ~45 Gy to avoid transverse myelitis. Some organs, such as the liver, can tolerate varying amounts of radiation, depending on the fraction of liver that is irradiated. The smaller the fraction of organ treated, the larger the dose that can be tolerated. Excess dose to the whole lung can produce fibrosis, resulting in a nonfunctional lung. However, even with these dose limits, local control rates, or control of the primary tumor that is being irradiated is not 100%, so it would be highly desirable to have a method for increasing the dose to the tumor without increasing the dose to the surrounding normal tissue. This is known as improving the therapeutic ratio, which, for a given dose fractionation, is defined as

$$\text{Therapeutic ratio} = \frac{\text{tumor control probability/normal tissue complication probability}}{\text{normal tissue complication probability}}$$

Ideally, this ratio should be as high as possible. One method to improve this ratio is through the use of intraoperative irradiation. In this technique, the area to be

treated, whether a tumor or tumor bed, is exposed during surgery and normal tissue (e.g., small bowel) is moved out of the way by using an applicator through which the radiation is delivered. By irradiating with electrons rather than photons (see section below for comparison between therapeutic photon and electron beams), the radiation can be safely and effectively limited to the area of the tumor. It has been firmly established using experiments on dogs (18,19) that a safe upper limit for single fraction doses is ~20 Gy. Doses delivered through IORT generally range between 7.5 and 20 Gy.

While IORT can increase the therapeutic ratio by excluding normal tissue and organs from the radiation field, it suffers from the fact that it is a single fraction procedure, which, as noted above is limited and therefore, by itself may not provide a curative dose of radiation for all tumors. The reason for this is that such processes as repair of sublethal damage, repopulation, redistribution, and reoxygenation (20), which contribute to enhancing the therapeutic ratio for fractionated radiation, are limited in single-dose therapy. While the initial Japanese study used IORT as the only source of treatment radiation, it has been customary in the United States to use IORT as a boost therapy. This means that the patient is treated using the standard fractionation scheme using external radiation and is given the IORT as an additional boost dose. Thus the tumor dose has been increased with only a small increase in dose to a fraction of the surrounding tissue. Table 1 (21) shows the radiobiological equivalent fractionated dose of single doses of radiation between 10 and 25 Gy. Assuming a conventional fractionation scheme for external therapy of 2 Gy per fraction, a single dose of 10 Gy is equivalent to 17 Gy for tumors. For late effects in normal tissue, this figure is much higher at 26 Gy, but is considerably lower if the dose per fraction is only 50%. Thus normal tissue toxicity determines the maximum dose that can be delivered intraoperatively.

Historical Review of IORT

The first cases of IORT were reported by Beck as far back as 1907 (22) and again in 1909 (23). He used X rays (probably 50 kVp, see below) to treat several cases of stomach and colon cancer. Several years later, Finsterer, in 1915 (24), reported on the treatment of stomach and colon cancer. He used doses between 2500 and 4500 R using X rays with filtration that varied between none, variable thicknesses of

Table 1. Equivalent Dose of a Single Dose of Radiation in Terms of the Fractionated Dose for Both Acute and Late Normal Tissue Reactions

IORT Single Dose, Gy	Equivalent Dose for Tumor and Acute Normal Tissue Reactions ^a	Equivalent Dose for Late Normal Tissue Reactions ^a
10	17	26
15	31	54
20	50	92
25	73	140

^a2 Gy fractions.

aluminum, or 0.25–0.45 mm Cu. The “R” stands for the roentgen, which is a unit of radiation exposure; 1 Gy is roughly equivalent to 87 R in air. Note that no external radiation was given to these patients. The choice of filtration was dictated by the thickness of the lesion. It was noted by Abe (25), in his historical review of the topic, that the practice of IORT then was different from that of today; however, it is different only in the sense that kilovoltage radiation is exponentially attenuated and delivers its maximum dose at the surface, whereas electron beams have a maximum dose below the surface and the dose falls sharply beyond a given depth, depending on energy. Thus, the intent then was the same as it is now, namely, to give additional dose to the tumor and spare normal tissues, even though there were more practical problems in dose delivery and differences in beam quality. However, there are photon beam modalities used in IORT that get around the problems mentioned above with X rays. High dose rate brachytherapy using radioactive ^{192}Ir sources, for example, is able to deliver a high tumor dose with a low dose to nearby critical structures.

Barth (26) gives a long account of the technique he used for many anatomical sites where he opened the skin and treated the underlying tumor with 50 kVp X rays at 2 cm SSD for very small field (~2.5 cm diameter) sizes. Many of these sites were in the head and neck region, just below the skin; more deep-seated tumors would have been hard to treat with this technique. Interestingly, Goin and Hoffman (27) describe instances, where IORT was delivered on more than one occasion. He reports on 13 patients, all but 1 of whom received IORT from 2 to 12 times, with overall doses ranging from 500 to 30,672 R. Thus surprisingly, fractionated IORT was practiced then, something that would not be countenanced today.

Current Status of IORT Application (User Surveys)

A survey of current institutions practicing IORT in the United States (28), whether using electron beams, orthovoltage X rays, the Photon RadioSurgery System (PRS) device or by High Dose Rate (HDR) was conducted in 2003. It was found that the number of institutions practicing IORT by each technique is shown in Table 2.

It can be seen that ~55% of the institutions perform IORT in the OR and 75% perform IORT with a dedicated unit. One of the centers using orthovoltage X rays and another using a dedicated linac in a shielded OR are converting to mobile linear accelerators. Approximately

Table 2. Number of Institutions in the United States Practicing the Various IORT Modalities (2003)

Type of IORT Practiced	Number of Institutions
Mobile linear accelerator	6 (15.8%)
Dedicated linac in shielded OR	4 (10.5%)
OR in radiotherapy department	8 (21.1%)
IORT by patient transport	9 (23.7%)
Orthovoltage X-rays	2 (5.3%)
Intrabeam (50 kV X ray)	5 (13.2)
HDR	4 (10.5)

72% of the institutions above responded to the survey, including all six with mobile units and all four dedicated units in the OR. The remaining eight responders consisted of six sites with an OR in the therapy department and two that used patient transport. By comparison with the 1992 survey, it was found that the greatest decline in the number of centers involved with IORT was the group performing IORT by patient transport, and these were primarily community centers. Thus a higher proportion of those sites still practicing IORT are academic centers. Of the institutions using nonmobile units, the average date inception of the program was 1986 (± 4 years), whereas all the mobile units were installed after 1998. The most commonly used energy was 9 MeV, followed by 12 MeV. This certainly justifies the choice of maximum energy of the mobile units. Slightly more institutions use the soft-docking (61%) rather than the hard-docking technique. The most commonly used field size is a 7 cm diameter applicator, followed by a 6 cm diameter applicator. Finally, the average number of IORT treatments performed by the reporting centers (72% of institutions responded) is 500, whereas the number of treatments performed in the last 12 months is 428. The respective numbers per institution are 31.2 ± 23.5 (range, 10–90) and 26.8 ± 22.7 (range, 0–70).

A similar survey was recently carried (29) out for European institutions and there are some similarities and differences. Table 3 shows the number in institutions performing each type of IORT.

The chief difference is that there are 40% more sites in Europe than in the United States and that more than one-half the European sites have mobile linear accelerators compared with 16% in the United States. Moreover, >60% of the treatments carried out are for breast cancer compared with almost no cases in the United States. Also, the average number of patients treated is in excess of 1000, compared with 500 in the United States, noted above. Of great importance is the fact that there are a number of clinical trials underway in IORT in Europe, whereas there are no trials in progress in the United States.

What the two continents have in common is the typical applicator sizes and energies. This result is what one would expect if each were treating the same distribution of disease sites. However, since those distributions are not the same, this commonality is quite remarkable. Finally, one other similarity is that a large percentage of institutions are performing a few cases and a few institutions

Table 3. Number of Institutions in Europe Practicing the Various IORT Modalities (2005)

Type of IORT Practiced	Number of Institutions
Linear accelerator in radiotherapy department	11 (31.6%)
Dedicated linac in shielded OR	5 (13.2%)
Mobile linear accelerator	21 (55.3)
Intraoperative interstitial brachytherapy	6 (15.8%)
Intraoperative HDR flaps	7 (18.4%)
Intrabeam (50 kV X ray)	2 ^a (5.3%)

^aThis figure may be substantially higher.

(primarily academic centers in the case of the United States) are performing a large number of cases.

IORT TECHNOLOGY

Early Technology

Radiotherapy prior to the 1960s used primarily X-ray machines for the treatment of cancer patients, although in the years just before that Cobalt-60 teletherapy units were coming into widespread use. Schultz (30) describes in great detail the history of the development of X-ray machines of increasing energy. Thus, in the early days, X-ray machines were the only modality to carry out IORT. In the earliest applications of IORT, Practitioners used 50 kV X rays with a focus-to-skin (FSD) distance of ~ 2 cm with a field size (diameter) of about the same dimension, varying the filtration to achieve sufficient penetration. Henschke and Henschke (31) provide considerable detail on the practical application of this technique to the treatment of large fields. They show that this can be accomplished either by increasing the FSD, primarily to increase the percent depth dose, or by use of multiple, overlapping fields at short FSD. Thus, until X-ray machines operating at higher energies with adequate dose rates at larger distances and covering larger fields became available, the role of IORT was relatively limited. Beginning in the late 1930s through the 1940s, such high energy units became available. Eloesser (32) described the use of an X-ray machine with filtrations between 0.25 and 0.5 mm Cu. He notes that the filtration depends on the thickness of the tumor being treated; the FSD was 30 cm. In 1947, Fairchild and Shorter (33) describe a technique using 250 kV X rays with a half value layer (HVL) of 1.7 mm Cu. At a FSD of 21.7 cm, a dose rate of $100 \text{ R} \cdot \text{min}^{-1}$ could be delivered over a 13 cm field diameter. Note that since these high energy X-ray beams were not collimated, lead sheets had to be placed around the surgical opening and internal viscera to provide adequate radiation protection.

Whereas the difficulty of using low energy X-ray beams is one of insufficient penetration, the problem with high energy X-ray beams is that, although their intensity is exponentially attenuated, tissues or organs beneath the tumor to be treated can receive a considerable dose of radiation. For this reason and many technical reasons associated with using an X-ray machine in an OR, it is clear that IORT could not have been anything other than an experimental technique pursued by a few investigators.

Recent IORT Technology: Photons versus Electrons

In the 1960s, the first linear accelerators appeared in clinical use for the treatment of cancer. One of the first units in the United States was a 6 MV machine (Varian Medical Associates, Palo Alto, CA), which produced a bremsstrahlung X-ray beam from a beam of 6 MeV electrons. However, it was not until the 1970s that linear accelerators were capable of producing clinical electron beams. Omitted from this historical review is a discussion of the betatron. Designed and built in the early 1940s by Donald Kerst, the first patient was treated with X rays in

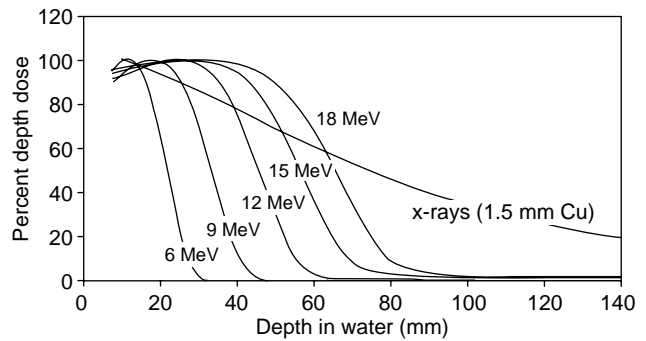


Figure 3. Comparison between electron (6, 9, 12, 15, and 18 MeV) and X ray (1.5 mm Cu X rays) percent depth dose curves.

1948; electron beams were a natural by product of this machine, although the dose rates were quite low. However, the betatron played no role in the field of IORT, and there are no betatrons currently in clinical use in the United States. Clinical electron beams are produced by passing the electron beam exiting the linear accelerator's waveguide through high atomic number scatterers (34). The great advantage of electron over X-ray beams is in the depth dose curve. Figure 3 shows a comparison of the percent depth dose curves, normalized to 100%, between electron beams of various energies and a 1.5 mm Cu X-ray beam. What is immediately apparent is that while the curve for the X-ray beams fall off more or less exponentially with distance beyond the depth of maximum dose, the curves for electron beams show a very sharply falling dose and have a finite range. This range is a function of the electron beam energy, so the higher the energy the greater the range. A rule of thumb is that for every 3 MeV of electron energy, the depth of the 80% dose changes by 1 cm. There are currently only one or two active IORT programs in the United States using orthovoltage X rays.

Dedicated IORT Units

The IORT programs can use one of two approaches. In the first approach, patients are transported from the OR to the radiation therapy department and the IORT treatment is delivered on one of the department's linear accelerators that normally treats outpatients with external beam therapy. Thus IORT is blended in with the other treatments. In the second approach, IORT treatments are delivered on a dedicated machine. Dedicated linear accelerators are those accelerators that are used exclusively for the treatment of patients through the intraoperative technique. The word dedicated is used here to mean that the linear accelerator, conventional (X rays plus electrons or electron-only) or mobile, is used exclusively for IORT. It is recognized here that no more electron-only linear accelerators of the conventional, non-mobile type, will be built in the future. Dedicated linear accelerators may be located in the OR itself or in a room in the radiotherapy department. If the former, the room is basically an OR that also contains a linear accelerator. If the latter, it is a radiotherapy room that is equipped as an operating room. The former is preferred since it is part of the total OR complex and the patient does not have to be moved outside the room. In the



Figure 4. Mevatron ME by Siemens Medical Systems located in an OR. Note that the machine is a conventional accelerator whose gantry (C-arm) support system is mounted in the end wall. Since the room is slightly oversized for an OR room at this hospital, this still leaves adequate room for surgery.

latter case, the patient undergoes surgery far from the hospital's main OR and issues of maintaining sterility of the operating area have to be addressed as well as the problem of what to do if a surgical emergency arises for which the satellite OR is not equipped to cope. There are currently very few dedicated units in the OR in the United States; there are slightly more located in the therapy department. For the non-dedicated linear accelerator, the need to transport the patient for each case from the OR to the radiotherapy area is a disincentive to the IORT program. For this reason, programs using nondedicated linear accelerators have seen a decline in numbers (see Fig. 4). It is highly unlikely that any more dedicated units of the conventional or electron-only type will be installed in the United States in the future for the reasons outlined above. An example of a dedicated electron-only linear accelerator located in the OR is shown in Fig. 4. The room measures $\sim 6.1 \times 8.0$ m and the unit is mounted into the wall at one end, ensuring an adequate space for surgery.

Experience has shown that centers with dedicated OR suites or mobile systems perform more IORT procedures than centers that use the patient transport technique. Details of the shielding aspects of an OR-based IORT machine have been published (13,35).

Mobile IORT Units

The cost of installing a dedicated linear accelerator in an operating room along with the required shielding is very high. A linear accelerator, even though it is run only in the electron mode, has a price tag of $\sim \$1.5$ – 2 M and the shielding costs for a room that is not located in a basement, which is often the case, can be a substantial fraction of the linear accelerator cost. Given that the number of patients to be treated with an intraoperative machine is unlikely to be >5 per week, it is clear that the economics for such a unit are not altogether favorable (10). Since 1996, several mobile linear accelerators have become commercially

available. These are the Mobetron, produced by IntraOp Medical (Santa Clara, CA), the Novac7 produced by Hitesys (Aprilia, Italy) and the Liac produced by Info&Tech (Udine, Italy). These are linear accelerators that have special features, in addition to being mobile, that reduce the need for extensive shielding in the adjacent walls, ceiling, and floor. This means that they can be used for treating patients in any OR room with little more than a portable lead shield to protect personnel in the surrounding areas, as well as above and below. (However, note that for commissioning and annual quality assurance purposes, a well-shielded room is required because of the high beam-on time needed for these tests.) This reduction in radiation leakage has been achieved in several ways. The first is to limit the maximum energy of the electrons that can be accelerated so that photoneutrons are not generated; the Mobetron has a maximum energy of 12 MeV, the Novac7 has a maximum energy of 9 MeV and the Liac has a maximum energy of 10 MeV (Note that according to European law, energies >10 MeV cannot be used in an OR without special protective shielding in the walls and floor). The second is to use an in-line beam, that is, the direction of the electron beam in the waveguide is the same as the direction of the beam that treats the patient. This avoids the use of a heavy magnet. Thus avoids the use of a heavy magnet. Conventional medium-to-high energy linear accelerators generally use a 270° bending magnet to bend the electron beam from the horizontal direction in the waveguide toward the patient at isocenter (34). The reason is that with the need to have a fully isocentric C-arm machine with an acceptable isocenter height, (Fig. 4), typically ~ 125 cm above the floor, the waveguide cannot be pointing toward the isocenter, given the length of the guide. This reduction of leakage also means that the shielding around the waveguide (usually lead or a tungsten alloy) is lower than for a conventional therapy linear accelerator, so the weight of the unit is reduced—essential for good mobility. Thus the major advantage of a mobile electron linear

accelerator is that minimal, in-room shielding is required for the unit so that it can be moved to any OR room, provided mechanical access is possible. A comparison of some of the other properties of these three mobile linear accelerators is in order since their method of beam generation and delivery are different. Major differences are that the Mobetron is a gantry-mounted unit that uses soft docking for alignments, whereas the two Italian models have the waveguide mounted on a robotic arm. Furthermore, the Mobetron waveguide operates in the X-band mode (8–12 GHz) whereas the Italian models operate in the conventional S-band mode (3 GHz). Note that the length of the accelerator's waveguide depends inversely on the frequency, so, for the same final electron energy, the X-band waveguide will be shorter.

Mobetron. The first Mobetron was installed and treated its first patient in 1997. There are currently 12 Mobetron units installed worldwide, 7 operating in the United States, 1 in Japan, and 4 in Europe. This unit operates with electron energies of 4, 6, 9, and 12 MeV. A view of this unit is shown in Fig. 5. The unit is mounted on a gantry, cantilevered with a beam stopper to intercept the X-ray component of the primary beam after it has passed through the patient. This ensures that the radiation exposure in the room below is sufficiently low as to be within the limits set for the general public and is the principal feature that allows this unit to be used in any OR without additional shielding. Although the unit is gantry mounted, the beam direction is not limited to the plane of gantry rotation, as



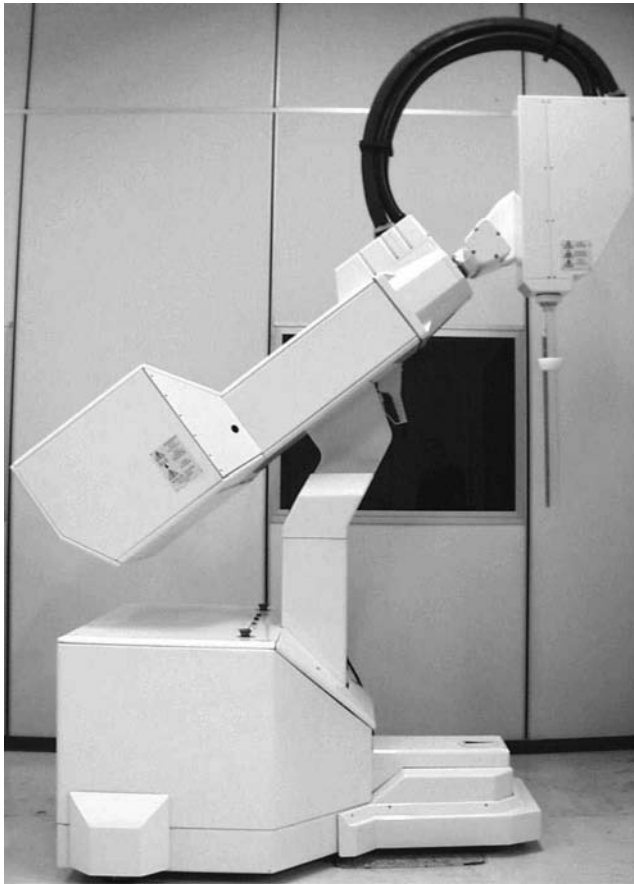
Figure 5. Mobetron mobile linear accelerator by IntraOp Medical Inc. Note that this unit has a gantry that also allows the linac head to pivot in a direction orthogonal to the gantry rotation plane.

with external beam linear accelerators, since the head can tilt in and out of this plane. This increases its versatility in setting up patients for treatment and reduces the amount of time needed to align the radiation field with the applicator. The beam stopper also moves in synchrony with this gantry motion to ensure that the primary beam is always intercepted.

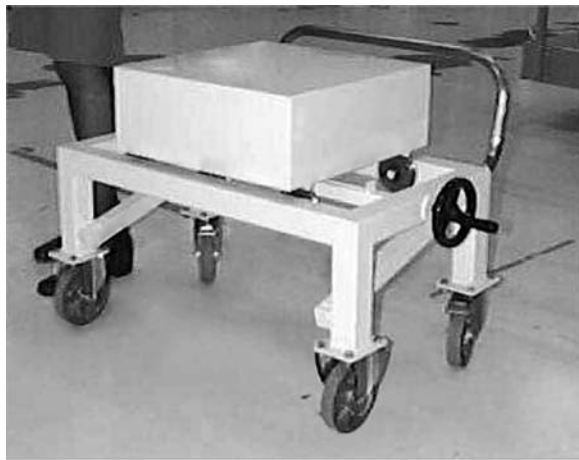
A review of the salient mechanical properties and operating parameters of the machine is in order. The weights of the treatment module and modulator are 1250 and 432 kg, respectively. The Mobetron is moved around through the use of a modified pallet jack. Thus the two units can be moved with reasonable ease between different ORs. The gantry can rotate off vertical by 45° in each direction, while the head tilt has a range of $\pm 30^\circ$. Since the distance between the machine target and the patient's skin is 50 cm, half that of a conventional linear accelerator, the maximum dose rate is $10 \text{ Gy}\cdot\text{min}^{-1}$. Thus the maximum treatment time is ~ 2 min. For setting up the machine to treat the patient, the unit has five degrees of freedom. There are two orthogonal translational motions whereby the stand can be moved relative to the base by up to ± 5 cm. There is also a translation motion of the head along the axis of the waveguide and, finally, there are 2 rotational degrees of freedom of the head, one the gantry rotation and the other tilt in and out of the gantry plane.

Meurk et al. (36) reviewed the physical properties of this device and Mills et al. (37) and Daves and Mills (38) provide a comprehensive review of the commissioning and shielding requirements of a Mobetron accelerator.

Novac7. The first Novac7 unit was installed in Rome in 1997. To date, there are 20 Novac7 systems operating in Italy, Germany, and Greece. The Novac7 operates with electron energies of 3, 5, 7, and 9 MeV. Unlike most conventional linear accelerators and the Mobetron, the Novac7 is mounted on the end of a cantilevered robotic arm, not a C-arm, has a total weight of 500 kg. Figure 6 shows the Novac7 unit as well as the beamstopper to attenuate the forward directed X rays. The beamstopper operates independently of the machine and must be put in place by the medical physicist responsible for the IORT procedure. In many situations, however, movable wall barriers must also be used. The applicators have diameters of 4, 5, 6, 7, 8, and 10 cm with available bevel angles of 0, 15, 22.5, 30, and 45° . Note, however, that while the unit can be adjusted mm by mm, the cylinder attached to the linac cannot be moved coaxially with the applicator; instead, movements are made through a combination of rotational movements. By using a robotic arm, the unit has more degrees of freedom available for aligning the electron beam with the cone set up in the patient than a conventional accelerator or the Mobetron. In addition to its movement across the floor, it possesses four rotational degrees of freedom, which provides flexibility in setting up the device for treatment. Interestingly, the machine does not use scattering foils to produce a broad, uniform beam, but, instead, relies on electrons scattering in the air within the tubes and from the walls of the tubes to produce the desired, flattened fields. As a result, the length of the applicator used depends on the applicator size selected.



(a)



(b)

Figure 6. (a) Novac7 mobile linear accelerator by Hitesys; (b) beam stopper for Novac7. Note that this unit is a robotic arm with several degrees of freedom. It also uses the hard dock procedure.

Partly because of this, the dose rate varies between 6 and 26 $\text{Gy}\cdot\text{min}^{-1}$, significantly higher than most linear accelerators. At these dose rates, treatments typically last <1 min. The lower dose rate corresponds to the largest applicator and lowest energy while the higher dose rate corresponds to the smallest applicator and highest energy.

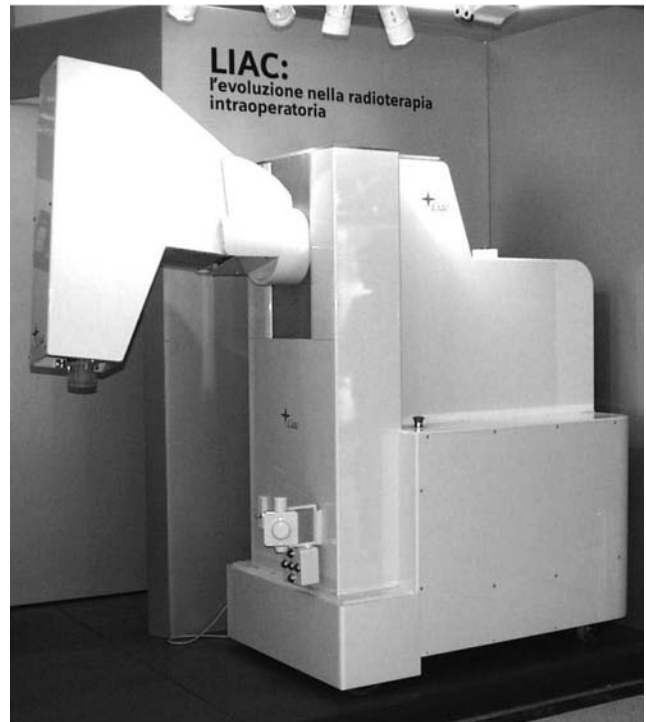


Figure 7. Liac mobile linear accelerator by Info&Tech. This unit looks similar to the Novac7 in terms of its movements and clinical set up, but there are differences in the design (see text for details).

These high dose rates result in substantial dosimetric problems, which have been studied and quantified by Piermattei et al. (39). A full technical description of the Novac7 has been given (40) and a review of the physical and dosimetric properties of the machine has been described by Tosi and Ciocca (41).

Liac. The first Liac unit (prototype) was installed in Milan in 2003. In its mechanical operation it is very similar to the Novac7 unit (see Fig. 7), with the accelerator guide mounted on a robotic arm. However, there are several major differences between the two units. First of all, the highest electron energy is now either 10 MeV (four energies of 4, 6, 8, and 10 MeV) or, potentially, 12 MeV (four energies of 6, 8, 10, and 12 MeV). This allows a greater penetration of tissue by the radiation, so that more deeply seated tumors can be treated. Second, the weight and dimensions of the machine are both lower, so the unit can be moved between rooms in the OR more readily, pass through doors more easily and fit into an elevator. The applicator system is the same as that of the Novac7, namely, it uses a hard docking procedure with no scattering foils. However, the applicator that defines the extent of the radiation field in the patient is now 30 cm in length. This length is standard for most commercial and noncommercial systems and allows the radiation oncologist and surgeon to view the radiation field directly, just before docking the applicator with the machine.

The Liac has the same high instantaneous dose rates as the Novac7, in this case 1.5–13 $\text{Gy}\cdot\text{min}^{-1}$ for the 10 MeV version and 3–22 $\text{Gy}\cdot\text{min}^{-1}$ for the 12 MeV version.

Table 4. Variation in the Depth of the 90% Dose with Electron Beam Energy

Electron Beam Energy, MeV	Depth of 90% Dose, cm ^a
6	1.7
9	2.6
12	3.7
15	4.5
18	5.1

^aFor a 7 cm diameter circular applicator that is commonly used in IORT; note that in IORT, the dose is normally prescribed to 90% level, taking into account that the surface dose is at or about that level.

Comparison between Conventional and Mobile Units

It is worthwhile comparing the pro’s and con’s of conventional linear accelerators in the OR versus mobile linear accelerators. The conventional machine has the advantage that since it is fully shielded for the highest energy available, ~18 MeV or higher, tumors with a greater depth can be treated. As a general rule, the depth of treatment increases by ~1 cm for every 3 MeV increase in energy at the 80% dose. Table 4 shows the depth of the 90% dose for one conventional unit.

However, the question arises as to how many treatments are delivered at high energy versus low energy. Figure 8 shows data from the MGH on the use of the various electron energies. It can be seen that all but 15% of the treatments were at 12 MeV or less, indicating that high energies are infrequently used. In contrast, the disadvantages are the cost of the machine, generally greater than for a mobile unit, and the shielding for neutrons as well as photons. At MGH, this shielding amounts to ~50 metric tons, which makes retrofitting existing ORs all but impossible. Moreover, should the machine malfunction, the OR becomes unavailable for the duration of the repair. Another fundamental difference is that conventional machines use a 270° bending magnet with momentum slits to focus the electrons onto the scattering foil, whereas the mobile units have no energy or momentum selection. While both units change energy by adjusting radio frequency (rf) power, the conventional units have the advantage of better energy selection. This leads to a higher surface dose and slower fall-off in the depth dose curves for mobile units (48).

Mobile linear accelerators, on the other hand, have distinct advantages over the conventional unit. First of all, the units are mobile so that they can be moved to any OR that has sufficient space. This is possible because the units require almost no additional shielding to operate

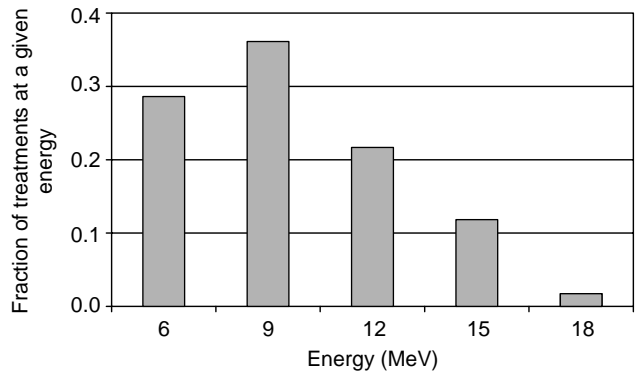


Figure 8. Fraction of treatments delivered at a given energy as a function of the energies available. From this graph, one can see that the majority of treatments are at 9 MeV or less and that <15% or patients are treated with an energy >12 MeV.

safely. However, a beam stopper is required for the primary beam in all mobile units. Whereas a dedicated unit is limited to isocentric movement, all mobile units have more degrees of freedom, allowing for greater flexibility in setting up the patient.

A summary of the chief differences between the three different modes of IORT is given in Table 5.

Treatment Applicators

There have been two types of applicators used in electron beam IORT, those made from poly methyl methacrylate (PMMA) and those made of metal, usually chrome-plated brass. The supposed advantage of the PMMA applicators was that one could see the tissues to be irradiated through the walls, although this turned out to be largely not the case. The disadvantage of using PMMA applicators is that they have to be sufficiently thick so as to prevent radiation penetrating through the walls and damaging normal tissue; this minimum thickness is ~6 mm. Metal applicators, on the other hand, do not have to be so thick because of their greater density. Since the pelvis is a tight area, anatomically, having a thick-walled applicator can restrict the treatment area for a pelvic side wall lesion. Thus, metal applicators are preferred, also because they can be flash (steam) sterilized at the time of the procedure, whereas applicators made from PMMA must be gas sterilized to avoid heat-related deformities and this requires ~12 h. Figure 9 shows examples of plastic applicators. Examples of metal cones are shown in Fig. 10.

Table 5. Comparison among the Three Methods of IORT Using Linear Accelerators

Mode of IORT	Advantages	Disadvantages
Linac in oncology department (patient transport)	Inexpensive; needs minimal additional equipment	For a busy outpatient machine, can do only 1–2 cases per week. Enthusiasm wanes due to effort required
Dedicated conventional linac in OR	Available on full-time basis. Maximizes convenience for surgical staff	Requires expensive shielded suite with low use factor
Mobile accelerator	Can be used in almost any OR; minimal additional shielding required	Limit on the maximum energy due to leakage X rays and neutrons

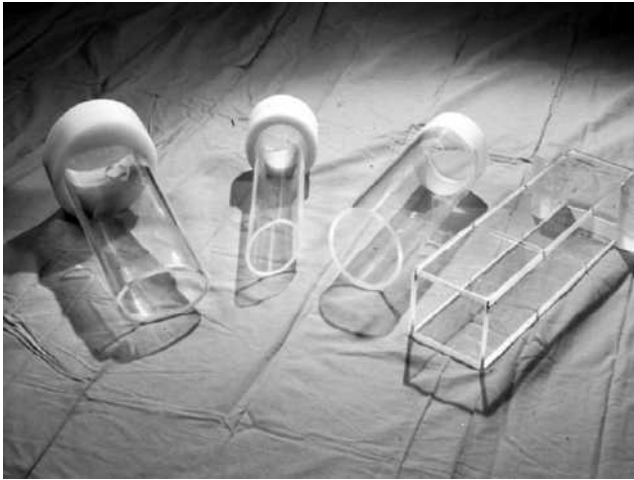


Figure 9. Examples of plastic applicators used in electron IORT. From left to right are shown an elliptical, a circular with beveled end, a circular, and a rectangular applicator.

Most manufacturers of dedicated linear accelerators provide applicators with a range of dimensions, usually circular in cross-section. A description of the design of treatment applicators for an OR-based IORT unit has been given by Hogstrom et al. (42)

Beam Alignment Devices

For nondocking machines, a system is required to align the electron beam from the waveguide with the axis of the treatment applicator in the patient (this is unnecessary for hard-docking systems since alignment is guaranteed with the tolerance between the two mating pieces). This is usually achieved optically. In the case of one dedicated, conventional linac (Siemens Mevatron ME, Concord, CA; this unit is no longer manufactured), beams from two lasers are split into four point sources and four line sources. When the points converge with the lines on a particular radius, then the distance is correct and the beam axis is correctly angled with respect to the treatment applicator. A metal disk with this radius drawn on it is placed on top of the treatment applicator for this purpose. A view of how



Figure 10. Examples of metal cones used in electron IORT. From left to right are shown a circular with bevel, a circular, an elliptical, and a rectangular applicator.



Figure 11. Example of the optical alignment used on the Mevatron ME linear accelerator. This test jig is used before each procedure to check the alignment of the lasers. Alignment is correct when the dots and lines (not easily seen here, but overlapping the radial lines) cross on the circle.

the correct alignment should look is shown in Fig. 11 for the test jig. Clearly shown is the circle with four laser dots at 90° intervals; the laser lines are in alignment with the radial lines shown, but on a black/white image, they are not visible. This system requires training to fully understand what movements should be made with the couch and gantry to align the treatment applicator with the beam when the dots and lines are not aligned. A more sophisticated system, that is simpler to use, is adopted with one of the mobile linear accelerators (Mobetron by IntraOp Medical, Inc.). This system also uses optical alignment, but provides visual indicators as to how to adjust the gantry to complete the alignment. As with the other system, laser light is reflected off a mirror on top of the applicator and is sensed by detectors in the head. This display is shown in Fig. 12, which shows the status of the alignment that is complete

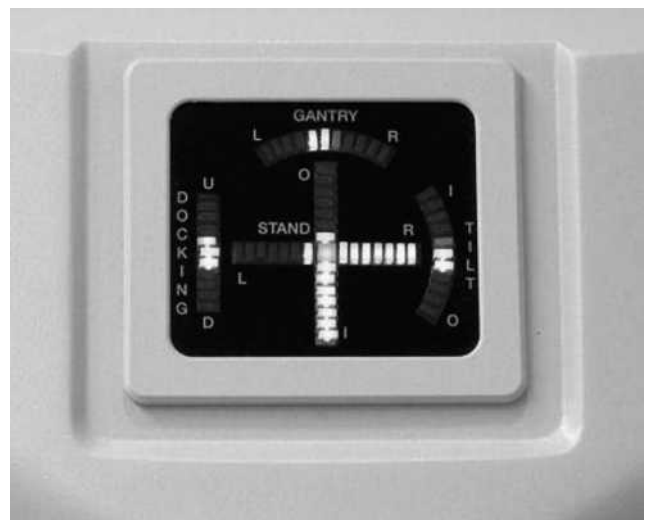


Figure 12. Display of the alignment system used with the Mobetron mobile unit. (Photo courtesy of G. Ezzel, Mayo Clinic.)

when each of four parameters is indicated by a single yellow bar either side of a green bar (colors not indicated). Parameters on the display are the gantry rotation, the docking or distance between the head and the applicator, the tilt of the head and the position of the stand (in-out and left-right). The display shown indicates that the stand has to be moved from in to out and from right to left to complete the docking. When the alignment is complete, the alignment is accurate to 2 mm along any one axis and 0.1° in angle.

IORT Treatment Logistics

Since the radiation is delivered through linear accelerators, there are several options for performing this technique. The first is that the patient is transported from the operating room (OR) to the radiation therapy department and set up and treated on one of the linear accelerators usually used for treating cancer patients with external beams. The second is to have a dedicated treatment room in the OR. In this case, the patient has to be moved only from one side of the room to the other where the linear accelerator is located. The third option is a mobile linear accelerator. A mobile linear accelerator is one that can be moved from one OR to another. A fourth option is a hybrid of the first and second methods in that a linac in the radiation therapy suite serves also as an operating room so that both the surgery and IORT are carried out in the same room. The problem with this option is that if this room is far from the other operating rooms and if difficulties are encountered during the surgery, assistance might be problematic, particularly if specialized equipment is required.

There is a significant difference between the first option and the other three options since the patient has to be transported from the OR to the radiation therapy area. Once there, the machine used for these treatments is sequestered for the duration of the set-up and treatment. Additional time is required before and after the IORT procedure to prepare and clean up the room. Thus there are two negative effects, one, the surgical procedure is lengthened because of the transport and reset-up of the patient, and two, the therapy department loses significant time on one of its linear accelerators. Clearly, a department having only one or two machines with a busy outpatient workload could not readily perform these procedures, or, at least, not more than once a week. Finally, while the patient is being transported or in the therapy area, the OR has to be held open for surgical closure or additional surgery and this has an impact on a busy surgical department and its resources.

IORT Dosimetry, Calibration, QA, and Radiation Safety Electron Beam Calibration, Dosimetry, and Quality Assurance. Important components of IORT dose delivery are the calibration of the of the electron beam, dosimetry and routine quality assurance. These three important areas of IORT will be addressed separately. Electron beams are calibrated using a recommended methodology (43) and according to national protocols. In the United States and Canada, the appropriate protocol has been established by the American Association of Physicists in Medicine (AAPM) (44). This protocol dictates that measurements be made using an

ionization chamber in a water phantom under specific conditions. The ionization chamber is typically a cylindrical chamber having an air volume of 0.6 cm^3 , although the protocol also recommends the use of plane parallel chambers for low energies. The specific conditions relate to the depth of measurement as a function of electron beam quality and chamber dimensions. The protocol provides factors to calculate the absolute dose from the measured chamber exposure. The chamber is calibrated at an AAPM certified laboratory whose calibration standards are, in turn, referenced to the National Institute for Standards and Technology (NIST). Thus users' chambers around the country have calibration factors that are traceable to NIST so that 1 cGy in one institute is equal to 1 cGy in any other institute. The 2σ uncertainty in the users' calibration is stated to be 1%.

Having calibrated the IORT electron beam under standard conditions, it is necessary to determine the dose delivered to the patient under the most general conditions. For this purpose, medical physicists perform a series of measurements on the machine, including the variation of dose with depth and applicator size. Using these measurements, medical physicists can recommend the optimal energy for treating a specific lesion given the depth of the lesion and will then set the parameters on the accelerator to deliver the dose prescribed by the radiation oncologist.

Quality assurance is a very important aspect of radiation therapy, both external and IORT. In general, quality assurance is mandated by state and federal agencies and recommendations are made by the AAPM (45). These laws and recommendations specify the types and frequency of tests that must be carried out on linear accelerators used in radiation therapy. For IORT, quality assurance is of particular importance since, unlike external radiation therapy where up to 42 fractional doses may be given over an 8 week period, the dose is given in a single fraction. Therefore, quality assurance tests must be particularly probing to ensure that the probability of any error is as low as possible (46). The AAPM has made specific recommendations for IORT (47,48) at the daily, monthly and annual level. Table 6 shows the data taken from the latest AAPM report (48).

In addition, additional, independent quality assurance on electron beam output and percent depth dose is provided by the Radiological Physics Center in Houston, using mailable TLDs (49).

Radiation Safety Issues

All radiation-producing equipment is subject to state and federal regulations. These regulations restrict the dose that a member of the general public can receive to 1 mSv year^{-1} or $0.02 \text{ mSv-week}^{-1}$. There is also a further restriction that such a person may receive no more than 0.02 mSv in any 1 h. For controlled areas (not accessible to the general public), the allowed limits are higher. Since many of the linear accelerators are located/used in the operating room environment, regulations for the general public apply.

As noted above, radiation safety requirements for electron beam IORT depend on the approach taken. For treatment

Table 6. Quality Assurance Tests for Mobile Linear Accelerators

Frequency	Parameter
Daily	Output constancy
	Energy constancy
	Door interlocks
	Mechanical motions
Monthly	Docking system
	Output constancy
	Energy constancy
	Flatness and symmetry constancy
Annual	Docking system
	Emergency off buttons
	Output calibration for reference conditions
	Percent depth dose for standard applicator
	Percent depth dose for selected applicators
	Flatness and symmetry for standard applicator
	Flatness and symmetry for selected applicators
	Applicator output factors
	Monitor chamber linearity
	Output, percent depth dose and profile constancy over the range of machine orientations
Inspection of all devices normally kept sterile	

^aAAPM task group 72.

with a linac in a radiotherapy department, the room is already well shielded for any number of IORT treatments. For a dedicated, conventional linear accelerator in the OR, the shielding is usually designed for a specific number of cases per week. A maximum number would be 10 cases per week, based on the length of an average surgical procedure. However, patient demographics and past IORT experience indicate that this limit is never reached on a continuing basis. However, the machine has to undergo acceptance testing and commissioning when first installed and extensive checks every year by the medical physicists that require extensive beam-on time (there are also daily and monthly checks, but these require much less beam-on time). Therefore, this work has to be carried out at night and over the weekend when personnel are generally not present. For corridors in the OR that are still in use, this may require installing temporary barriers with appropriate radiation warning signs; rooms above and below have to be checked to ensure that they are not occupied. For mobile linear accelerators, it was noted above that, due to their low leakage, considerably less shielding is required for their use in the OR, apart from a primary beam stopper. Lead shields placed strategically around the patient provide secondary shielding. However, before any OR can be used for treatment with a mobile linear accelerator, extensive surveys have to be carried out to determine radiation levels in the immediate vicinity with and without the secondary shielding. Based on these readings, the number of cases that can safely be performed on a weekly basis without exceeding the maximum permissible dose can readily be calculated. Clearly, this exercise has to be performed for every OR in which IORT cases may be performed. As noted above for the dedicated, conventional linear accelerator, there are occasions when the medical physicist has to perform extensive testing of the equipment. This can also only be done at night or on weekends, provided the occupancy of nearby areas can be controlled. If this cannot be done, the unit

has to be taken to a shielded area for testing, such as the radiotherapy department. Issues relating to the radiation safety aspects of mobile linear accelerators are fully covered in the AAPM's report on mobile linear accelerators (48). In all cases, consultation with the hospital's radiation safety officer is highly recommended.

OTHER IORT TECHNIQUES

INTRABEAM System for Intracranial Lesions

Another device that has come into use for the intraoperative treatment of intracranial and other lesions is the INTRABEAM System. This device was originally manufactured by Photoelectron Corporation, but is now marketed by Care Zeiss Surgical GmbH, Germany. This system is a 50 kV device that fits into a standard neurosurgical stereotactic frame. Figure 13 shows the INTRABEAM device; electrons are accelerated down a 10 cm long 3.2 mm diameter evacuated tube, striking a gold target at the tip. The thicker section between this tube and the body of the device contains coils to steer the electron beam. Built into the body of the device is a scintillation detector that detects backward emitted photons in a fixed geometry. This



Figure 13. INTRABEAM X-ray tube. The body of the device contains the high voltage electronics, power for the steering coils and the electron gun. It also contains the internal radiation monitor for integrating the radiation output of the device, similar to the internal ionization chambers in a linear accelerator. Electrons are accelerated to the tip of the probe where they strike a gold target and produce the X rays. The large diameter section contains the steering coils.



Figure 14. The INTRABEAM device inserted into a stereotactic head frame demonstrating how the device is used for the treatment of intracranial lesions.

is known as the internal radiation monitor. Figure 14 shows the X-ray device inserted into a CRW neurosurgical frame so that the probe tip can be placed at the location of the stereotactic site to within an accuracy of better than 1 mm. Thus an intracranial lesion is first imaged using Computerized tomography (CT) and the coordinates of the center of the lesion determined so that, in addition to the neurosurgeon performing a stereotactic biopsy, the tip of the X-ray source, where the X rays are generated, can be accurately located at this point and a tumoricidal dose of radiation delivered. Figure 15 shows a series of isodose curves generated using radiochromic film (Gafchromic film, type MD-55, ISP Technologies Inc., Wayne NY). The outline of the probe tip in the north-south direction

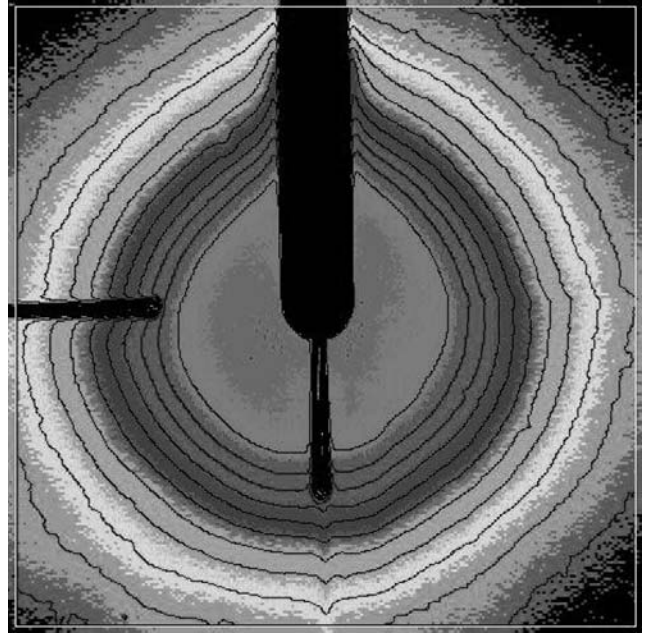


Figure 15. Isodose curves for the INTRABEAM X-ray device. It can be seen that the tip of the probe is the center of the X-ray source and that the source is nearly isotropic in intensity except in the backward direction close to the tube.

can easily be identified as the source of the radiation. The horizontal line and the line extending from the probe tip represent cuts in the film. The dose depth curve for 50 kV X rays in water is shown in Fig. 16. The dose rate from this device is about $2 \text{ Gy} \cdot \text{min}^{-1}$ at 1 cm from the probe tip in water. The dose falls off as the inverse third power of the distance and, hence, the dose delivered outside the tumor drops very rapidly. At this dose rate and for this dose fall-off, typical treatment times last from 15 to 30 min, depending on the treatment radius. A fuller description of the apparatus and the clinical testing of the device has been given by Beatty et al. (50). A Monte Carlo has been performed (51) to predict both the spectrum and output of the X-ray beam with good accuracy.

Although initially developed for intracranial lesions, the INTRABEAM system is approved by the U.S. Food and

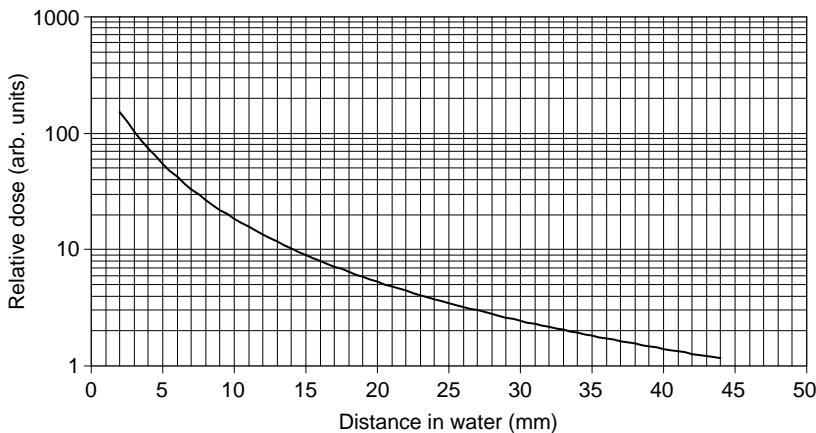


Figure 16. Distance-dose curve for the INTRABEAM device for 50 kVp X rays. The output falls off as the inverse of the third power of the distance from the probe tip. In addition to inverse square, there is an additional factor due to the attenuation of the X-ray beam, which is significant at these energies.

Drug Administration (FDA) for application to any location in the body. To that end, a set of spherical applicators was developed that could be used to treat meningiomas or breast cancer, after surgical resection. For breast and other extracranial treatments, the stereotactic frame is not used. Instead, a stand developed by Zeiss is used to support the INTRABEAM device. It can be set and locked into any position so that it remains stationary during the procedure. In Europe, the treatment of breast cancer is one of the principle uses of this device. There are currently ~30 units operating in the United States and Europe. Of note is the fact that since this device delivers X rays at very low energies, the shielding requirements are minimal. All personnel, except for the operator and perhaps the anesthesiologist, leave the room. A portable diagnostic X-ray shield is used to protect the operator and anesthesiologist. All entry doors except one, beyond which all the other OR personnel wait during the treatment, are locked.

High Dose Rate (HDR) Brachytherapy

Brachytherapy is a technique whereby sealed radioactive sources are placed inside the body or in body cavities, usually for a short duration of time or, in some cases, permanently. Most of the sources for temporary implants have activities such that the dose rate to an isodose line surrounding the tumor is $\sim 50 \text{ cGy}\cdot\text{h}^{-1}$. With the typical doses given in conventional, low dose brachytherapy, this means that the treatment time is often several days and the patient has to be hospitalized. However, by using a HDR source (typically 10 Ci of ^{192}Ir), the treatment time can be shortened to minutes. Figure 17 shows an HDR (Nucletron Microselectron HDR) with one catheter inserted. The other end of the catheter is placed in a well counter, a chamber used by medical physicists to measure the strength of the radioactive source. A total of 18 catheters can be attached to the unit and a computer control system (not shown) guides the single source down each catheter to a given position for a predetermined dwell time. By varying the dwell time of the source at each location, the dose distribution can be optimized for each treatment. In



Figure 17. View of the HDR system with one catheter hooked up to a well chamber for calibration of the source by a medical physicist.

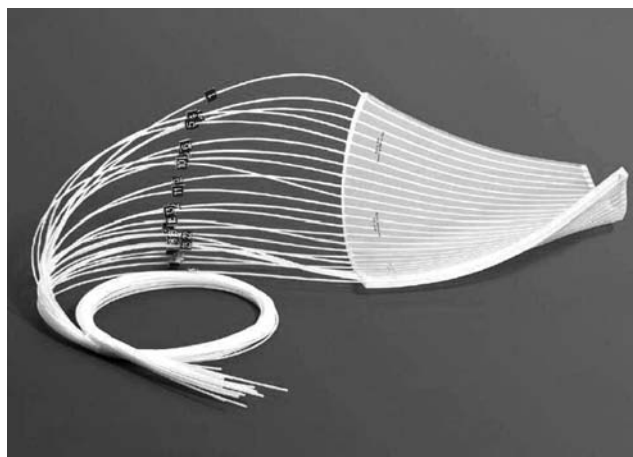


Figure 18. Sample H.A.M. applicator. The catheters are embedded in a 1 cm thick silicone rubber mat. Note the clear identification numbers on each catheter. The manufacturer supplies applicators with varying numbers of catheters. (Photo courtesy of Felix Mick.)

this procedure, as many catheters as are required to cover the tumor or tumor bed are inserted into the patient or placed on the outside for skin treatments; a single source is used and this is mechanically driven along each catheter, programmed to dwell at preselected positions for lengths of time determined by the treatment planning program. The HDR treatments took some time to be accepted by the medical community because it was initially thought that dose rate effects in brachytherapy were crucial and that HDR treatments would be less effective than those using conventional, low dose rate techniques. Because of this shortened treatment time, it became possible to perform exploratory surgery in the OR, lay out the catheters in the area of residual disease, treat the patient while still on the operating room table and then close up the patient. Figure 18 shows a Harrison-Anderson-Mick (H.A.M.) applicator for HDR-IORT. This applicator consists of a number of catheters embedded in a 1 cm thick silicone rubber mat. The catheters are located mid-plane and therefore 0.5 cm from each surface. Applicators are available with the edges curved in a half circle (not shown here) to help maintain a more uniform dose at the surface. Applicators are available in sizes from 3–24 channels in groups of 3, namely, 3, 6, 9, and so on. Because this is a planar application of ^{192}Ir sources, the depth of treatment is almost always taken to be 1 cm, that is, 0.5 cm from the surface of the applicator, to avoid an excessively high dose at the skin-applicator interface. This restricts the number of applications for which this approach can be used, typically lesions that have been resected with positive margins, compared with electron beams, which, as shown above, can penetrate much greater depths. Care must be exercised in the planning dosimetry for these applications since the dose depends on the curvature of the applicator (52) and the lack of backscatter material (53). Figure 19 shows an applicator in place in a patient prior to treatment. Note the curvature of the applicator.

Shielding, in the form of concrete or lead is required in the walls and floor if a significant number of patients are

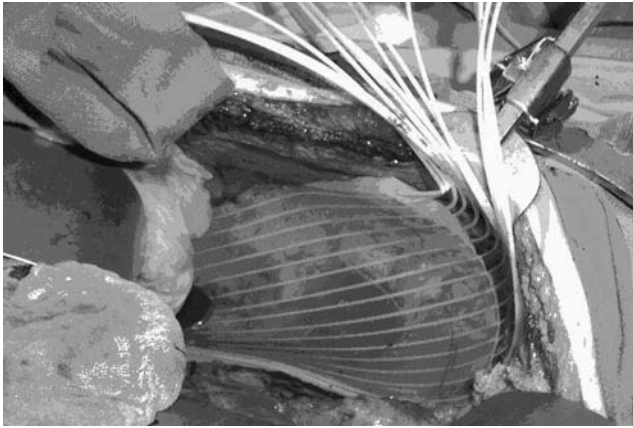


Figure 19. View of a H.A.M. applicator inserted in a patient and ready for treatment. (Photo courtesy of Dr. C. Willett.)

treated. The exact thicknesses will depend on the number of cases treated per week.

At present, there are four institutions in the United States performing HDR in the OR. Excellent technical descriptions of the HDR technique applied in the operative setting have been given by Harrison et al. (54) and Nag et al. (21).

CLINICAL RESULTS

Disease sites that have been treated with IORT include pelvic lesions such as pancreas, rectum, colon, retroperitoneal sarcoma, bladder, kidney, sacrum, liver, endometrium, cervix, and ovary. Extra-pelvic sites include head and neck, breast, and distal limbs. At the MGH, the most commonly treated disease sites include the pancreas, rectum, colon, and retroperitoneal sarcoma. UCSF reported in 2003 that a majority of treatments had been in the area of head and neck and thorax. A considerable number of pediatric patients had also been treated in a few centers. Thus the type of sites treated in different institutions reflects the interests of the surgical and oncology departments of those institutions.

One of the problems encountered when comparing the efficacy of IORT combined with external beam radiotherapy (EBRT) with EBRT alone in terms of overall survival is that distant metastases play a major role in many disease sites. While it is true that IORT combined with EBRT provides a result as good as any EBRT alone series, the difference in survival is often not significant. Moreover, due to the rapid progression of distant metastases, it is often not possible to know the exact pathologic stage of each patient under treatment, and so comparisons among different series is difficult. Data from MGH (unpublished) shows that overall survival for patients with pancreatic cancer depends on the field size of the treated field, the smaller the field, the greater the survival. What this says is that patients with bulkier disease will experience more rapid development of distant metastases.

One of the success stories of IORT, however, is colorectal cancer since the effect is statistically significant. Table 7 shows the results from MGH for primary and recurrent

Table 7. Clinical Results for IORT Treatment of Primary and Recurrent Rectal Cancer Using Electrons

Primary Colorectal Cancer					
	No. Pts	Median (month)	Survival, years		
			2	3	5
EBRT	17	18	35%	24%	24%
EBRT+IORT	56	40	70%	55%	46%
Recurrent Colorectal Cancer					
	No. Pts	Median (month)	Survival, years		
			2	3	5
No IOERT	64	17	26%	18%	7%
IORT+/-EBRT	42	30	62%	43%	19%

colorectal cancer. This was not a randomized trial, but a comparison between EBRT with electron beam IORT and historical controls, EBRT alone. In each case, the survival is greater when IORT is used, even up to 5 years. While this does not have the power of a randomized trial, the difference in the results in each case is both statistically significant and impressive. Moreover, similar results have been observed at other institutions, for example, the Mayo Clinic (55). Similar results have been obtained with HDR-IORT.

Current results in the treatment of gastric cancer indicate that while some benefits may accrue from IORT, further studies are needed to demonstrate a significant benefit. Interestingly, gastric cancer was one of the sites first chosen by the Japanese for IORT, which is quite understandable, given the high rate of incidence of that disease in Japan. For locally advanced primary and recurrent gynecological malignancies IORT has shown that results are comparable to historical results with standard salvage therapy. In one of the rare randomized trials involving IORT, a study of retroperitoneal sarcoma accumulated 35 patients between two arms, EBRT and EBRT+IORT. With a minimum follow-up of 5 years and a median follow-up of 8 years, a significant difference in local control was found between the two groups (56). However, as with studies for other diseases, there was no difference in median survival between the two groups. Two major centers, the Mayo Clinic and the Denver Children's Hospital, have been involved with pediatric malignancies. They have shown that IORT is effective in local control for locally advanced pediatric malignancies. Interestingly, the Denver Children's Hospital used only IORT whereas the Mayo Clinic used combined IORT and EBRT. In bladder cancer, IORT has been shown to be effective in bladder preservation. Investigators found few complications related to these treatments.

Perhaps the greatest thrust at the moment in the field of IORT is the treatment of breast cancer. This is actively being pursued in Europe, particularly Italy, followed by Germany and the United Kingdom (UK), and, to a lesser extent, in the United States. The drive for this is related to the need to shorten the overall duration of the conventional

external therapy treatment time. In this area, IORT is competing with several other techniques, including partial breast irradiation and temporary implants using ^{192}Ir brachytherapy seeds and the INTRABEAM system. The IORT and the INTRABEAM systems would, of course, require only a single fraction treatment. Although a large number of breast cancer patients have been treated intraoperatively to date, the follow-up time is too short to draw any conclusions.

At MGH, interstitial radiosurgery with the INTRABEAM has been performed for intracerebral metastases, primary malignant gliomas and as adjunct to resective surgery in meningiomas. Local control of metastases was possible in 80–85% of cases ($n = 73$), which is consistent with other forms of external radiosurgery. In a handful of cases with malignant gliomas, tumor recurred as expected beyond the areas of high dose irradiation. In 12 parasagittal meningiomas where residual tumor was left along the sagittal sinus, there has been no evidence of local recurrence with mean follow-up of 3 years, which is too early to determine efficacy.

FUTURE DIRECTIONS

If one looks at the history of IORT, one sees that the procedure has been practiced almost as long as there have been X rays. One of the driving forces for IORT in the early days was undoubtedly the need to overcome the limitation in dose delivery due to skin desquamation, a breakdown of the skin, resulting from the highest dose being at the surface. Certainly, the conditions of using X-ray equipment available at that time could not have been conducive to ease of use in an OR setting. It is clear, however, that the history of its practice has been punctuated by the introduction of new technology, first the change from low kilovoltage X rays (50 kV) through orthovoltage X rays to electrons. However, it was clearly the introduction of linear accelerators with electron beams that provided the greatest impetus over the last century, although the technique was not universally adopted because of cost, time, and many other factors. Indeed, there was a decline after the 1990s associated with these factors. It was not until the introduction of mobile linear accelerators that the field regained its strength. Part of the reason for this is the desire, for logistical reasons, particularly in Europe, to speed up the radiation therapy treatment process. Granted that this is also possible using external beam techniques, such as partial breast irradiation, but IORT certainly has a role to play in this change in therapeutic practice. The increase in the number of units attests to this: In 1997, the first unit was installed and at the time of this writing, 33 have been installed worldwide with units available from three manufacturers. This technology has enabled more centers to participate in IORT procedures, allowing both surgeons and radiation oncologists more convenient ways to treat patients. Moreover, several photon modalities have come into increasing use in IORT, including HDR-IORT and the Zeiss INTRABEAM device. These are playing a greater role in IORT, in part because the capital investment in these technologies is considerably less than a dedicated mobile linear accelerator.

However, before the widespread acceptance of IORT is possible, proponents need to demonstrate clearly that IORT provides a therapeutic advantage for their patients. While this has been shown for a subset of patients with rectal cancer, based on retrospective studies, it is clear that randomized trials are needed for other disease sites to provide unequivocal evidence for IORT's benefit. This is currently being done for breast cancer in Europe.

Thus, while the field is still small, it continues to grow, thanks to better technology. More widespread use of IORT will determine whether the procedure confers a benefit to patients across a broad class of diseases.

BIBLIOGRAPHY

1. Abe M, et al. Intra-operative radiation in abdominal and cerebral tumors. *Acta Radiol* 1971;10:408.
2. Abe M, et al. Intraoperative radiotherapy of gastric cancer. *Cancer* 1974;34:2034.
3. Abe M, et al. Techniques, indications and results of intraoperative radiotherapy of advanced cancers. *Radiology* 1975;116:693.
4. Abe M, Takahashi M. Intraoperative radiotherapy, The Japanese experience. *Int J Radiat Oncol Biol Phys* 1981;7: 863–868.
5. Goldson A. Preliminary clinical experience with intraoperative radiotherapy (IORT). *Semin Oncol* 1978;8:59–65.
6. Gunderson LL et al. Intraoperative irradiation. A pilot study combining external beam photons with "boost" dose intraoperative electrons. *Cancer* 1982;49:2259–2266.
7. Tepper J, Sindelar WF, Glatstein E. Phase I study of intraoperative radiation therapy combined with radical surgery for intra-abdominal malignancies. *ASCO Proc* 1980;21:395.
8. Gunderson LL et al. Intraoperative and external irradiation with or without resection: Mayo pilot experience. *Mayo Clin Proc* 1984;59:691.
9. Coia LR, Hanks GE. The need for subspecialization, intraoperative radiation therapy. *Int J Radiat Oncol Biol Phys* 1992;24:891–893.
10. Wolkow H. The economics of intraoperative radiation therapy. In: Dobelbower RR, Abe M, editors. *Intraoperative Radiation Therapy*. Boca Raton, (FL): CRC Press; 1989.
11. Papillon J. Conservative treatment by irradiation as an alternative to surgery. *Rectal and Anal Cancers*. New York: Springer Verlag; 1982.
12. Rich TA, Piontek RW, Kase KR. The role of orthovoltage equipment for intraoperative radiation therapy. In: Dobelbower RR, Abe M, editors. *Intraoperative Radiation Therapy*. Boca Raton, (FL): CRC Press; 1989.
13. Biggs PJ, et al. Dosimetry, field shaping and other considerations for intra-operative electron therapy. *Int J Radiat Oncol Biol Phys* 1981;7:875–884.
14. Palta JR, Suntharalingam N. A nondocking intraoperative electron beam applicator system. *Int J Radiat Oncol Biol Phys* 1989;17:411–417.
15. Kharrati H, Aletti P, Guillemain F. Design of a non-docking intraoperative electron beam applicator system. *Rad Oncol* 1999;33:80–83.
16. Jones D, Taylor E, Travaglini J, Vermeulen S. A non-contacting intraoperative electron cone apparatus. *Int J Radiat Oncol Biol Phys* 1989;16:1643–1647.
17. Björk P, Knöös T, Nilsson P, Larsson K. Design and dosimetry characteristics of a soft-docking system for intraoperative radiation therapy. *Int J Radiat Oncol Biol Phys* 2000;47: 527–533.

18. Sindelar WF, Johnstone PAS, Hoekstra HJ, Kinsella TJ. Normal tissue tolerance to intraoperative irradiation. In: Gunderson LL, Willett CG, Harrison LB, Calvo FA, editors. *Intraoperative Irradiation: Techniques and Results*. Totowa, (NJ): Humana Press; 1999.
19. Gillette EL, Gillette SM, Powers BE. Studies at Colorado State University of normal tissue tolerance of beagles to IOERT, EBRT or a combination. In: Gunderson LL, Willett CG, Harrison LB, Calvo FA, editors. *Intraoperative Irradiation: Techniques and Results*. Totowa, (NJ): Humana Press; 1999.
20. Hall EJ. *Radiology for the radiologists*. Hagerstown, (MD): Harper and Row; 1978.
21. Nag S et al. Intraoperative irradiation with electron-beam or high-dose-rate brachytherapy: Methodological comparisons. In: Gunderson LL, Willett CG, Harrison LB, Calvo FA, editors. *Intraoperative Irradiation: Techniques and Results*. Totowa, (NJ): Humana Press; 1999.
22. Beck C. Über Kombinationsbehandlung bei bösartigen Neubildungen. *Berl Klin Wochenstr* 1907;44:1335.
23. Beck C. On external Roentgen treatment of internal structures (eventration treatment). *N Y Med J* 1909;89:621–622.
24. Finsterer H. Zur Therapie inoperabler Magen- und Darmkarzinome mit Freilegung und nachfolgender Röntgenbestrahlung. *Strahlentherapie* 1915;6:205–218.
25. Abe M. History of intraoperative radiation therapy. In: Dobelbower RR, Abe M, editors. *Intraoperative Radiation Therapy*. Boca Raton, (FL): CRC Press; 1989.
26. Barth G. Erfahrungen und Ergebnisse mit der Nahbestrahlung operativ freigelegten Tumoren. *Strahlentherapie* 1953;91: 481–527.
27. Goin LS, Hoffman EF. The use of intravesical low voltage contact Roentgen irradiation in cancer of the bladder. *Radiology* 1941;37:545.
28. Biggs PJ, Noyes CG, Willett CG. Clinical physics, applicator choice, technique and equipment for electron intraoperative radiation therapy. In: Petrelli N, Merrick III HW, Thomas Jr CR, editors. *Surgical clinics of North America*. Philadelphia, PA: W.B. Saunders; 2003.
29. Hensley FW, Ciocca M, Petrucci A, Biggs PJ. Survey of IORT activities in Europe. IVth International meeting of the International Society of Intraoperative Radiation Therapy. Miami, (FL): Mar 17–19, 2005.
30. Schultz MD. The supervoltage story. *Amer J Roentgenol Rad Therapy Nucl Med* 1975;124:541–559.
31. Henschke G, Henschke U. Zur Technik der Operationsbestrahlung. *Strahlentherapie* 1944;74:228–239.
32. Eloesser L. The treatment of some abdominal cancers by irradiation through the open abdomen combined with cauterization. *Ann Surg* 1937;106:645–652.
33. Fairchild GC, Shorter A. Irradiation of gastric cancer. *Br J Radiol* 1947;20:511.
34. Podgorsak E, Metcalfe P, Van Dyk J. *Medical Accelerators*. In: Van Dyck J, editor. *The modern technology of radiation oncology—a compendium for medical physicists and radiation oncologists*. Madison, (WI): Medical Physics Publishing; 1999.
35. Mills MD et al. Shielding consideration for an operating room based intraoperative electron radiotherapy unit. *Int J Radiat Oncol Biol Phys* 1990;18:1215–1221.
36. Meurk ML, Schonberg RG, Haynes G, Vaeth JM. The development of a small, economic mobile unit for intraoperative electron beam therapy. *Am J Clin Oncol* 1993;16: 459–464.
37. Mills MD et al. Commissioning of a mobile electron accelerator for intraoperative radiotherapy. *J App Clin Med Phys* 2001;2: 121–130.
38. Daves JL, Mills MD. Shielding assessment of a mobile electron accelerator for intraoperative radiotherapy. *J Appl Clin Med Phys* 2001;2:165–173.
39. Piermattei A et al. The saturation loss for plane parallel ionization chambers at high dose per pulse values. *Phys Med Biol* 2000;45:1869–1883.
40. Fantini M et al. IORT Novac7: A new linear accelerator for electron beam therapy. In: Vaeth JM, editor. *Intraoperative radiation therapy in the treatment of cancer*. Vol. 31. *Front. Radiation Therapy Oncology Basel*: Karger; 1997.
41. Tosi G, Ciocca M. IORT with mobile linacs: The Italian experience. *Oncologia* 2004;27:350–354.
42. Hogstrom KR et al. Design of metallic electron beam cones for an intraoperative therapy linear accelerator. *Int J Radiat Oncol Biol Phys* 1990;18:1223–1232.
43. AAPM, American Association of Physicists in Medicine, Task Group 25, Radiation Therapy Committee, Clinical electron beam dosimetry, Report of AAPM radiation therapy committee Task Group 25. *Med Phys* 1991;18:73–109.
44. Almond PR et al. AAPM's TG-51 protocol for clinical reference dosimetry of high-energy photon and electron beams. *Med Phys* 1999;26:1847–1870.
45. Kutcher GJ et al. *Comprehensive QA for radiation oncology*, Report of AAPM Radiation Therapy Committee Task Group 40. *Med Phys* 1994;21:581–618.
46. Davis MG, Nyerick CE, Horton JL, Hogstrom KR. Use of routine quality assurance procedures to detect the loss of a linear accelerator primary scattering foil. *Med Phys* 1996;23:521–522.
47. Palta JR, et al. Intraoperative electron beam radiation therapy, technique, dosimetry, and dose specification, report of Task Group 48 of the radiation therapy committee, American Association of Physicists in Medicine. *Int J Radiat Oncol Biol Phys* 1995;33:725–746.
48. Beddar AS et al. Intraoperative radiation therapy using mobile electron linear accelerators: Report of the AAPM Radiation Therapy Committee Task Group 72. To be published.
49. Hazle JD, Chu JCH, Kennedy P. Quality assurance for intraoperative electron radiotherapy clinical trials: Ionization chamber and mailable thermoluminescent dosimeter results. *Int J Radiat Oncol Biol Phys* 1992;24:559–563.
50. Beatty J, et al. A new miniature x-ray device for interstitial radiosurgery: dosimetry. *Med Phys* 1996;23:53–62.
51. Yanch JC, Harte KJ. Monte Carlo simulation of a miniature radiosurgery x-ray tube using the ITS 3.0 coupled electron-photon transport code. *Med Phys* 1996;23:1551–1558.
52. Beddar SA. Dose delivery for HDR-IORT using curved HAM applicators. IVth International meeting of the International Society of Intraoperative Radiation Therapy. Miami, (FL): Mar 17–19, 2005.
53. Raina S et al. Quantifying IOHDR brachytherapy underdosage resulting from an incomplete scatter environment. *Int J Radiat Oncol Biol Phys* 2005;61:1582–1586.
54. Harrison LB, Cohen AM, Enker WE. High-dose-rate intraoperative irradiation (HDR-IORT): Technical factors. In: Gunderson LL, Willett CG, Harrison LB, Calvo FA, editors. *Intraoperative Irradiation: Techniques and Results*. Totowa, (NJ): Humana Press; 1999.
55. Gunderson LL et al. Locally advanced primary colorectal cancer: Intraoperative electron and external beam irradiation ± 5-FU. *Int J Radiat Oncol Biol Phys* 1997; 37:601–614.
56. Gieschen HL et al. Electron or Orthovoltage IORT for retroperitoneal sarcomas. In: Gunderson LL, Willett CG, Harrison LB, Calvo FA, editors. *Intraoperative Irradiation: techniques and results*. Totowa, (NJ): Humana Press; 1999.

See also RADIOSURGERY, STEREOTACTIC; X-RAY THERAPY EQUIPMENT, LOW AND MEDIUM ENERGY.

RADIO THERAPY, THREE-DIMENSIONAL CONFORMAL

GEORGE STARKSCHALL
The University of Texas
Austin, Texas

INTRODUCTION

The goal of radiation therapy is to deliver a sufficient dose of radiation to a tumor site to control the disease while keeping doses to uninvolved tissue within tolerable limits. In order to achieve this goal, radiation beams are typically directed to the tumor target from several directions. Consequently, multiple beams irradiate the target, but only a few beams irradiate the uninvolved tissue surrounding the target, thus reducing the dose to the uninvolved tissue. The portals used to define the radiation fields are carefully shaped to fit the target in order to further reduce the amount of radiation reaching the tissue surrounding the target. In three-dimensional conformal radiotherapy (3DCRT), the radiation dose distribution is designed to conform to the contours of a target volume. The method by which the beam geometries and treatment portals of 3DCRT are designed differentiates 3DCRT from earlier methods of radiation treatment planning and delivery.

Prior to the 3DCRT era (mid-1990s), radiation treatment planning was based on a small number of computed tomography (CT) images of the patient and radiographic images from a simulator, a diagnostic X-ray machine whose geometries simulated the geometries of the radiation treatment machine. Beam configurations were determined based on a CT image of the patient in a single transverse plane, typically the plane containing the central axes of the radiation beams. Treatment portals were determined from the simulation radiographs; the design of these portals was based primarily on bony landmarks that were visible on the radiographs. Radiation doses were computed in the plane containing the beam central axes, and plans were evaluated on the basis of information displayed in that plane (1).

With the development of fast CT scanners and fast treatment planning computers with inexpensive memory, radiation therapy moved into the 3DCRT era. Whereas in the past, radiation treatments were planned and evaluated on the basis of information in a single or limited number of planes, radiation treatment planning is now based on patient volume. Patient information is acquired over a volume of the patient, treatment portals are designed to irradiate a target volume, doses are calculated in volumes of tissue, and treatment plans are evaluated based on dose-volume considerations. The 3DCRT process currently used in contemporary radiation therapy consists of the following steps: (1) accurately imaging the tumor and normal tissue in three dimensions, (2) precisely defining the target volumes, (3) optimizing beam geometries and beam weights, (4) translating the treatment plan to produce accurate delivery of radiation according to the treatment plan, and (5) verifying the accuracy of the delivery of radiation.

IMAGING FOR RADIATION ONCOLOGY

In the current state of practice in 3DCRT, treatment planning is based on virtual (CT) simulation. A 3D CT

image data set of the patient is acquired with the patient in the treatment position. Patients are typically positioned in an immobilization device, and marks are placed on the patient's surface to assist in maintaining setup reproducibility. The desired setup accuracy may vary from submillimeter accuracy for treatments in the head, where the tumor may lie very close to highly critical, radiosensitive, normal anatomic structures, to accuracies of 0.5–1 cm in areas of the thorax or abdomen.

Various types of CT scanners are used to acquire patient information. The third generation CT scanner consists of a single radiation source emitting X rays in a fan-shaped radiation pattern, which is detected by a single array of detectors. The gantry containing the radiation source makes a single (whole or partial) rotation around the patient with the detectors on the opposite side of the patient. Following acquisition of projection data from a single transverse slice, the table is indexed and another slice is acquired. This procedure continues until the entire region of interest is scanned.

Technologic developments in CT image acquisition include helical, or spiral, CT. This technology combines gantry rotation with table translation so that the path of the radiation source with respect to the patient forms a spiral trajectory. Image acquisition times using helical CT are significantly faster than times using third generation CT. Even faster is a new technology in which a multislice X-ray detector is used. Typically, such detectors can acquire from 4 to 16 image slices simultaneously. Multislice helical CT has the potential for increased axial resolution as well as scan times that can be as short as 3–5 s (2).

The ability to increase the axial resolution of CT image data sets brings up the issue as to the desired axial resolution for planning 3DCRT. Prior to the 3DCRT era, CT scans were typically acquired at planar separations of ~10 mm. This axial resolution should be compared with the typical picture element (pixel) dimensions in each plane of $<1 \text{ mm}^2$. The ideal axial resolution would be comparable with the planar resolution; however, a submillimeter axial resolution would result in a very large number of CT images. Because the contours of tumors as well as those of some normal anatomic structures have to be manually delineated on the CT images for treatment-planning purposes, requiring such delineation on so many CT images would result in an extraordinarily time-consuming, labor-intensive task. Consequently, axial resolutions presently used in 3DCRT represent a compromise and are typically in the vicinity of 3 mm.

A key component in CT simulation is the production and display of digitally reconstructed radiographs (DRR). A DRR is a radiographic image that models the transmission of an X-ray beam through the patient (3). A DRR is obtained by tracing the path of X rays from the location of the radiation source through the 3D CT patient image data set to a plane beyond the data set and calculating the X-ray transmission through the CT data set. Generating a DRR offers several advantages over obtaining radiographic images from a conventional simulator. Whereas tumors are sometime difficult to visualize on a conventional radiograph, they can often be visualized more easily on a CT image. The contours delineating the tumor can be drawn on

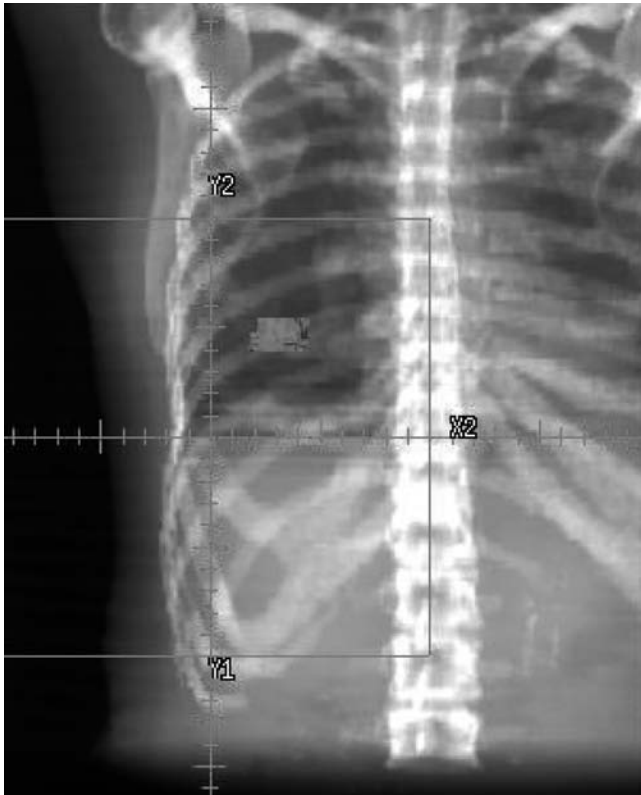


Figure 1. A digitally reconstructed radiograph illustrating a tumor in a patient's right lung.

the CT image data set and projected onto the DRR. Treatment portals are then designed based on the projected contours. The design of a treatment portal based on the volumetric extent of the tumor is a key component of the 3DCRT process. In addition, many soft-tissue anatomic structures may also be difficult to visualize on a conven-

tional radiograph. Contours delineating these structures can be drawn on the CT image data set and projected onto the DRR as well. Display of these projected contours can assist the treatment planner in selecting radiation-beam geometries that will irradiate the tumor while avoiding critical anatomic structures. Finally, the DRR can be compared with a radiographic image of the treatment field acquired on the treatment machine to verify that the patient is correctly set up on the treatment machine. Figure 1 illustrates a DRR with a target volume drawn on it.

Computed tomography image data sets are not the only images used in planning the 3DCRT. They are the images of choice for performing actual radiation-dose calculations because the values of the CT volume elements (voxels) can best be correlated to the extent of interactions of the X-ray beam with the patient. However, other imaging modalities may be more effective than CT in imaging the tumor. For example, magnetic resonance (MR) imaging is often more reliable for imaging soft tissue, especially for planning radiation treatment of brain tumors. Positron emission tomography (PET) is a very useful modality for imaging tumor metabolism and often provides additional information about the presence and extent of lung tumors. Figure 2 illustrates a CT image and a PET image illustrating the differences in the information provided by the two imaging modalities.

A major issue encountered in the use of multimodality imaging for 3DCRT planning is that of image registration, that is, the geometric correlation of information from the various imaging modalities with that of the CT image data set. The coordinate systems used in the acquisition of images from various modalities are typically different, requiring some sort of registration so that information delineated on one image data set is correctly displayed on the treatment-planning CT data set. Moreover, MR images are sometimes subject to geometric distortion,



Figure 2. Coronal CT and PET images of a patient with lung cancer.



Figure 3. A combination PET/CT scanner.

especially near the periphery of such images, making the registration problem more complex. The PET and CT images are very difficult to register manually, as the two modalities image very different properties of the patient. Registration of PET and CT image data sets is sometimes accomplished by registering the CT image data set with a transmission scan taken on the PET scanner and used to make corrections in the PET scan for patient attenuation. A new device, in which a PET and a CT scanner are mounted together, has recently been introduced. An example of such a device is shown in Fig. 3. Using this device, the patient first undergoes a CT scan. With the patient immobilized, the support table is translated into the PET scanner, and a PET scan is obtained. In this manner, the geometries of the PET scan and CT scan differ only by a known table translation, making image registration a simple process.

DEFINITION OF TARGET VOLUMES

Because dose prescriptions used in 3DCRT are based on irradiating volumes of tissue, more uniformity in terminology has been required in the specification of dose prescriptions. The International Commission on Radiation Units (ICRU) has developed a series of documents that describe a methodology for specification of target volumes (4,5). The ICRU definitions constitute a methodology for reporting radiation treatments in an unambiguous manner.

The ICRU defines the gross tumor volume (GTV) to be the “gross demonstrable extent and location of the malignant growth” (4). The GTV consists of the primary tumor, involved lymph nodes, and metastatic disease that can be visualized. The specific method of visualization is not specified; the GTV might be identified and delineated on clinical examination, radiographs, CT images, MR images, or any other visualization method. Moreover, the extent of the GTV may be different for different examinations, depending on the method of visualization. One of the aims of the radiation therapy is to deliver a tumoricidal dose to the entire GTV. The patient may have several GTVs depending on the nature of the malignant disease, and each GTV may have a different therapeutic aim. In some

cases, in particular after surgical intervention, the GTV may not exist.

In addition to clinically demonstrable disease, the patient is likely to have subclinical disease that can only be visualized under pathologic examination. This subclinical disease must also be eliminated for the tumor to be controlled. The combination of GTV and subclinical disease constitutes the clinical target volume (CTV). The radiation oncologist establishes the margin that defines the CTV on the basis of clinical experience. In a few cases, such as nonsmall cell lung tumors, pathology studies have helped the radiation oncologist define the CTV margin. In the case of surgical intervention, a CTV might exist without a GTV. The existence of GTVs and CTVs are based on general oncologic principles and are independent of any therapeutic approach. The GTVs and CTVs can be identified and delineated prior to the selection of the treatment modality and the treatment-planning procedures.

Treatment planning for 3DCRT is typically based on information obtained from a CT image data set acquired a few days before treatment is scheduled to begin. An inherent assumption in basing the treatment plan on the CT data set is that the data set accurately models the patient during the entire course of radiation therapy. In reality, the patient moves, exhibiting intrafractional motion such as respiratory and cardiac motion, as well as interfractional motion such as bladder and rectum filling and tumor regression. In order to account for motion, an internal target volume (ITV) is defined. The ITV is the CTV with an internal margin (IM). In many cases, the ITV is assumed to be the CTV with a uniform, isotropic IM, established from clinical experience; but in some cases, the ITV is explicitly defined from multiple CT image acquisition.

A final margin needs to be added to the ITV to account for the fact that there is some degree of nonreproducibility in the day-to-day setup of the patient for radiation treatment. Consequently, a planning target volume (PTV) is defined to include the ITV along with a setup margin (SM) to account for these setup uncertainties. The PTV is thus defined to be a region that, if irradiated to the prescription dose, makes sure that the entire CTV will be irradiated to the prescription dose, regardless of internal motion or setup uncertainty. The extent of the SM is based on the device used to immobilize the patient. When invasive devices are used, the SM may be reduced to <1 mm; typically, however, the SM is in the range 0.5–1.0 cm. Figure 4 illustrates an axial CT image of a patient's lung with the GTV, CTV, and PTV identified.

In 3DCRT, a DRR with the delineated PTV is normally displayed to the treatment planner. A radiation-treatment portal is established to include the entire PTV along with an additional 5–7 mm margin that accounts for the fact that the edge of the radiation beam is fuzzy due to the finite size of the radiation source as well as effects of scattered radiation. Thus, a GTV is expanded by a margin to account for microscopic disease to generate a CTV, which is expanded by a margin to account for motion to generate an ITV, which, in turn, is expanded by a margin for setup uncertainty to generate a PTV.

In addition to identifying and delineating target volumes, the 3DCRT planning process must also identify

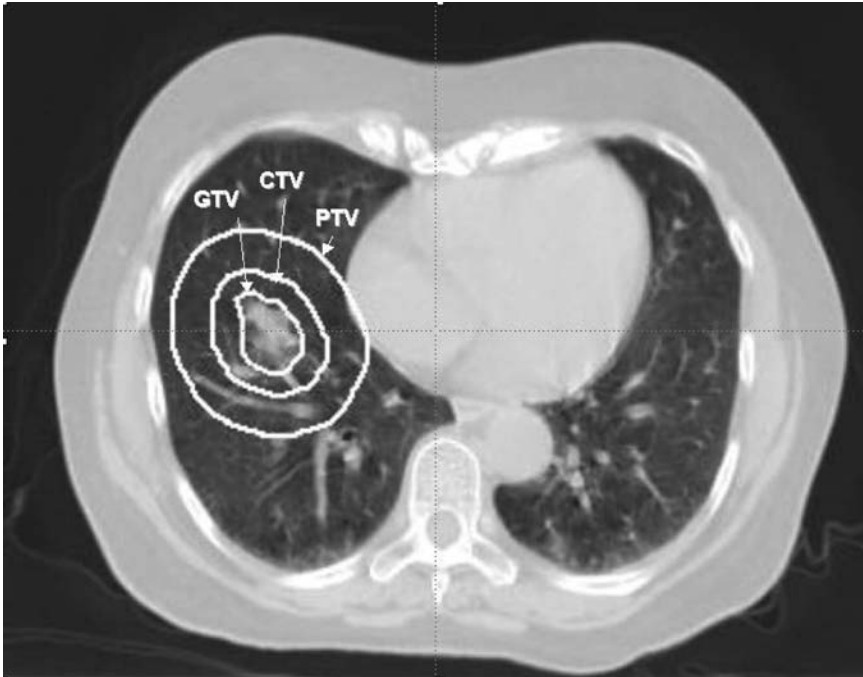


Figure 4. An axial CT image of a patient's thorax, illustrating a GTV, CTV, and PTV. The CTV is obtained by expanding the GTV by a uniform 0.8 cm margin, and the PTV is obtained by expanding the CTV by a uniform 1.5 cm margin.

and delineate organs at risk (OAR), which are defined to be “Normal tissues whose radiation sensitivity may significantly influence treatment planning and/or prescribed dose” (5). For example, in planning the radiation treatment of lung tumors, the OARs typically include the right and left lung, heart, esophagus, and spinal cord, whereas in planning the radiation treatment of prostate tumors, the OARs typically include the bladder, rectum, and femoral heads. Setup uncertainty and OAR motion are accounted for in defining the planning organ at risk volume (PRV), which contains the OAR along with an internal margin and a setup margin (5). Thus, in the planning of 3DCRT, the planner must determine one or more beam geometries and treatment portals that will adequately irradiate the PTV, while keeping doses to the various PRVs within tolerable limits.

DOSE PRESCRIPTION

Prior to the development of 3DCRT, radiation-dose prescriptions were generally based on point-dose methodology. A dose prescription might have read, “66 Gy to the isocenter”, where the isocenter was the point around which the gantry of the radiation machine rotated. Another prescription might have read, “72 Gy to the 95% isodose”, where the 95% isodose represented the locus of those points receiving 95% of the dose to isocenter. In both cases, the dose prescriptions described doses to a single point, typically the point at isocenter.

In 3DCRT, one is concerned with doses delivered to volumes, in particular, the target volumes. The goal of a 3DCRT treatment plan is to irradiate the entire CTV to the prescription dose. Because the CTV may move anywhere within the PTV, it would also be desirable for the entire PTV to be irradiated to the prescription dose as well. However, because the CTV does not lie in the entire PTV for the entire treatment, one may allow a small

amount of the PTV (typically 5%) to receive a dose less than the prescription dose without adversely compromising the goals of the radiation treatment, especially if doing so would significantly improve organ sparing. If <100% of the CTV receives the prescription dose, the radiation oncologist may have to address the potential consequences that could occur if tumor cells do not receive a tumoricidal dose. However, compromises in CTV dose may be necessary to avoid excessive morbidity. In many cases, the 3DCRT planning process represents a compromise between the dose needed for adequate tumor control and the dose that would cause unacceptable side effects of the radiation.

BEAM DETERMINATION

Determination of radiation beam geometries for 3DCRT is, to an initial approximation, a trial-and-error process based on class solutions. For example, a pair of parallel-opposed beams delivers approximately a uniform dose to the entire volume irradiated by the beams. Tumors, however, rarely extend completely through the patient; so in 3DCRT, more than two beams are likely to be used. Placing two more beams at right angles to a parallel-opposed pair of beams generates what traditionally has been called a “four-field box”. This configuration delivers ~50% of the target dose to regions that are irradiated by only two of the four beams. More complicated beam geometries are sometimes used based on the need to avoid irradiation of critical structures. Figure 5 illustrates a multiple-field beam configuration used to treat prostate cancer. Incorporating couch rotations resulting in beams whose axes are noncoplanar can sometimes be used to improve avoidance of critical structures. Wedge-shaped metallic filters placed in the beam to generate dose gradients can be used to compensate for dose nonuniformities generated from certain beam configurations.

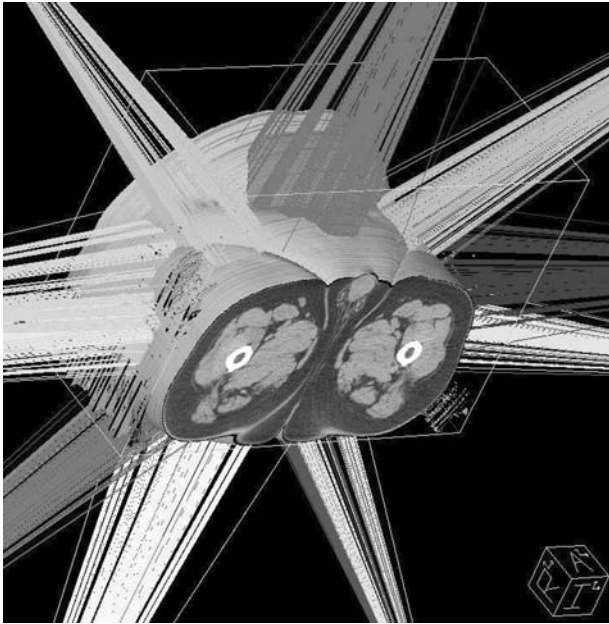


Figure 5. A multiple field beam configuration used to treat prostate cancer.

Treatment portals are determined for each beam geometry and designed to encompass the PTV along with a margin to allow for the fact that the beam does not end abruptly at the geometric edge but rather the dose distribution has a finite slope at the beam edge. Caused both by the fact that the radiation source is not a point source and that radiation can scatter from the irradiated region to the blocked region, this penumbra requires that the treatment portal surround the PTV with a margin of from 0.5 to 1 cm in order to deliver a high dose to the PTV. The treatment planning typically designs the treatment portal based on viewing the BEV projection of the PTV on a DRR.

In addition to optimizing the beam geometries, the planning often must also optimize the beam weights. Typically, this optimization consists of manual fine tuning of the beam weights to deliver a more uniform dose across the target volume, decrease the dose to specific critical structures, or both of these outcomes. However, if specific prescription goals are established, such as desired doses to target volumes and maximum allowed doses to critical structures, various computer algorithms can be used to optimize the beam weights. If additional conformality of the dose distribution to the target volume is desired, techniques exist for the planning and delivery of intensity-modulated radiation therapy (IMRT). This technique is particularly useful if the target volume is concave.

DOSE CALCULATION

In most current radiation treatment planning systems, X-ray dose distributions are calculated using a convolution algorithm (6,7). In this algorithm, the radiation dose is obtained by convolving an incident fluence distribution with a dose-spread array. The dose-spread array represents the distribution of radiation dose around a point in

which a unit amount of X-ray fluence is seen. Because the dose-spread array is dependent solely on the X-ray energy and distance from the point of fluence deposition, dose-spread arrays, one for each energy, need only be computed once, and then stored in the treatment-planning computer. The in-air fluence exiting the linear accelerator is determined once for each linear accelerator and is modeled by a set of parameters. The X-ray fluence is calculated by propagating the in-air fluence through the CT model of the patient, attenuating the fluence based on exponential attenuation, characteristic of X-ray propagation through matter.

More accurate calculations of X-ray dose distributions in the patient can be achieved by more accurate modeling of the interactions that X rays undergo with the patient using Monte Carlo techniques (8). Originally limited in applicability by limitations in computing power, newer high speed processors are making such calculations clinically feasible.

CONFORMAL PARTICLE THERAPY

The 3DCRT can also be achieved using beams of charged-particle radiation, such as electrons or protons. In some respect, achieving 3DCRT is easier with charged-particle beams than it is with X-ray beams. One feature that differentiates the dose distribution from charged-particle beams from that of X-ray beams is that the dose distribution from charged-particle beams, at least to a first approximation, is constant from the patient surface to a depth equal to the range of the charged particle, and then goes rapidly to zero. This distribution results from the charged particle's property of depositing a fixed amount of energy per interaction until the particle energy is exhausted; beyond that depth, no energy is deposited. Consequently, the lateral extent of a charged-particle beam can be shaped based on the BEV projection of the PTV, just as are X-ray beams, but the depth of penetration of a charged-particle beam can be made to track the distal surface of the PTV by spatially modulating the energy of the beam, which shifts the range of the charged particles. A simple method for range-shifting involves placing a sufficient amount of attenuating material in the path of the beam to reduce the local energy of the beam so that its range is slightly greater than the depth of the distal surface of the PTV. Alternatively, one could modify the spatial distribution of the energy of the beam by modifications in the beam transport in the head of the particle accelerator.

Particle therapy, on the other hand, has some limitations. The margins of electron beam dose distributions are not as sharp as those for X-ray beams because of multiple scattering. In addition, heterogeneities and surface irregularities cause significant perturbations in the electron beam dose distributions. Widespread use of proton beams may be limited because of the significantly higher cost of production of therapeutically useful beams.

RADIATION DELIVERY

Once a 3DCRT treatment plan has been developed, it is necessary to translate the treatment plan accurately to the

radiation-delivery system. Although this can be done manually, the complexity of modern treatment plans demands that plan information be transferred automatically from the treatment-planning computer to a linear accelerator. Because of the variety of treatment-planning systems and linear accelerators, significant effort has been expended to standardize the exchange of information among the imaging systems, treatment planning computers, and treatment delivery systems.

VERIFICATION OF DELIVERY

The final step in 3DCRT is that of quality assurance, verification that the radiation beams delivered to the patient are consistent with those planned for delivery. All components of the planning and delivery process need to be reviewed as part of a quality assurance process, including the validity of the treatment plan, the accuracy of the machine settings, the validity of the transfer of data from the planning system to the radiation machine, the accuracy of the portal for delivery of the radiation beam, the accuracy of the patient setup, and the accuracy of the radiation delivery.

The first step in verification is that of verifying that the treatment plan accurately reflects the radiation dose that is actually designed to be delivered to the patient. This is achieved by implementation of a quality assurance program for the treatment planning system. In such a program, one verifies that the treatment-planning system accurately depicts the patient images, contours, and beams that are provided as input into the dose calculations, and that the beam models are accurate, so that the doses calculated by the treatment planning system are accurate (9).

The next component of verification is that the machine setting determined by the treatment-planning system will deliver the correct dose to the target. Accuracy of machine setting is verified initially when the beam model is first commissioned and introduced into the treatment-planning system and is verified for each dose calculation by and independent calculation of radiation doses to individual points within the patient.

Next, it is necessary to verify that the machine settings, such as beam geometries, and monitor units are accurately transferred from the treatment-planning system to the machine. In many radiation oncology clinics, a record and verify (R&V) system is used to record the machine settings used to deliver the radiation beams to the patient, ensuring that machine settings lie within clinically accepted tolerances of the values established on the treatment plan. A component of a comprehensive quality assurance program verifies that the machine settings determined by the treatment-planning system are accurately transferred into the R&V system.

Another key step is to verify that the treatment field actually delivered to the patient is identical to the treatment portal as determined on the treatment plan. In order to accomplish this task, a radiographic image of the delivered treatment portal is acquired. The portal image is then compared to a DRR of the treatment field extracted from

the treatment-planning system. Ideally, the geometric relationship between the tumor target and the treatment portal would be compared, but tumors are rarely visible on portal images. Consequently, bony landmarks in the vicinity of the tumors are typically used as surrogates for the tumors. Originally, the portal image was recorded on radiographic film, but many radiation oncology clinics are acquiring this information using an electronic portal imaging device (EPID). The use of digital images produced by EPIDs allows for rapid display of the portal images as well as off-site review of the portal images. Figure 6 illustrates a DRR and a portal image used to verify the accuracy of the treatment portal.

One important issue regarding the use of portal images for treatment portal verification is the substantial difference in image quality between simulation and portal images. Simulation images are typically acquired in the energy range of 50–100 keV. In this energy range, the primary interaction between the incident X rays and the patient results in absorption of the incident X rays. Consequently, the only X rays that reach the detector are those that are not absorbed in the patient. Moreover, in this energy range, bone absorbs significantly more radiation than soft tissue; hence, the contrast between bony anatomy and the surrounding soft tissue anatomy results in clear, sharp radiographic images. Portal images are acquired at much higher X-ray energies, typically several megaelectronvolts, the energies that are used in radiation therapy. In this energy range, the difference in absorption of radiation between bone and soft tissue is significantly less, resulting in lower contrast. Moreover, the presence of scattered radiation resulting from X-ray interactions with the patient at these energies results in a significant amount of radiation reaching the detector that gives no indication of the point of origin, resulting in a noisy image.

In addition to verifying the geometry of the treatment portal, acquisition of a portal image is one of several techniques used to verify that the patient has been set up in a reproducible manner for each treatment. Because the treatment field is often positioned adjacent to critical uninvolved anatomic structures, accurate and reproducible patient positioning is essential so that these uninvolved anatomic structures do not get unnecessarily irradiated. Perhaps the simplest method of position verification is through the use of external markings. Lasers in the walls and ceilings are all directed to a particular point in space, the machine isocenter, where the axis of gantry rotation coincides with the axis of collimator rotation. In 3DCRT, the patient is often positioned so that the isocenter lies in the approximate center of the tumor volume. The points at which the lasers intersect the patient surface are marked and used on a daily basis to assist in ensuring that the patient is set up reproducibly.

In conjunction with external markings, patients are often placed in immobilization devices to assist in reproducible positioning. Immobilization devices have many forms. For example, invasive devices, such as stereotactic head frames, which are screwed into the patient's skull, can achieve submillimeter reproducibility and are necessary when the geometric tolerances are very small. When

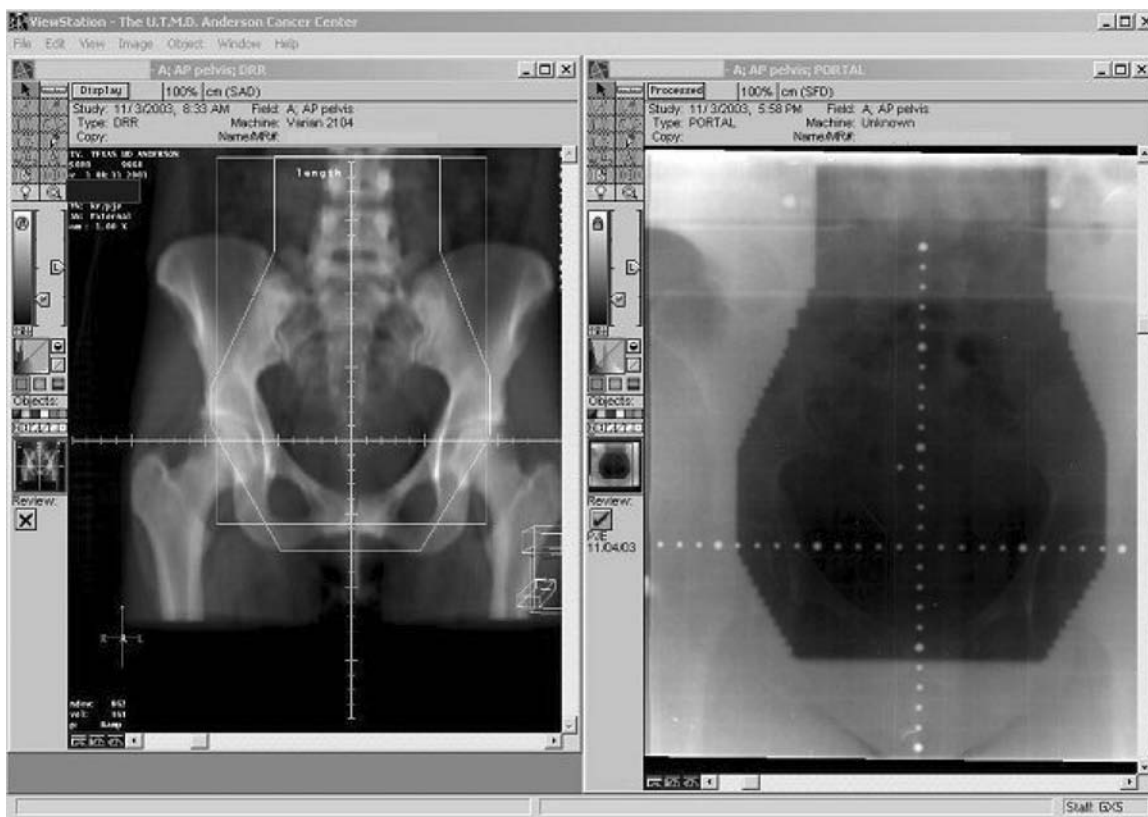


Figure 6. A DRR of an anterior pelvic treatment field compared with a portal image of the same field used to verify that the patient is set up accurately and that the location and extent of the irradiated region is as per plan.

submillimeter tolerance is not as crucial, less invasive devices can be used, including thermoplastic masks, molds, and vacuum bags.

Portal-imaging studies have demonstrated differences between true setup errors and random uncertainties in patient setup, which are a consequence of the degree of patient immobilization, and have indicated guidelines as to when and how to correct for setup inaccuracies. In the treatment of some cancer sites, for example, the prostate, daily variations in patient anatomy have led to the development of more sophisticated methods of setup verification. Differences in bladder and rectal filling on a daily basis combined with the tight treatment margins typical of 3DCRT may cause one or more radiation beams to miss part of the target volume. One technique that enables correction for daily anatomic variations is to scan the patient just prior to treatment using an ultrasound technique. Figure 7 illustrates an example of the unit used in such an ultrasound process. The outlines of the target volumes and anatomic structures, which are extracted from the radiation treatment plan, are superimposed on the ultrasound scan, and the patient is moved so that the image of the target volume on the scan is superimposed on the outline of the target volume on the treatment plan, thus ensuring accurate delivery of radiation to the target volume. Figure 8 illustrates the use of the ultrasound images in realigning a patient.

Another promising technique involves placing a CT scanner in the treatment room in a configuration that

allows both imaging and treatment on the same patient couch. The patient is set up in the treatment position and then scanned. It is then possible to compare the CT image data set thus acquired with the data set from which the treatment plan is based, allowing for the patient to be repositioned or even replanned. Rather than allowing the patient table to move through the CT scanner, as is conventionally done, this device moves the CT scanner while the patient remains stationary, hence the device is referred to as “CT-on-rails”.

In addition to verifying that the radiation beams accurately irradiate the target volume and spare surrounding normal-tissue anatomy, it is essential to verify that the radiation dose actually delivered to the target is identical to the radiation dose prescribed on the treatment plan. This is achieved by means of a thorough quality assurance system, including a regular schedule of well-defined daily, monthly, and annual evaluations of the output of the radiation machine, including accurate measurements of the magnitude of the radiation emitted from the linear accelerator as well as the energy of the radiation (10). In addition, techniques exist for measurement of doses to accessible parts of the patient during treatment, using small radiation detectors such as thermoluminescent dosimeters (TLD) or diode detectors, which can be placed either on external surfaces of the patient or in accessible cavities within the patient. Current studies are underway to assess the safety and accuracy of radiation detectors that can be implanted into the tumor inside the patient.

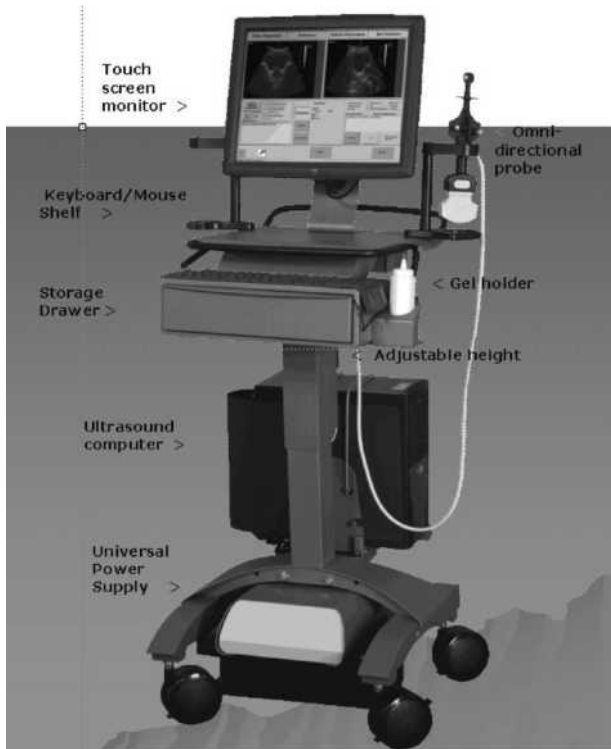


Figure 7. An ultrasound imaging device used to verify reproducibility of patient position for radiation treatment of the prostate. (Figure courtesy of NOMOS Radiation Oncology division of North American Scientific.)

CLINICAL CONSEQUENCES OF 3DCRT

Finally, it is necessary to determine whether 3DCRT represents an improvement in the treatment of cancer.

Treatment-planning studies can readily demonstrate that 3DCRT allows higher doses to be delivered to target volumes than was possible with conventional radiation therapy. Higher radiation doses have been shown to lead to increased rates of tumor control. An early study comparing 3DCRT to conventional radiation therapy in the treatment of prostate cancer was conducted by Pollack (11) in the mid-1990s. In this study, patients were randomly selected to receive radiation therapy based on conventional treatment planning or a dose ~10% higher based on 3DCRT. Pollack demonstrated a significant increase in freedom from failure for intermediate-to-high risk patients, but with some increase in rectal toxicity.

CONCLUSIONS

Three-dimensional conformal radiotherapy is a radiation treatment planning and delivery technique in which the design of the radiation-treatment portals is based on images of the tumor target. Planning is based on high resolution CT images that explicitly display the tumor as well as provide a mathematical model for the patient. Additional imaging modalities are often used to define the extent of tumor involvement more clearly. Sophisticated algorithms are used to calculate the magnitude of radiation dose deposited in the patient, allowing the assessment of the potential for tumor control as well as toxicity of the treatment plan. In order to make certain that the radiation is delivered to the tumor and not the surrounding uninvolved tissue, as indicated in the treatment plan, extensive quality assurance procedures are required that verify the accuracy of the geometry as well as the radiation dosimetry.

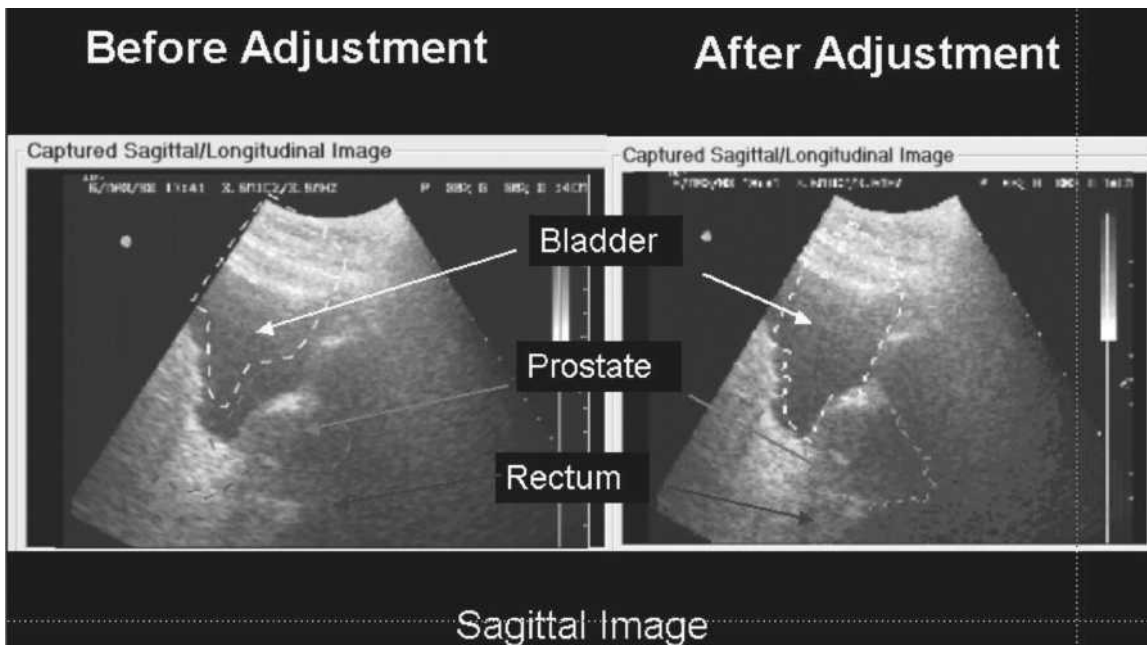


Figure 8. Sagittal ultrasound images of a patient before and after realignment of the patient. (Figure courtesy of Lei Dong, Ph.D., U. T. M. D. Anderson Cancer Center.)

BIBLIOGRAPHY

1. Starkschall G. What is 3-D radiotherapy treatment planning. In: Purdy JA, Starkschall G, editors. *A Practical Guide to 3-D Planning and Conformal Radiation Therapy*. Madison, WI: Advanced Medical Publishing; 1999. p 1–16.
2. Bushberg JT, Seibert JA, Leidholdt EM, Boone JM. *The Essential Physics of Medical Imaging*. Philadelphia, PA: Lippincott Williams & Wilkins; 2003. Chapt. 13.
3. Goitein M, Abrams M, Rowell D, et al. Multi-dimensional treatment planning: II Beam's eye-view, back projection, and projection through CT sections. *Int J Radiat Oncol Biol Phys* 1983;9:789–797.
4. International Commission on Radiation Units and Measurements (ICRU) Report 50: "Prescribing, Recording, and Reporting Photon Beam Therapy". ICRU (1993) Bethesda MD.
5. International Commission on Radiation Units and Measurements (ICRU) Report 62: "Prescribing, Recording, and Reporting Photon Beam Therapy (Supplement to ICRU Report 50)". ICRU (1999) Bethesda MD.
6. Mackie TR, Scrimger JW, Battista JJ. A convolution method of calculating dose for 15 MV x-rays. *Med Phys* 1985;12:188–196.
7. Boyer AL, Mok EC. A photon dose distribution employing convolution calculations. *Med Phys* 1985;12:169–177.
8. Ma CM, Mok EC, Kapur A, Pawlicki T, Findley D, Brain S, Forster K, Boyer AL. Clinical implementation of a Monte Carlo treatment planning system. *Med Phys* 1999;26:2133–2143.
9. Fraass BA, Doppke KP, Hunt MA, et al. Task Group 53: Quality assurance for clinical radiotherapy treatment planning. *Med Phys* 1998;25:1773–1829.
10. Kutcher GJ, Coia L, Gillin M, et al. Comprehensive QA for radiation oncology: Report of AAPM Radiation Therapy Committee Task Group 40. *Med Phys* 1994;21:581–618.
11. Pollack A, Zagars GK, Starkschall G, et al. Prostate cancer radiation dose response: Results of the M.D. Anderson phase III randomized trial. *Int J Radiat Oncol Biol Phys* 2002;53:1097–1105.

Further Reading

Khan F. *The Physics of Radiation Therapy*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2003. especially Chapt. 19.

See also RADIATION DOSIMETRY FOR ONCOLOGY.

RADIOTHERAPY TREATMENT PLANNING,
OPTIMIZATION OF

ANTHONY B. WOLBARST
Georgetown Medical School
Washington, DC

LEE CHIN
Harvard Medical School
Boston, Massachusetts

PAVEL STAVREV
Cross Cancer Institute
Edmonton, Alberta, Canada

INTRODUCTION

Treatment of solid tumors typically involves some combination of surgery, chemotherapy, and radiation therapy. With any of these approaches—indeed, as with so much of

medicine—the objective is to eradicate or control the disease without producing unacceptable side effects. Radiation therapy, in particular, is used for one-half of all cancer patients in the United States, and can frequently contribute to the cure of a malignant tumor, or at least to a significant improvement in quality of life.

Intense localized irradiation of the region of a solid tumor can both incapacitate or kill cancer cells directly, and strangle them through damage to tumor-bed microvasculature. High doses of ionizing radiation will kill not only the cells of tumors, however, but also those of healthy tissues. So a successful treatment must target the tumor accurately, with little radiation ending up elsewhere; it would be of questionable benefit to eradicate a tumor of the esophagus if this also led to severe functional complications in the spinal cord.

Radiation therapy is therefore preceded by a careful, and sometimes extensive, planning process, and a search for a best treatment plan.

This article is based on material found in Ref. 1.

ASSESSING BENEFITS AND RISKS

Deciding upon a medical strategy, like most other activities in life, involves an optimization process. In some situations, the proper approach is so obvious (Take two aspirin and call me in the morning!) that one hardly gives thought to the matter. In the treatment of cancer with radiation, however, the decisions are usually much more difficult. It may not even be clear, in fact, how the various possible criteria should be balanced in choosing the best treatment.

Ultimately, the desired result of radiotherapy is the complete disappearance of the disease, or at least long-term palliation, without the onset of unacceptable side effects. For some types of cancer, such as carcinomas of the skin, the malignant cells can be exposed directly to radiation, with limited risk to any critical organs. Also, they may be inherently more susceptible to radiation damage than are the adjacent healthy tissues. The reasons for this, while not completely understood, involve the tendency of tumor cells to divide faster than healthy ones—if both kinds are irradiated, a greater fraction of the cancer cells will be undergoing division, during which time they are more vulnerable to radiation damage, and a greater fraction of them will die. They may also have a diminished ability to repair radiation damage. In any case, complete cure without complication is frequently achievable.

For diseases in which the neoplastic (cancerous) cells are relatively radioresistant or less accessible to irradiation, however, the odds are greater that a curative dose of radiation would cause extensive damage to surrounding healthy tissues. The amount of radiation to be delivered to the tumor region and the means of physically delivering it must then be determined by assessing the likelihood of effecting a cure or prolonged survival against that of inducing unacceptable complications.

Thus any new patient presents the radiation oncologist, in theory, with a fundamental, three-part problem. For

every reasonable treatment strategy (including the various reasonable dose distributions and their timing), the physician has to

1. Estimate the probabilities of eradicating the disease, on the one hand, and of inducing complications of a range of severities, including death, on the other; such numbers are, in general, not well known.
2. Assuming that they have somehow arrived at plausible values for these probabilities, the physician and staff must then assign a weight, or measure of the seriousness or undesirableness, perhaps relative to death by the disease, to all the possible complications; such an assignation cannot help but be highly subjective.
3. Finally, they must combine the probability and undesirability information for every strategy in such a way as to arrive at a medically significant overall figure of its merit; the treatment plan with the best such score would then indicate the regimen of choice. Unfortunately, there is no obvious or simple way of doing this.

Thus the physician and patient must together decide how much they would be willing to compromise the quality of life in an attempt to preserve it. Unfortunately, the currently available data on the probable outcomes of the different therapeutic approaches are both sparse and crude. The response of liver to non-uniform irradiation has been studied, for example, but this kind of work is still in an early stage (2,3). And even if the essential statistical information were available, there is no unequivocal way of weighting it with the necessary value judgments on harm in a systematic treatment-selection process. In other words, there is considerable art in treatment planning, in addition to the science.

A PRELIMINARY EXAMPLE

Let us begin by considering some of the difficulties to be found in even the simplest of idealized cases.

A hypothetical patient presents with pockets of lethal disease within a single vital organ, and will die if either the disease spreads or the organ fails. Suppose that it is possible to irradiate this tumor-bearing organ uniformly and without risk to other tissues. And finally, assume that the dose-response characteristics both of the tumor and of the organ itself have separately been determined, Fig. 1.

The probability that the patient will survive the ravages of the disease, and that the tumor will be controlled, $S_{\text{tumor}}(D)$, will improve with increasing dose, D , until every viable tumor stem cell has been eliminated. On the other hand, the odds that the patient will escape unacceptable (e.g., life-threatening) complications to the organ, $S_{\text{organ}}(D)$, will be a decreasing sigmoidal function of D . (Either way, S is used to denote the probability of patient survival, or well being.) The parameters $S_{\text{tumor}}(D)$ and $S_{\text{organ}}(D)$ are also known as the tumor control probability (TCP) and normal tissue complication probability (NTCP), respectively, and by other names as well.

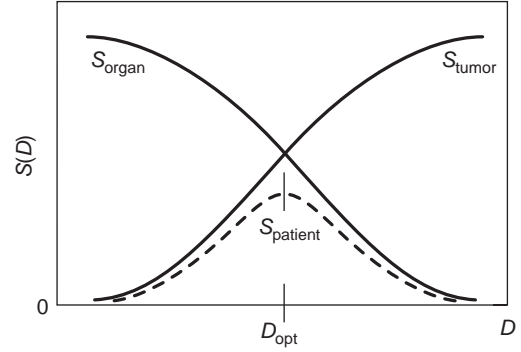


Figure 1. For this idealized, hypothetical example, the tumor is spread throughout a single critical organ, and only that organ is irradiated. The probability of the patient surviving the ravages of the disease, $S_{\text{tumor}}(D)$, increases with dose, D , to the tumor volume. Unfortunately, the likelihood of escaping an unacceptable complication that might arise from the treatment, $S_{\text{organ}}(D)$, decreases with D . The compound odds of both eradicating or controlling the disease and avoiding the complications is then $S_{\text{tumor}}(D) \times S_{\text{organ}}(D)$. This overall survival probability, $S_{\text{patient}}(D)$, reaches its maximum for some optimal value of the dose, D_{opt} .

At any level of uniform dose, D , the overall likelihood of uncomplicated survival, $S_{\text{patient}}(D)$, is then given by the law of multiplication of probabilities for independent events,

$$S_{\text{patient}}(D) = S_{\text{tumor}}(D) \times S_{\text{organ}}(D) \quad (1)$$

shown as the dashed line. At low doses, failure of the treatment through continuing tumor growth is likely. At high doses, an unacceptable complication in the organ is liable to occur. But between these two extremes, the patient may well survive both the disease and the effects of the treatment. At some particular optimal value of the dose, D_{opt} , the probability of that happening is greatest, and $S_{\text{patient}}(D)$ passes through a maximum.

Figure 1 illustrates a case in which there is a fairly broad range of dose over which the probability of uncomplicated cure is high, a situation said to be of good therapeutic ratio. When this is not the case, radiation treatment is a less promising option.

If the sole consideration in selecting D were the maximization of the probability of complication-free survival, then the appropriate value would be D_{opt} . This theoretical criterion, however, is often difficult to apply in practice, since it is generally accepted that the incidence of severe complications must be kept to a "tolerable" level. Hence, the dose actually chosen for treatment would normally be somewhat $< D_{\text{opt}}$, and the odds of its failure would likely be correspondingly greater.

This example is highly idealized, of course, but it can serve as a backdrop against which to contrast some of the issues that arise in real cases:

1. It is frequently difficult to determine in 3D, even with computerized tomography (CT), positron emission tomography (PET), or magnetic resonance imaging (MRI), the location and extent of a tumor; and it is virtually impossible to track down nearby microscopic clusters of neoplastic cells which, if untreated,

might themselves grow into tumors. The temptation is therefore to treat the region thought to harbor disease with generous spatial margins of healthy tissue; the risk of inducing complications, on the other hand, will increase with the volume of healthy tissue taken to high dose.

2. Commonly in clinical practice, one standard (but not necessarily optimal) treatment objective is to irradiate the tumor as uniformly as possible (4–10). But it would be better to sculpt a unique, patient-specific distribution of dose, non-uniform within the region of suspected disease, that takes into account the spatial variations in the density and responsiveness of tumor cells, the uncertainty in the physician's estimate of their whereabouts, and the sensitivity and criticality of the infiltrated and surrounding healthy tissues (4,5,9). The kinds of information necessary are not readily obtainable.
3. Equation 1 is meaningful only when any possible complication is of a seriousness comparable to that of death. The various complications that can arise in practice, however, differ greatly in severity. (And who is best able to assess the severity? The patient? The doctor? The insurance companies?) In any case, the gravity of any one acute or chronic complication (not only its probability of occurrence) is likely to vary with dose.
4. Equation 1 is valid only so long as $S_{\text{tumor}}(D)$ and $S_{\text{organ}}(D)$ are completely independent of one another; if the tumor reacted to intense irradiation by emitting a toxin or other biochemical that compromises the resistance of healthy organ tissue (or of the entire person) to radiation damage, for example, then equation 1 would begin to break down. Likewise, a number of organs doubtless will be at least partially irradiated in any real treatment, and some of these might behave as coupled systems, affecting one another's radiation responses; here, too, the generalization of equation 1 becomes nontrivial.
5. Even if one could apply equation 1 directly in some situations, information on $S_{\text{organ}}(D)$ is difficult to obtain from patient or radiobiological data, and little of it is yet available (11,12). Most healthy organs in a patient undergoing treatment will be irradiated non-uniformly, moreover, making the prediction of their responses considerably more complex.
6. The above points have concerned the spatial distribution of dose. The timing of treatments is equally important, adding yet another facet to the problem. It has been found that healthy tissues fare better than do tumor cells if the dose is fractionated, or spread out over an extended period; they seem to retain more ability to repair damage to DNA over time. Since the normal tissue is irradiated heterogeneously, there may be, in addition to the cellular repair between the irradiations, organ repair on a macrolevel; that is, tissue rescuing units (cells migrating from the healthier parts of the organ, which received much lower doses) may help the organ to reestablish some of its damaged functionality.

Treatment is therefore delivered typically in 20–30 smaller fractions, rather with a single shot.

These and other issues, together with patient-specific factors, leave the radiation oncologist with a formidable, multidimensional treatment-planning optimization problem. The usual solution in real clinical situations is to sublimate most aspects of the probabilistic issues described above, and employ empirically established, time-tested protocols. It is in modifying these procedures, as warranted by the condition and response of the individual patient, that the acquired skills and informed judgment of a radiation oncologist are essential.

THE TREATMENT PLANNING PROCESS

The objective of curative radiotherapy is to deliver a tumoricidal dose to a clinically determined target volume (i.e., the volume of suspected gross disease plus a suitable margin), while depositing safe doses in the neighboring healthy tissues. The optimal strategy depends on the sensitivity of the tumor to radiation and its accessibility to treatment, and on the sensitivity and criticality of adjacent tissues that will be unavoidably dosed in the treatment.

For a superficial, radiosensitive lesion, it is often appropriate to employ an electron beam from a linear accelerator, which deposits most of the dose near the surface and leaves the underlying tissues unaffected.

The most common means of treatment of a deep-seated tumor in a modern medical setting is by means of highly penetrating, high-energy photons from a linac. (This article will use external photon-beam examples, but exactly the same issues are of concern for any form of radiotherapy.) Typically two to five X-ray beams cover a wide enough area (tumor region plus a *margin*) in their cross-fire region to eliminate any small clusters of cancer cells that may have extended into tissues adjacent to the primary target. The beams are arranged so as to deposit an adequate and uniform dose to the lesion, but without exceeding the tolerance levels in healthy tissues elsewhere. (It may often be helpful later to add a local *boost*, with a smaller-field photon or electron beam, or with brachytherapy, to increase the dose locally in the immediate vicinity of the primary tumor.) The resulting Compton (which predominate at high energies in soft tissue) and photoelectric interactions ionize the tissues they traverse. The production of free radicals and other molecular instabilities leads to damage to DNA and to the tissue microenvironment. That, in turn, is intended to kill the tumor cells.

In the first step in the treatment planning process itself, the radiation oncologist and treatment planning staff identify the tumor, through some combination of CT, MRI, PET, or by other means, and also the healthy critical structures to be avoided; multimodality imaging, in particular CT–PET fusion, is playing an increasingly important role in this. The oncologist then determines, based on experience and generally accepted treatment protocols, the dose that should be delivered to the tumor, and limits (to the known tolerance levels) the doses to be allowed at the normal tissues and critical organs. The radiation oncologist also

decides upon the *fractionation* schedule (the timing of the treatments.)

After the position, size, and shape of the lesion are determined, an *isocenter* (the single point at which the central axes of all treatment fields intersect) is chosen within it. Information on the patient's external shape, variations in tissue density, tumor geometry, and so on, is entered into the treatment planning computer, and one or more transverse planes are selected for the future display of the organs and superimposed topographical *isodose map*, with its contour lines of constant dose deposition.

The radiotherapy physicist and/or dosimetrist (a professional especially trained in treatment planning) then make an educated guess as to the energy, number, orientations, sizes, and weightings (the relative strengths, or relative contributions to the total dose at isocenter) of the beams, which can either be stationary or rotate through an arc, taking into account ways to block or otherwise modify parts of each one.

The computer then draws upon previously stored data to generate a representation of the spatial distribution of dose imparted by each of these beams, correcting for the patient's body shape, the lower density of lung tissue, and so on. Finally it sums, at each point of interest, the doses delivered by all the beams together, thereby producing a 3D isodose map. Although some work remains to be done for tissue inhomogeneities (especially lung and bone) for both photon and electron beams, most of the necessary dose calculation software has been developed. With our current abilities to model the interaction of high energy photon and charged particle beams with matter, and to localize healthy and pathological body structures, the computer can generate such maps with uncertainties in dose of less than a few percent.

In all likelihood, the plan can be improved by changing some of the beam parameters, so the process is repeated iteratively several times with different treatment configurations until a good (or at least clinically acceptable) dose distribution is obtained, as determined visually. (Numerical information, such as that from dose-volume histograms and the other approaches discussed below, is also being used increasingly as a guide.) This may involve the manipulation of a fair number of variables, and the plan ultimately selected may be only locally, rather than globally, optimal. Indeed, the limiting factors in the selection process may be the experience, creative skills, patience, and available time of the staff. This process is usually referred to as *forward treatment planning*.

The radiotherapy community has established a few empirical ground rules for judging treatment plans:

1. The entire tumor and a small margin should all receive at least the prescribed dose, and irradiation of the target volume should be reasonably uniform, to ensure the absence of any cold spots.
2. The highest dose isodose lines or surfaces should, when superimposed upon a set of CT or MRI scans, conform closely to the target volume.
3. The dose should decrease rapidly away from the target. Although a fair amount of healthy tissue

may have to be irradiated (and may thereafter become nonfunctional), the volume of it taken to high dose should be as small as possible, and dose falloff within it should be as fast as possible.

4. Dose to most of the lung, to all of the spinal cord, and so on, must be kept below their presumed tolerance levels; the tolerance dose of a healthy organ, however, may depend on the fraction of it that is being irradiated—although for some organs, such as the spinal cord or some of the brain, the dose limit applies to any small volume of tissue—which clearly complicates matters.

In summary, dose throughout the target should achieve the prescribed level, the volume of healthy tissues should be minimized, and doses to critical structures should be below their tolerance levels.

EXAMPLE OF PLAN OPTIMIZATION

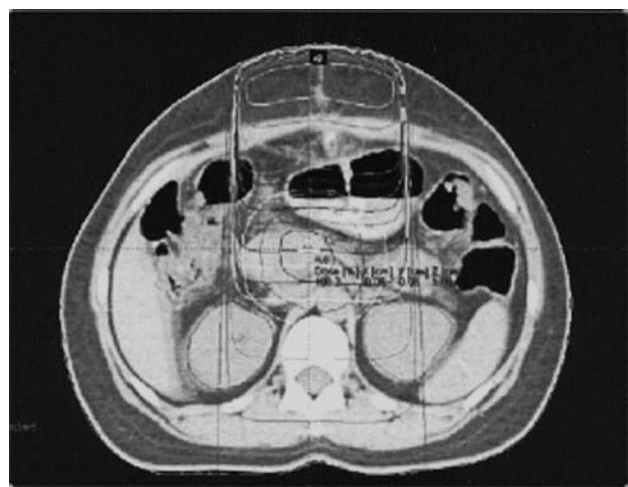
A unique treatment plan of beam configurations must be designed for each patient, based upon the specific clinical situation and requirements. Figure 2, for example, presents three possible plans and dose distributions for treating an oblong tumor of the pancreas, following surgery. The primary objectives are to take the entire tumor to a curative dose, typically 50–68 Gy, while making certain that the normal tissues of concern, the kidneys, spinal cord, and small bowel, remain functional. We shall examine a single-field plan first, then a pair of parallel-opposed fields, and finally a four-field plan.

For the single anterior field configuration, the dose across the target region ranges from 125% to 80% of the value at isocenter, taken to be at the center of the tumor, Fig. 2a. The spinal cord is at ~ 65% of the treatment value, which it can tolerate, and the maximum dose of 160% is situated in the anterior subcutaneous tissue layer. The lateral aspects of the kidneys are clear of any substantial dose. But this plan is not satisfactory because of both the cold area within the tumor and the anterior hot region, even though it has good cord- and kidney-sparing characteristics.

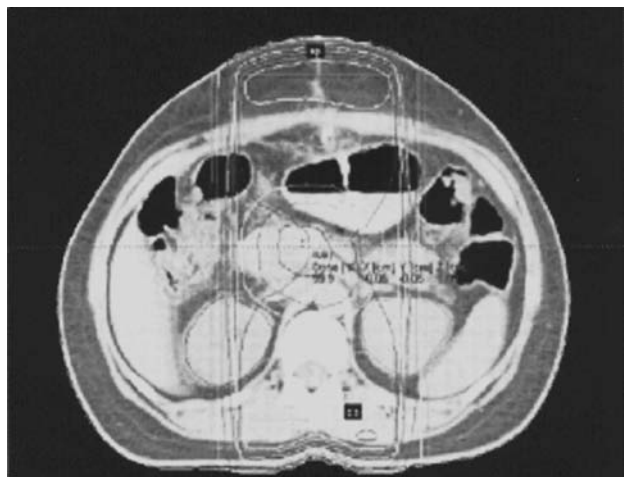
When a posterior beam is added to the anterior beam, the dose uniformity improves greatly, Fig. 2b. The dose gradient across the target region for this parallel-opposed plan is of the order of 5%, and the kidneys are still okay. The spinal cord now receives 105% of the target dose, however, which is too high.

Figure 2c is a four-field “box” plan that shows dose uniformity across the target and a high dose encompassing it. The spinal cord gets 60% of the tumor dose, and the kidneys and anterior bowel are spared reasonably well.

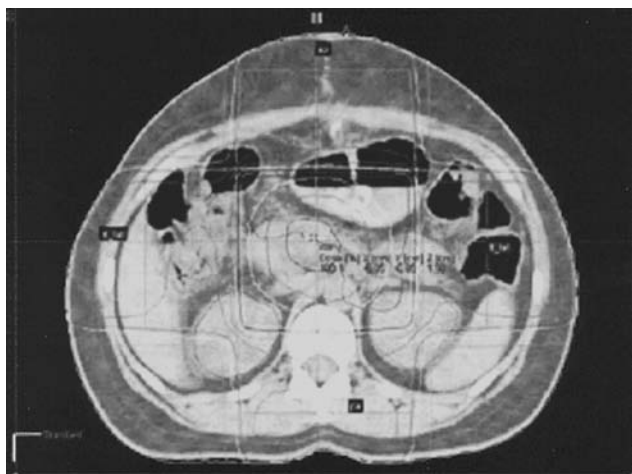
After going through such an analysis for a number of potential plans, it is usually apparent which ones work best for the patient. In the case of Fig. 2, an experienced physicist or dosimetrist would know immediately to begin with the box, but still would have to play with it to find the best weightings, or relative contributions, of the four fields. For some more difficult cases, it takes considerable effort to find a set of fields, and their settings, for which the ratio of



(a)



(b)



(c)

Figure 2. As possible plans for treatment of the pancreas with 10 MV X-ray beams, the physicist and/or dosimetrist consider the isodose maps for (a) a single AP field; (b) a pair of parallel-opposed fields; (c) a standard four-field box plan that covers the tumor relatively uniformly, spares most healthy tissue, and does not exceed the tolerance levels of either the kidney or the spinal cord.

dose in tumor to dose in the surrounding tissues is high; tumor dose is reasonably uniform; relatively little normal tissue is taken to high dose; and critical organs are not in serious jeopardy.

THE NEED FOR QUANTITATIVE OPTIMIZATION TOOLS

At present, visual inspection is still the standard process for optimizing isodose plans, as is done in this example, but this can be highly time and energy consuming. Meanwhile, the patient population is growing rapidly as the baby boomers reach older ages. Also, treatment strategies are becoming significantly more complex and demanding, especially so with the advent of intensity-modulated radiation therapy (IMRT), in which the beam direction, shape, intensity, and so on, may vary continuously over time during irradiation. So efficiencies in treatment planning that once were desirable are now becoming essential.

A means of assigning reliable numerical scores to competing plans would be of great value, not only in speeding up the through-put of routine cases, but also in facilitating decisions on the more complicated ones. Such scoring capability, moreover, forms the backbone of any computer-based expert system developed for automated treatment plan decision making. Much of the heavy mathematical machinery to find a maximum in an *objective function*, or *Figure of Merit*, is already available; the major problem, rather, is to find clinically meaningful criteria by which plans may be judged, and quantitative forms in which these criteria can be expressed.

Two fundamentally different approaches have arisen to the use of computers in judging treatment plans.

The first draws upon some of the physical and geometrical rules of thumb that are commonly adopted in visual plan assessment. How constant is the dose throughout the target volume, and does the prescribed-dose isodose surface conform closely to its surface? How much dose is being deposited in healthy tissues, and is the tolerance level of any critical organ exceeded? Such criteria have the benefits of being easily understood and clinically widely accepted, and of being easy to implement in an automated optimization system. On the downside, it is not clear that the plan with the most uniform tumor dose is necessarily most likely to effect a cure, or that the one with the lowest average dose to healthy tissue will most probably avoid complications. The obvious physically-based criteria, either individually or in combination, do not necessarily reveal the strategy most likely to work.

The second is somewhat more abstract, and more removed from traditional clinical practice. It attempts to draw upon various kinds of available dose-response information on normal tissues to estimate the probabilities that any particular dose distribution in an organ will lead to complications (13–15). This biologically and probabilistically based mathematical modeling work is still in early stages of development, and obtaining and interpreting the necessary data on radiation damage is difficult. Nonetheless, some researchers feel that the approach has perhaps greater potential for eventual success.

Either way, three aspects to the optimization process must be addressed:

1. First, it is necessary to decide what would constitute an ideal treatment. A plan judged to be nearly perfect according to one criterion might well be far from it according to another; the best overall optimization criterion to be found might, indeed, be some balanced amalgam of several others.
2. One must then find a quantitative measure of the deviation of any computer-generated plan from the ideal.
3. Finally, an algorithm is needed to search out quickly the plan that deviates least from the ideal, as beam weights, field sizes, and other parameters are varied.

OPTIMIZATION USING PHYSICAL/GEOMETRICAL CRITERIA

Several traditional criteria for the visual optimization of plans are in common use. One would like the prescription-dose isodose surface to wrap as closely as possible around the target volume. Also, some have argued, the dose throughout the tumor should be uniform. It is desirable to keep the amount of dose to healthy tissues to a minimum, and doses to all critical structures must be maintained below their tolerance levels. All of these ideas, it turns out, have been easy to incorporate into computer-based optimization programs.

For all of what follows, we shall partition the region of interest into J small voxels, all of volume Δv , and assume that a particular configuration of beams delivers the dose D_j to the j th such voxel. For simplicity, we shall also consider situations in which all beam parameters have already been chosen by the treatment planner, with the exception of the best values of the relative weightings, w_k , of the K beams (w_k refers to the relative dose at isocenter produced by the k th field).

For any possible treatment plan, the dose in the j th voxel is now fixed completely by

$$D_j = \sum_k^K d_{jk} w_k \quad (2)$$

where $w_k \geq 0$. For each particular configuration of beams, the matrix of elements d_{jk} determines the dose deposited at the j th voxel by the k th field. A standard treatment planning program is designed to calculate such matrices. Changing the energy, orientation, blocking, and so on, of the beams, however, will result in a revised matrix of d_{jk} values.

Healthy Tissue Integral Dose

The simplest single measure of the irradiation of a block of healthy tissue is the *integral dose* (ID). Integral dose is defined through the function

$$f_{ID} = \sum_j^J D_j \Delta v \quad (\text{healthy tissue}) \quad (3a)$$

where the sum is confined to healthy tissues outside the target volume. Equation 3a serves as the objective function that is to be minimized, by manipulating various beam parameters, when integral dose is selected as the optimization criterion. Note that apart from a factor of $J\Delta v$, the integral dose is just the average voxel dose. By equation 2, the integral dose may be expressed as

$$f_{ID} = \sum_j^J \sum_k^K d_{jk} w_k \Delta v \quad (\text{healthy tissue}) \quad (3b)$$

It is intended, moreover, that the distribution deliver at least a tumoricidal dose, $D_{\text{prescribed}}$,

$$D_l \geq D_{\text{prescribed}} \quad (\text{tumor}) \quad (4a)$$

at previously selected points within the target volume. Similar constraints,

$$D_{l'} \leq D_{\text{tolerance}} \quad (\text{point } l' \text{ in critical organ}) \quad (4b)$$

protect any critical structure, such as the spinal cord.

The task now is to find those weights that deliver the prescribed dose to the target volume, but minimize the integral dose objective function, while keeping irradiation of certain voxels below specified limits. The objective function is linear in the weights w_k , the variables of interest, and equations 3 and 4 define a routine linear programming problem whose solution can be obtained in seconds with the Simplex method.

Despite the mathematical attractiveness of the integral dose objective function, its application does not necessarily lead to results in accord with clinical experience. Taking an entire organ to 30 Gy, for example, yields the same f_{ID} as does taking half of it to 60 Gy; but the two irradiation schemes could lead to vastly different results if the organ's tissues exhibit a response threshold at 45 Gy. This may serve as a reminder of a fundamental, but sometimes overlooked truth: The availability of powerful mathematical optimization algorithms does not ensure the clinical value of an objective function upon which they can be employed.

Optimal Point Doses

In perhaps the most widely explored and successful of the physical-geometric approaches, the physician prescribes the doses to be delivered at a designated set of points within the patient; some of these, in the tumor and in critical organs, act as rigid constraints. The treatment planning system then uses available mathematical techniques to find that plan (i.e., set of beam sizes, angles, relative weightings, etc.) for which the calculated doses at these points come closest, on the whole, to the prescribed values. Points might be selected, for example, on the tumor border, to force the prescribed-dose isodose surface to envelope it closely, while also satisfying the constraint requirements. Alternatively, they might be placed throughout the target volume, to maximize tumor dose uniformity.

Again, let the variable parameters be the relative beam weightings, and now suppose that the objective is to maximize tumor dose uniformity (TDU). The M prescription

points will be scattered throughout the tumor, and, by our criterion, doses there should all be nearly the same. The deviation of any plan from the ideal may be expressed as

$$f_{TDU} = \sum_m^M (D_m - D_{av})^2 \quad (5)$$

where the D_m provided by equation 2 are functions of the w_i values, and D_{av} is the average dose for the selected points.

Similar objective functions can be devised to force the prescribed isodose line to lie as near as possible to the target volume boundary, or to cause certain discrete anatomic regions to receive prescribed doses. Methods such as quadratic programming and constrained least-squares algorithms are available for rapidly finding extrema in functions of the form of equation 5, subject to constraints like equations 4. While optimization with such objective functions may not alone yield clinically appropriate plans, they can often provide starting points for further manual searching.

Developments in computer-controlled radiation treatment have stimulated renewed interest in such approaches. Computers have long been employed to monitor a linear accelerator's treatment parameters, verifying and recording every field's gantry angle, jaw setting, dose delivered, and so on. More recently, with the growing adoption of IMRT systems, dedicated computers are controlling variables such as collimator configuration, dose rate, gantry angle, table movements in real time. While such flexibility allows the implementation of highly refined treatment plans, the associated planning process can be very cumbersome and time consuming. The problem would be totally unmanageable were it not for the availability of various objective functions and search programs (known as inverse planning algorithms) to help in finding suitable time-dependent linac output parameters (16,17).

References that describe optimization by way of physical-geometric criteria may be found in Ref. 18.

BIOLOGICAL-PROBABILISTIC APPROACHES TO TREATMENT PLAN OPTIMIZATION

The physical-geometric approaches just discussed could perhaps best be called pragmatic. They reflect mathematically some of the conventional clinical criteria for plan optimization; as such, they can discriminate among plans only as effectively as do the standard clinical guidelines that they mimic. They cannot choose, for example, between a treatment that irradiates a sizable portion of lung to a dose near its tolerance level and another treatment that takes even more lung to a somewhat lower dose, other aspects of the case being equal.

The biological-probabilistic methods, admittedly in their infancy, attempt to address directly the fundamental question: How probable is it that some given spatial and temporal distribution of dose will eradicate or control a tumor, and do so without resulting in severe complications? Asking this returns us to the thinking illustrated in Fig. 1.

Limited theoretical progress has been made on several fronts. Among them, we can derive some dose-response curves of the general form of S_{tumor} and S_{organ} from more

basic radiobiological principles. We have elementary ideas on how to use such information to estimate the probability of complications arising in a healthy organ irradiated non-uniformly, and we can crudely model the radiation response of a tumor.

Shape of a Tumor Dose-Response Curve, S_{tumor}

Of an initial cluster of $n(0)$ of any kind of cells, only $n(D)$ of them will remain viable after receiving an initial dose of D Gy; the rest will die before or while attempting to divide. If an incremental dose, ΔD , soon follows (before repair or repopulation occurs), the number of surviving cells diminishes by an additional

$$\Delta n = -c(D)n(D)\Delta D \quad (6a)$$

equivalently,

$$c(D) = -\Delta n/n(D)\Delta D \quad (6b)$$

is the probability per unit dose of inactivating a cell that has already been given D . There is no physics or real biology in the picture yet—everything so far is just bookkeeping.

For the purposes of this article, the discussion will be greatly simplified by restricting the biophysics being considered. Although low Linear Energy Transfer (LET) radiations, namely high-energy photons and electrons, are almost always employed in practice, here the treatment is by way of a beam of energetic, heavy charged particles, such as cyclotron-produced protons. Every proton leaves a dense trail of ionization in a cell it traverses, and damage to any DNA molecule it passes sufficiently close to is likely to involve irreparable breaks in both sugar-phosphate strands (19). Cells in which this occurs will probably be completely inactivated by a single hit of an incident particle; most cells more distant from its track, on the other hand, will be oblivious to its passage. Cell killing in this situation is thus an all-or-nothing affair. Surviving cells retain no memory of earlier irradiations, and $c(D)$ becomes a constant, commonly called $1/D_0$, independent of dose.

In addition, it is assumed that the tumor cells of concern do not communicate with one another in any significant way; that is, radiation damage in one voxel does not notably affect (by releasing toxins, cutting off the influx of nutrients, etc., as in "bystander" phenomena) the dose-response characteristics of an adjacent voxel.

After replacing $c(D)$ in equation 6 with $1/D_0$, the associated differential equation

$$dn/dD = -(1/D_0)n(D) \quad (7a)$$

can be integrated to yield

$$n(D)/n(0) = e^{-D/D_0} \quad (7b)$$

This is the fraction of cells, either in the tumor or in any small part of it, that survives irradiation to dose D . It may be viewed equally well as the survival probability for an individual tumor cell. Expressing $n(0)$ as the product (ρV) , where ρ is the tumor cell density and V is the tumor volume, this becomes

$$n(D) = n(0)e^{-D/D_0} = (\rho V)e^{-D/D_0} \quad (7c)$$

This is the number of cells statistically expected to survive when the entire tumor receives a dose of D . All the relevant radiobiology and radiation physics reside within D_0 , which can be a complex function of any number of interesting parameters that may, or may not, be understood. (With X rays or electrons, single-strand DNA breaks occur; some degree of repair can take place, and $c(D)$ is not independent of D . Integration of equation 7a then leads to a cell survival curve with a shoulder, rather than equation 7c.)

It is believed that a tumor can repopulate from a single clonogenic cell. The desired outcome of a curative treatment must therefore be the total eradication of all tumor cells. That is, the probability that *no* cells survive must be high. Poisson (or binomial) statistics provides a way of assessing that probability: if the average number of times some event takes place in a situation of interest is μ , then the probability of exactly zero such events occurring, $P_\mu(0)$, is $e^{-\mu}$,

$$P_\mu(0) = e^{-\mu} \quad (\mu = \text{average number of events}) \quad (8a)$$

In the present case, the average number of cells expected still to be clonogenic after an irradiation of D is given by equation 7c. From equation 8a, then, the probability, $P_D(0)$, that there will be *no* tumor cells left viable is

$$S_{\text{tumor}}(D) = P_D(0) = e^{-(\rho V)e^{-D/D_0}} \quad (8b)$$

This can be rewritten in terms of the geometrical characteristics of the curve, namely D_{50} the dose resulting in 50% probability of complication, $S_{\text{organ}} 0.5$ (see also Ref. 20), and slope, γ_{50} :

$$S_{\text{tumor}}(D) = (1/2)^{e^{2\gamma_{50}(1-D/D_{50})/\ln 2}} \quad (8c)$$

This fits clinical data sufficiently well (21,22); indeed, a number of tumor cell populations characterized by different parameter values result in almost indistinguishable sets of dose response curves, in part because of strong correlations between the parameters. The D_{50} and γ_{50} values for a number of sites have been published (21).

Equations 8b and 8c provide, and Fig. 1 displays, one particular form for the probability, $S_{\text{tumor}}(D)$ of equation 1, that the patient will survive the disease. It increases sigmoidally with dose, as certainly expected. It also decreases exponentially with tumor size; equation 8b indicates that the doses required to eradicate (with equal probabilities of success) two otherwise identical tumors that differ only in pre-irradiation volume are related as

$$D_2 = D_1 + D_0 \ln (V_2/V_1) \quad (8d)$$

This model is simple but, not too surprisingly, it agrees with the clinical finding that larger tumors require more dose to achieve a cure than do smaller tumors of the same histological type.

Equation 8 describe the irradiation characteristics of a tumor exposed uniformly. The result is of legitimate clinical interest since it is common practice to attempt to impart a fairly flat dose across the target volume. In the case of non-uniform irradiation, the tumor can be viewed as

consisting of J voxels that receive doses D_j , with control probabilities of $s(D_j)$, again assuming their statistical independence. The parameter S_{tumor} is then given by

$$S_{\text{tumor}} = \prod_j^J s(D_j) \quad (8e)$$

One of the main characteristics of a tumor is its comparatively fast repopulation rate. To account for this in conjunction with the Poisson (or binomial) distribution, several authors in the 1980s proposed setting the initial number of clonogens to $n(0)$, after which $n(t) = n(0)e^{-\lambda t}$, where λ and t are the rate constant and repopulation time, respectively (23–26). This approach predicts a $S_{\text{tumor}}(t)$ that always tends to zero for large post-treatment times, which is incorrect. Later, taking clonogen repopulation between fractions (27–30) into account lead to the Zaider–Minerbo model (31) of $S_{\text{tumor}}(t)$, which is applicable for any temporal treatment protocol. An expression for $S_{\text{tumor}}(t)$ with different time intervals between consecutive irradiation fractions and with varying cell survival probability per fraction was obtained by Stavreva et al. (32) based on the Zaider–Minerbo approach. Animal experiments support the validity of the Zaider–Minerbo approach (32).

Other approaches to the determination of S_{tumor} are discussed in the literature (33–36).

Integral Response for a Healthy Organ

Several ways have been devised to address the non-uniform irradiation of healthy tissues. One of these is an extension of the integral dose idea, introduced in equations 3 and 4. It does not focus on the spatially varying dose distribution within a tissue *per se*, as do the physical–geometric approaches, but rather on the local biological response that the dose elicits. Variation on the approach, of a range of levels of complexity, have been discussed by a number of researchers (18,37–41).

Imagine an organ or biological compartment that produces some physiologically important substance (such as a critical enzyme or hormone) or that performs an important task (like phagocytosis or gas exchange). If too many of its cells are inactivated by irradiation, then the organ cannot do its job adequately, and the organism runs into trouble. As before, mathematically partition the organ into J small volume elements of size Δv , and assume that the radiation response of any such voxel (or of the cells within it) is nearly independent of its neighbor's response. If, moreover, high LET radiation is again involved, the dose–response relationship is of the form of equation 7b, where D_0 contains all the important biophysics.

Of the entire organ, only the fraction

$$v = (1/J \Delta v) \sum_j^J e^{-D_j/D_0} \Delta v \quad (9a)$$

of its tissue will remain functional, where $(J \Delta v)$ is its volume. (The parameter $(1 - v)$ thus provides a direct measure of the amount of radiation damage to the organ.) For the case of nearly uniform irradiation to the level D ,

v reduces to the

$$n(D)/n(0) = e^{-D/D_0} \quad (9b)$$

of equation 7b. This observation, along with the form of equation 9a, suggests that v be viewed as a generalized, or spatially averaged, dose–response parameter, prompting adoption of the name “integral response” (18,34).

The probability that the patient will escape serious complications, S_{organ} , is a nondecreasing function of the relative amount of organ that remains intact,

$$S_{\text{organ}} = S_{\text{organ}}(v) \quad (9c)$$

the shape of which must be obtained experimentally or by other means (39,40). The parameter $S_{\text{organ}}(v)$ and equations 2 and 9a together define the integral response (IR) objective function, f_{IR} , and the optimization problem is bound by the constraints of equations 3 and 4.

If Q radiobiologically independent organs are irradiated, and if the possible complications are all of comparable severity, then the overall survival probability may be given by the product

$$S_{Q \text{ organs}} = \prod_q^Q S_{\text{organ } q} \quad (9d)$$

Automated treatment plan optimization may then be carried out, in principle, with a combination of equations 2, 4, and 9.

Another possible method of handling non-uniform irradiation of such an organ is based on its N -step dose–volume histogram. The idea is to reduce it in such a fashion as to yield a revised histogram that corresponds to the same complication probability, but that contains only an $(N - 1)$ steps; this calculation is repeated $(N - 1)$ times, until there is left a single-step histogram, the S_{organ} of which can be obtained from experiment or clinical observation.

Several algorithms have been proposed for carrying out the histogram-reduction process (42–47).

Healthy Organ Composed of Separately Critical Voxels

The integral response objective function is based on the effectiveness of operation of a healthy organ taken as a unit (48). A radically different approach is needed for an organ, such as the spinal cord or certain regions of the brain, that behaves like a chain or computer program, in which serious complications in any single small part can spell disaster for the whole.

Once again, consider an organ made up of J radiobiologically independent voxels. Let $s(D_j)$ refer to the probability that the organ will suffer no serious complications when the j th voxel is taken to dose level D_j and the rest is left unirradiated. If the entire organ is to escape damage, each of its J parts must do so separately, and

$$S_{\text{organ}} = \prod_j^J s(D_j) \quad (10a)$$

If the small-volume tolerance dose is exceeded in even a single voxel, then the objective function S_{organ} can become perilously low. If the organ is irradiated uniformly,

incidentally, equation 10a reduces to

$$S_{\text{organ}} = [s(D)]^J \quad (10b)$$

The methods just described for dealing with nonuniform irradiation make use of the assumption of the radiobiological independence of adjacent voxels. While this assumption clearly is not valid for most organs, nor perhaps for tumors, it may apply to some (e.g., the blood pool). More importantly, models built upon it may serve as jumping-off points for the development of more realistic pictures.

Other approaches to the determination of S_{organ} are discussed in the literature (49–52).

CONCLUSION

The radiation response of a tissue depends in an extremely complex way on a number of parameters, some of which the radiation oncologist can control directly, some indirectly, and some not at all. One can choose the modality (X rays, electrons, gamma-rays, protons, neutrons); the volume to be irradiated; the dose per fraction and number of fractions; the administration of response-modifying drugs; and the spatial dose distribution in healthy tissues. Some of the generally uncontrollable (or weakly controllable) parameters are radiosensitivity differences of different constituent parts of a tissue; tissue concentrations of oxygen, drugs, toxins, and other compounds; numbers of cells in each portion of the mitotic cycle; differential cell population kinetics; and repair of radiation-induced cellular injury. Manipulation of directly controllable variables may lead to changes in others: reoxygenation; differential cell cycle phase redistribution; differential recruitment of proliferative cells; and differential repair.

In view of this complexity, it is not clear how effective empirical or *ab initio* mathematical modeling can be in providing clinically useful descriptions of processes as intricate as the kinetics of irradiated tissues. At the present time, the subject is largely of academic interest. But it is to be expected that over the next several decades, computer-based expert systems will continue to spread throughout all of medicine. This development, together with the increasing automation of radiation delivery systems, will doubtless cause the search for quantitative methods of treatment plan optimization to expand.

BIBLIOGRAPHY

1. Wolbarst AB. *Physics of Radiology*. 2nd ed. Madison (WI): Medical Physics Publishing; 2005.
2. Lawrence TS, Ten Haken RK, Kessler ML, Robertson JM, Lyman JT, Lavigne ML, Brown MB, Duross DJ, Andrews JC, Ensminger Wd, Lichter AS. The use of 3-D dose volume analysis to predict radiation hepatitis. *Int J Radiat Oncol Biol Phys* 1992;23:781–788.
3. Jackson A, Ten Haken RK, Robertson JM, Kessler ML, Kutcher GJ, Lawrence TS. Analysis of clinical complication data for radiation hepatitis using a parallel architecture model. *Int J Radiat Oncol Biol Phys* 1995;31:883–891.

4. Fischer JJ. Theoretical considerations in the optimisation of dose distribution in radiation therapy. *Br J Radiol* 1969;42: 925–930.
5. Brahme A, Agren AK. Optimal dose distribution for eradication of heterogeneous tumours. *Acta Oncol* 1987;26:377–385.
6. Webb S, Nahum AE. A model for calculating tumour control probability in radiotherapy including the effects of inhomogeneous distributions of dose and clonogenic cell density. *Phys Med Biol* 1993;38:653–666.
7. Webb S, Evans PM, Swindell W, Deasy JO. A proof that uniform dose gives the greatest TCP for fixed integral dose in the planning target volume. *Phys Med Biol* 1994;39:2091–2098.
8. Stavreva NA, Stavrev PV, Round WH. A variational approach to the problem of optimizing the radiation dose distribution in tumours. *Australas Phys Eng Sci Med* 1996;19:9–11.
9. Stavreva NA, Stavrev PV, Round WH. A mathematical approach to optimizing the radiation dose distribution in heterogeneous tumours. *Acta Oncol* 1996;35:727–732.
10. Ebert MA, Hoban PW. Some characteristics of tumour control probability for heterogeneous tumours. *Phys Med Biol* 1996;41:2125–2133.
11. Sanchez-Nieto B, Nahum AE. Bioplan. Software for the biological evaluation of radiotherapy treatment plans. *Med Dosim* 2000;25:71–76.
12. Warkentin B, Stavreva N, Stavrev P, Field C, Fallone BG. A TCP-NTCP estimation module using DVHs and known radiobiological models and parameter sets. *J Appl Clin Med Phys* 2004;5:1–14.
13. Brahme A. Biologically based treatment planning. *Acta Oncol* 1999;38(13 Suppl): 61–68.
14. De Gersem WR, Derycke S, De Wagter C, De Neve WC. Optimization of beam weights in conformal radiotherapy planning of stage III non-small cell lung cancer: Effects on therapeutic ratio. *Int J Radiat Oncol Biol Phys* 2000;47:255–260.
15. Stavrev P, Hristov D, Warkentin B, Sham E, Stavreva N, Fallone BG. Inverse treatment planning by physically constrained minimization of a biological objective function. *Med Phys* 2003;30:2948–2958.
16. Brahme A, Roos JE, Lax I. Solution of an integral equation encountered in rotation therapy. *Phys Med Biol* 1982;27:1221–1229.
17. Webb S, Convery D, Evans PM. Inverse planning with constraints to generate smoothed intensity-modulated beams. *Phys Med Biol* 1998;43:2785–2794.
18. Wolbarst AB, Chin LM, Svensson GK. Optimization of radiation therapy: Integral-response of a model biological system. *Int J Radiat Oncol Biol Phys* 1982;8:1761–1769.
19. Hall E. *Radiobiology for the Radiologist*. 5th ed. Baltimore: Lippincott Williams & Wilkins; 2000.
20. Stavreva N, Stavrev P, Warkentin B, Fallone BG. Derivation of the expressions for gamma50 and D50 for different individual TCP and NTCP models. *Phys Med Biol* 2002;7:3591–3604.
21. Okunieff P, Morgan D, Niemierko A, Suit HD. Radiation dose-response of human tumours. *Int J Radiat Oncol Biol Phys* 1995;32:1227–1237.
22. Goitein M, Niemierko A, Okunieff P. The probability of controlling an inhomogeneously irradiated tumor: A stratagem for improving tumor control through partial tumor boosting. 19th L H Gray Conference: Quantitative Imaging in Oncology. Newcastle (UK): 1995; p 25–39.
23. Thames HD, Rozell ME, Tucker SL, Ang KK, Fisher DR, Travis EL. Direct analysis of quantal radiation response data. *Int J Radiat Oncol Biol Phys* 1986;49:999–1009.
24. Tucker SL, Travis EL. Comments on a time-dependent version of the linear-quadratic model. *Radiother Oncol* 1990;18: 155–163.
25. van de Geijn J. Incorporating the time factor into the linear-quadratic model. *Br J Radiol* 1989;62:296–298.
26. Yaes RJ. Linear-quadratic model isoeffect relations for proliferating tumor-cells for treatment with multiple fractions per day. *Int J Radiat Oncol Biol Phys* 1989;17:901–905.
27. Tucker SL, Thames HD, Taylor JM. How well is the probability of tumor cure after fractionated irradiation described by Poisson statistics? *Radiat Res* 1990;124:273–282.
28. Yakovlev AY. Comments on the distribution of clonogens in irradiated tumors. *Radiat Res* 1993;134:117–120.
29. Kendal WS. A closed-form description of tumour control with fractionated radiotherapy and repopulation. *Int J Radiat Oncol Biol Phys* 1998;73:207–210.
30. Tucker SL, Taylor JM. Improved models of tumour cure. *Int J Radiat Biol* 1996;70:539–553.
31. Zaider M, Minerbo GN. Tumour control probability: A formulation applicable to any temporal protocol of dose delivery. *Phys Med Biol* 2000;45:279–293.
32. Stavreva N, Stavrev P, Warkentin B, Fallone BG. Investigating the effect of cell repopulation on the tumor response to fractionated external radiotherapy. *Med Phys* 2003;30:735–742.
33. Wolbarst AB, Sternick ES, Curran BH, Dritschilo A. Optimized radiotherapy treatment planning using the complication probability factor (CPF). *Int J Radiat Oncol Biol Phys* 1980;6:723–728.
34. Roberts SA, Hendry JH. A realistic closed-form radiobiological model of clinical tumor-control data incorporating intertumor heterogeneity. *Int J Radiat Oncol Biol Phys* 1998;41:689–699.
35. Fenwick JD. Predicting the radiation control probability of heterogeneous tumour ensembles: Data analysis and parameter estimation using a closed-form expression. *Phys Med Biol* 1998;43:2159–2178.
36. Brenner DJ. Dose, volume, and tumor-control predictions in radiotherapy. *Int J Radiat Oncol Biol Phys* 1993;26:171–179.
37. Kallman P, Agren A, Brahme A. Tumour and normal tissue responses to fractionated non-uniform dose delivery. *Int J Radiat Biol* 1992;62:249–262.
38. Withers HR, Taylor JM, Maciejewski B. Treatment volume and tissue tolerance. *Int J Radiat Oncol Biol Phys* 1988; 14:751–759.
39. Jackson A, Kutcher GJ, Yorke ED. Probability of radiation-induced complications for normal tissues with parallel architecture subject to non-uniform irradiation. *Med Phys* 1993; 20:613–625.
40. Niemierko A, Goitein M. Modeling of normal tissue response to radiation: The critical volume model. *Int J Radiat Oncol Biol Phys* 1993;25:135–145.
41. Olsen DR, Kambestad BK, Kristoffersen DT. Calculation of radiation induced complication probabilities for brain, liver and kidney, and the use of a reliability model to estimate critical volume fractions. *Br J Radiol* 1994;67:1218–1225.
42. Lyman JT, Wolbarst AB. Optimization of radiation therapy, III: A method of assessing complication probabilities from dose-volume histograms. *Int J Radiat Oncol Biol Phys* 1987;13:103–109.
43. Lyman JT, Wolbarst AB. Optimization of radiation therapy, IV: A dose-volume histogram reduction algorithm. *Int J Radiat Oncol Biol Phys* 1989;17:433–436.
44. Kutcher GJ, Burman C. Calculation of complication probability factors for non-uniform normal tissue irradiation: The effective volume method. *Int J Radiat Oncol Biol Phys* 1989;16:1623–1630.
45. Kutcher GJ, Burman C, Brewster L, Goitein M, Mohan R. Histogram reduction method for calculating complication probabilities for three-dimensional treatment planning evaluations. *Int J Radiat Oncol Biol Phys* 1991;21:137–146.
46. Niemierko A, Goitein M. Calculation of normal tissue complication probability and dose-volume histogram reduction

- schemes for tissues with a critical element architecture. *Radiother Oncol* 1991;20:166–176.
47. Niemierko A. A generalized concept of equivalent uniform dose. 41th AAPM Annual Meeting, Nashville, 24–29 July, 1999. *Med Phys* 1999;26:1100.
 48. Wolbarst AB. Optimization of radiation therapy II: The critical-voxel model. *Int J Radiat Oncol Biol Phys* 1984;10:741–745.
 49. Lyman JT. Complication probability as assessed from dose-volume histograms. *Radiat Res (Suppl)* 1985;8:S13–S19.
 50. Yaes RJ, Kalend A. Local stem cell depletion model for radiation myelitis. *Int J Radiat Oncol Biol Phys* 1988;14:1247–1259.
 51. Stavrev P, Stavreva N, Niemierko A, Goitein M. Generalization of a model of tissue response to radiation based on the idea of functional subunits and binomial statistics. *Phys Med Biol* 2001;46:1501–1518.
 52. Stavreva N, Niemierko A, Stavrev P, Goitein M. Modeling the dose-volume response of the spinal cord, based on the idea of damage to contiguous functional subunits. *Int J Radiat Biol* 2001;77:695–702.

See also RADIATION DOSE PLANNING, COMPUTER-AIDED; RADIATION THERAPY TREATMENT PLANNING, MONTE CARLO CALCULATIONS IN; RADIATION THERAPY, QUALITY ASSURANCE IN; RADIOLOGY INFORMATION SYSTEMS.

RANGE OF MOTION. See REHABILITATION AND MUSCLE TESTING.

RECORDERS, GRAPHIC

HERMAN VERMARIEN
Vrije Universiteit Brussel
Brussels, Belgium

INTRODUCTION

Graphic recorders, as considered here, are essentially measuring instruments that produce in real-time graphic representations of biomedical signals, in the form of a permanent document intended for visual inspection. Recorded data are thus fixed on a two-dimensional (2D) medium which can simply be called “paper” (although the material applied may differ from ordinary writing paper); A so-called hard copy of the information is generated. Equivalent names are paper recorder, direct writing recorder, plotter, chart recorder, and strip chart recorder (if long strips of paper are used); In some cases, the more general term hard-copy unit is also applied. The nomenclature oscillograph is sometimes used (in correspondence with the other display instrument, the oscilloscope). By the aspect of visual inspection the graphic recorder is distinguished from other data storage devices (e.g., magnetic or optical disk, solid-state memory), whereas in the property of permanence graphic recording differs from visual monitoring as realized by the oscilloscope or the computer screen. Real-time paper recordings have the benefit of direct visual access to signal information, allow immediate examination (and re-examination) of trends (as long strips of paper can be used), present better graphic quality than

most screens and can be used as a document for scientific evidence.

The graphic records are inspected through the human visual waveform recognition abilities; Moreover, discrete parameter values can be derived that are further evaluated. For the latter purpose, the measuring ruler still continues to be an intensively used instrument. Values for the physical variables presented, including time, can be derived, usually by comparing different related biomedical signals that were simultaneously recorded. For this purpose most biomedical recorders are multichannel recorders, which are able to process a set of signals. As the information has to be stored, the content of the records cannot be limited to the naked signal tracings; It is obviously essential that additional information concerning the recorded signals (identification, calibration, etc.) and the experimental circumstances (date, subject, experiment description, stimulus type, etc.) be kept in file, by preference directly on the recorded charts.

Two primary aspects describing the performance of the graphic recorder are (1) the properties of the instrument as a recorder of information, that is, measuring accuracy and the ability to display additional information, and (2) the quality of graphics, implying the clearness of the tracings and the overall graphic layout (e.g., including identification of the curves by color difference) in relation to ergonomics in visual examination. Also, the stability in time (permanence) of these graphic qualities can be catalogued under item 2. Secondary aspects (but not necessarily less important to the user) are ease of control of the apparatus (adjustments, calibration, adding alphanumeric information, remote control, computer connection and communication possibilities, paper loading), input amplification and signal conditioning facilities, unavoidable maintenance (ink, pens, mechanical parts), costs (the cost not only of the apparatus, but also of maintenance and, not negligibly, paper), and service life.

In early physiological experiments on mechanical functioning, such as muscle contraction, recording was performed with the aid of a directly coupled writing stylus; Suspended with minimal weight and friction, the stylus arm was applied as a cantilever, one end connected to the contracting muscle, the other end provided with a tip writing on a rotating drum. A directly coupled method is found in the spirometer: The low weight air cavity of the spirometer, which moves up and down as the subject expires and inspires, is mechanically linked to a pen writing on calibrated paper that moves at constant velocity, thus generating a trace of the course of lung volume versus time. Nevertheless, most recorders use a transducer that converts electrical information (signal voltage or current) into mechanical data, more specifically, position on the paper. This can be achieved by moving the writing element to the specific position (analog transducers) or by activating the correct point in a large array of stationary, equally spaced writing elements (digital transducers).

At present, sophisticated graphic recorders that offer a broad variety of possibilities to the operator are available. The measuring quality of transducers has continuously been improved and, as digital techniques have been applied, the possibilities of automatic control and addition

of alphanumeric information have been largely extended. A typical example is automatic recording of standard ECG derivations (frontal bipolar and unipolar, and precordial derivations) with calibration, identification of the curves, and determination of typical parameters (such as the duration of specific electrocardiographic intervals) and generation of diagnostics. Nowadays, the borderline between graphic recorders, digital oscilloscopes, data-acquisition systems, and PC-based virtual measuring instruments become less clear. There is a decreasing interest in analog recorders except for the multipen recorders having the benefit of simplicity, low price, and excellent identifiability of curves by the use of different colors. They are used in laboratory applications (and process monitoring) and handle slow varying signals. Digital recorders are generally provided with an LCD screen allowing monitoring with different colored tracings. Moreover, signal processing and data extraction, storage of data (and replay of original or processed signals), computer connection for handling data and control of recording settings are possible. Measurements where immediate visualization of the tracings on a long strip of paper is not required and data are to be digitally analyzed and stored, can be performed by PC with virtual instrument software and printer; Nowadays the majority of data collection is accomplished using digital PC approaches.

In giving an overview of graphic recorder function, one inevitably refers to technology of transducers. Especially in analog recorders the capacities and the limitations of the transducer is of major importance to the quality of the complete recorder. A number of mechanisms were applied in commercially available devices, but in many applications the analog types have been replaced by the digital type or by the PC system. Recorders can be called special purpose when built within a complete measuring system [e.g., an electrocardiograph (ECG), an electroencephalograph (EEG)], or they can be general purpose. The latter may be provided with specific signal conditioning modules to be connected to biophysical sensors (e.g., manometers for blood pressure measurements, Doppler probes for blood velocity assessment, thermal sensors for temperature recording) or, even more generally, to standard amplifiers allowing amplification levels, direct current (dc) adjustment, filtering, and so on. In most applications the graphic transducer, not the electronic signal conditioning hardware, is the crucial stage in the measuring chain. This article is intended to cover primarily the essentials of graphic recording, focusing on the principal aspects mentioned before: recording of information and graphic quality. For extensive practical details on recorders and signal conditioning modules, waveform monitoring, digital storage, processing and communication facilities, the manufacturers' data sheets should be consulted.

FUNDAMENTAL ASPECTS OF GRAPHIC RECORDING

Graphics: Images or Tracings

As the paper used for graphic recording is a 2D medium, two coordinates, x and y , can be defined: y represents the ordinate corresponding to paper width; x is the abscissa,

corresponds to paper length. If an exception is made for the strip chart recorder, the distinction between x and y is merely a matter of definition. In the most general case an image can be presented on the paper: At each point (x, y) a gray scale or color scale value is displayed. For example, in speech analysis, the so-called spectrograph displays amplitude in the form of a gray scale value versus frequency (y) and time (x). Such types are image recorders. Nevertheless, in most biomedical recordings the content of z is limited to some discrete values, sometimes two (black and white or, more generally, marked and not marked), or more, in case different colors or marking intensities are applied. As such, the image is reduced to a set of tracings. This implies that for each value of x a set of y values is marked. Marking intensity (gray or color scale) and line thickness are insignificant with respect to signal representation and can be beneficially used for trace distinction and identification in relation to the quality of graphic layout.

In most applications, x corresponds to time and y to the physical magnitudes recorded (a t - y recorder or y - t recorder). Most attention will be paid to these types. The number of y magnitudes then stands for the number of channels (signal inputs) of the recorder. If the recorder is designed for an arbitrary abscissa input, the indication x - y recorder is used. The ideal multichannel t - y recorder produces a graphic representation of a set of time signals (Fig. 1). The effect of time is generally originated by pulling a strip of paper at constant velocity; A site has to be marked to indicate the time reference. Sensitivities and reference (e.g., zero) levels have to be known for each channel. Paper velocity, time reference, sensitivities, and reference levels are important scale factors that, apart from a few exceptions, are absolutely necessary when tracings are examined. Indeed, in specific applications some of these parameters are not required as they carry no information. A typical example is the zero level in electrophysiological measurements using skin electrodes, such as ECG and EEG. In these signals, the dc level is not significant as it is not generated by the physiological source (the heart in ECG, the brain in EEG); Moreover, they are usually high pass filtered to eliminate baseline disturbance from electrode-tissue interface potentials. Evidently, the same does not apply for such information as blood pressure and blood velocity, where the zero level is indispensable for evaluation. An accurate time reference is required only if the recorded signal is of the evoked type, that is, a response to a specific stimulus. For the ideal apparatus the recording parameters are constants; Obviously, this is but an approximation for the real recorder and its quality as a measuring instrument is determined by the constraints imposed on the deviations of the parameters with respect to their nominal values.

Analog and Digital Recorders

Depending on the transducing device applied, two categories of recorders can be considered: analog and digital. In analog recorders, a physical displacement occurs: Positioning (y) of the pen (or ink jet, light beam or other), toward the site on the paper to be marked. In digital recorders, as

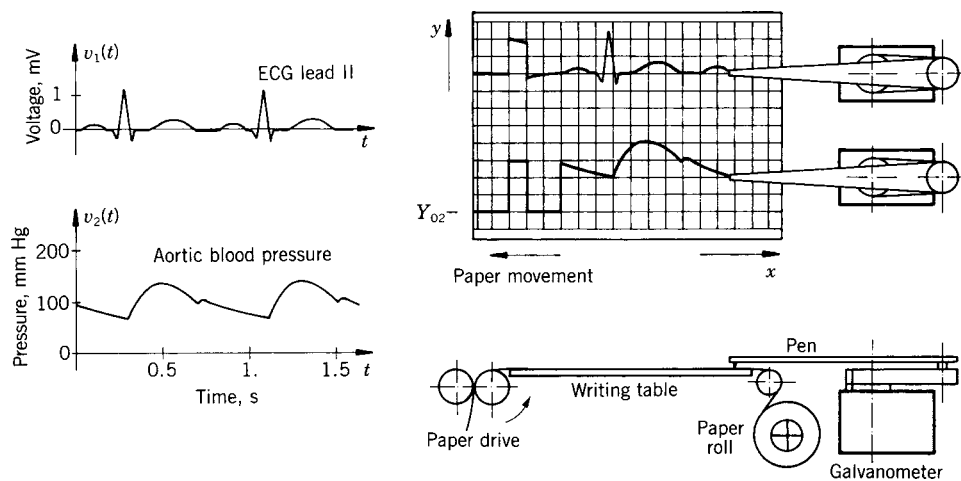


Figure 1. Time signals and graphic representation. On the left, two time signal examples are given: ECG (bipolar lead II) and aortic blood pressure. On the right, the t - y recording is sketched as executed by galvanometric rectilinear pen-writing devices on calibrated paper. For ECG and blood pressure the recording sensitivities have to be known. This is accomplished by recording a calibration pulse: 1 mV for ECG, 100 mmHg (0.133 kPa) for pressure. For pressure the zero level is obviously necessary. The time effect is generated by pulling a paper supplied from a paper roll over the writing table. Remark: Nowadays these types of signals are mostly recorded with a digital recorder (thermal array) or a PC application.

considered here, no moving parts are present. The writing device consists of a linear array of stationary writing styli, positioned at equal distances. The array is directed according to the y -axis and covers the complete width of the paper. The location y to be marked is identified by giving the address of the point in the array situated at the corresponding location.

Three elementary functions can be considered: x positioning, y positioning, and intensity control of the writing process. The most striking difference between analog and digital recorders is that in digital transducers the y positioning does not exist, as the function of indicating the y value to be marked is carried out by the intensity control on the stationary styli in the writing array. Analog recorders are capable of drawing continuous lines; digital recorders essentially put separate dots on the paper at discrete y levels corresponding to the positions of styli in the array and at discrete x values according to the incremental steps at the x input. Besides the continuous mode, analog transducers can be used in a discontinuous mode (scanning mode). In this case, the x drive is essentially similar to the one used for digital systems (a progression with equal steps) and at each step the transducer scans the paper width at high speed, putting dots (or dashes) at locations corresponding to the different signal values. As such, the digital transducers construct tracings by setting dots, the analog transducers used in the scanning mode generate tracings by setting dots or dashes directed according to the y axis. Only the analog transducer in continuous mode behaves as a real curve tracer, implying that the device is unable to generate images.

Digital transducers as well as analog transducers in the scanning mode can produce images that are composed as indicated (dots or lines). Applied as waveform recorders, both instruments have the special advantage that a single

transducer can handle a set of signals, the number being limited only by properties of resolution (dependent on the number of writing points of the digital transducer and on the recording width) and of identifiability of possibly overlapping curves. In the continuous mode a separate analog transducer is required for each channel to be recorded. To complete the discussion of analog and digital transducers, note that analog types can be used in a digital way, when positioning is bound to equal incremental steps, as applied in printers and plotters for computer graphic output. Moreover, analog recorders may be used in connection with digital systems, provided the necessary digital-to-analog conversion facilities are present. In such cases, simple digital transducers may be added to the apparatus design. A typical example is the analog apparatus equipped with printing heads for adding alphanumeric information to each tracing.

With these technological possibilities in mind, the problem of reporting the recording parameters (sensitivities, reference levels, paper velocity, timing reference) can be further discussed. Transducers capable of handling more than one signal (analog transducers in scanning mode and digital transducers) may draw calibration lines as well (ordinate and abscissa). The same is not possible for analog transducers in continuous mode. For example, for drawing a zero-level line an additional transducer would be necessary. It is obvious that this solution would be expensive. Moreover, it does not feature the benefits of accuracy as in the method where signals and calibration lines are generated by the same transducer. The problem is solved by using calibrated paper. Nevertheless, in some apparatus a separate timing marker is provided, writing small dashes that indicate time divisions (e.g., each second), or an event marker can be used to identify the start of an experiment (e.g., the stimulus if applied). Such timing devices are

evidently low cost, low performance elements and cannot be used as measuring instruments.

In the case of analog transducers and calibrated paper a zero line and a level corresponding to a standard value of the physical magnitude can be drawn before recording signal information (Fig. 1). For example, in the ECG the 1 mV pulse is commonly used. Ideally, this kind of calibration should cover the complete measuring chain: The zero line and standard level are to be applied at the biophysical sensor input. This can easily be performed in electrophysiological measurements, but for other physical magnitudes it is quite problematic, as it requires the continuous availability of a calibration setup (e.g., blood pressure measurements). In these cases one uses previously calibrated transducers and electrical calibrations corresponding to zero and standard levels of the physical magnitude are then simply provided at the recorder input (Fig. 1). Alternatively, one can draw a piece of zero line and keep the sensitivity values (physical unit/paper division) in file (written on the record). Evidently, both methods, with and without self-generation of coordinate lines, will feature different recording accuracy properties.

In the technique of analog transducing there are essentially two methods. In the direct method, the transducer generates a positioning directly determined by the electrical input. In the second method, the actual position of the writing device is measured by a sensor system and the result is compared with the recorder input value. The difference between both values is (following amplification and signal conditioning) fed to the transducer, which originates a movement tending to zero the position deviation. This second type is called the feedback, compensation, or servo method and corresponds to the well-known null detection technique in general measurement theory. Both methods, direct and feedback, are applied in analog recorders.

Chart Abscissa Generation

An essential feature of the graphic recorder is the ability of making recordings in real time. The time effect is generated by pulling a strip of paper at constant velocity and the name t - y recorder is applicable (Fig. 1). In digital recorders chart abscissa generation is performed essentially in small identical increments with the aid of a stepping motor that is controlled by a stable stepping frequency. In analog recorders, a continuous movement is envisaged and different types of motors are applied in the apparatus design (synchronous, direct current and also stepping motors). The motor can be combined with a tachometer: In this case, the actual velocity is measured through the tachometer, and with the result the motor velocity is corrected to the desired reference velocity value via a feedback circuit (null detection technique).

Paper can be fed in Z-fold or on a roll. The pulling of the paper can be achieved by sprocket wheels operating within equally spaced perforations at both sides of the paper. This method is seen mostly in low paper speed applications. At higher speeds, the paper is pressed between two rollers driven by the motor system (Fig. 1). In most cases as a result of the pulling force, the paper is pulled to the writing

table (e.g., pen writing) or to the writing head (e.g., digital systems), which is essential for thorough graphic recording. If necessary, additional pressing facilities are provided to ensure optimal contact with the writing device. Evidently the use of graphic recorders is not limited to real-time applications of the time signals. Delayed and time-expanded or compressed recordings can be realized if computing and memory facilities are provided (for digital as well as for analog recorders). For example, signals with a higher frequency spectrum in comparison with the bandwidth of the transducer, which thus fail to be plotted in real time without considerable distortion, can be recorded in this way: Data are stored at high speed in memory and are released afterward (off-line) at low speed to the recorder input.

Recorders equipped with identical input hardware for both x and y coordinates are called x - y recorders. In this case, the chart is stationary during recording. It can be supplied as separate sheets or from a roll. Different methods are used to stabilize the paper during recording (e.g., mechanical clamps, magnetic parts on an iron plate writing table), but electrostatic attraction to the writing table seems to be the most elegant method of fixation. An x - y recorder can also be used in the t - y mode by applying a linearly increasing voltage to the x input.

Recording Accuracy

Recording accuracy is the first principal quality of the graphic recorder to be discussed. This quality comprises the recorder's performance as a measuring instrument and its ability to display additional information concerning the measured tracings or images. With respect to measuring performance, the larger part of the discussion can be formulated as for other measuring instruments (1-3). Such aspects as accuracy, precision, resolution, static linearity, noise content (including drift), dead zone and hysteresis, dynamic behavior comprising frequency domain and time domain responses, and sampling and digitization effects are typical performance indicators. Parameters for both x and y axes have to be considered. It is known that energy transducing devices usually represent the most delicate functions in the chain. This is valid for biophysical sensors and it applies equally to electromechanical recording transducers. Moreover, the power amplifier driving the transducer may be critical (e.g., with respect to saturation effects). Signal amplification and conditioning modules generally play no limiting role in the overall performance. Properties of the biophysical sensors are not discussed, as this subject falls outside the scope of this article. Analog recording transducers suffer from the limitations and errors typical of analog systems, which are excluded from digital systems. In the latter case, limitations are determined by sampling (sampling frequency) and digitization (number of bits).

Accuracy is an overall parameter defined as the difference between the recorded and the true value, divided by the true value, regardless of the sources of error involved. In digital systems a number of error sources are excluded. Moreover, accuracy is largely dependent on how calibration is performed and how the recording parameters are

reported. It is evident that the recording method in which the transducer itself generates the coordinate lines (x and y) as well as the signal tracings is less subject to error effects than the method using precalibrated paper. In the first type, accuracy is simply related to the correctness of the coordinate values generated; In the second type, there are found a number of additional error sources caused by the mechanical positioning of the writing device (inherent to the analog electrical-to-mechanical conversion) and also the paper positioning. Two methods of analog transducing have been mentioned: It is known that higher accuracy can be achieved with the feedback type (null detection) than with the direct type. The same applies for the accuracy of paper velocity.

Precision of the graphic recording is related to the preciseness with which a value can be read on a tracing. It is thus dependent on the line thickness (a sharp line implies a high precision) and the paper width covered by the tracing (maximal deflection in the y direction) and, additionally, paper velocity (with respect to time readings, in the x direction).

Resolution, being the smallest incremental quantity that can be measured, is determined by the digitizing step in digital systems and the dead zone in analog systems. The dead zone usually originates as a consequence of static friction (e.g., in the case of moving, paper-contacting devices, such as pens, the friction between the paper and the writing element) or backlash in mechanical moving parts. The phenomenon also gives rise to hysteresis: A curve recorded in continuously increasing coordinates will not exactly fit the same curve recorded in continuously decreasing coordinates.

Error sources can be found in the instability of the recording parameters (sensitivity and reference level). There can be small alterations comparable to electronic noise. If the alterations occur very slowly they are called "drift". Drift can result from temperature variations. Digital transducers are not subject to drift. Nevertheless, a mechanical source of drift on the reference levels can be the shifting of the paper along the y axis. Also, the paper velocity might not be stable as a consequence of motor speed variations or paper jitter from the mechanical pulling system. It is obvious that such errors have minimal effect if the transducer itself generates the coordinate lines. This applies for drifts of the sensitivities as well as the reference levels. Noise due to electric mains interference can occur with poor channel isolation and/or poor grounding techniques. Gain and phase distortion due to impedance loading can occur if transducers are driving multiple data collection systems (i.e., chart recorder, medical monitor, VHS tape recorder, computer). As for any instrumentation system the use of appropriate preamplification may help to avoid these inconveniences.

Furthermore, the recording sensitivities may be slightly dependent on the values of the signals processed, implying that there is a deviation from strict static linearity. A specific nonlinearity problem arises in the case of galvanometric transducers where rotation has to be converted into translation. In digital transducers, linearity is determined by the accuracy of the construction of the array of stationary writing styli. Moreover, nonlinear effects are less

inconvenient if the transducer generates its coordinate lines, as these are consequently subject to the same nonlinearity as the signals. Note that in transducers with moving parts a specific nonlinearity is introduced for safety purposes. By electrical means (saturation levels) or by mechanical stops the deflection of the moving part is limited (e.g., in multichannel galvanometric pen recorders).

The dynamic behavior of the recorder refers to its response to sine waves or to transients (pulses or step functions). For analog transducers, one describes the frequency dependence of the sensitivities; A general discussion can be found in linear system theory (1–3). Some typical properties of linear systems can be mentioned. For example, a sine wave finds itself, in any case, reproduced as a sine wave at the output of the system, possibly with altered amplitude and phase. Another typical feature is that the spectral bandwidth of the system is independent of the signal amplitude; likewise the sensitivity in the bandwidth. Most linear analog recording transducers act as low pass systems: Frequencies from 0 Hz (dc) up to a certain cutoff frequency (the bandwidth) are about equally recorded. Beyond the cutoff frequency, amplitudes are progressively attenuated as frequency increases. When a step input is applied, the recorder will not exactly follow: There will be some delay, a limited rise time, and possibly an overshoot with respect to the steady-state level. These parameters (cutoff frequency, delay, rise time, overshoot) are significant in characterizing the dynamic behavior of the analog recorder. Real analog transducers behave as a linear system only within restricted limits (of deflection, slew rate, and also acceleration). Nevertheless, their performance is characterized by the same parameters. Limitations and errors are due mostly to the electromechanical transducer itself and possibly to the driving power amplifier. As for any measurement system, insufficient bandwidth will give rise to gain and phase distortion, delay and decreased slew rate. In digital transducers, where no moving parts are present, there are no errors connected to electromechanical positioning, moreover, coordinate lines are easily reproduced. In this case, limitations are determined primarily by the act of sampling and digitization. The phenomenon of overshoot is nonexistent. Delay and rise time correspond to the interval between two writing (printing) actions (the writing (printing) period). The bandwidth in analog systems determined by the -3 dB frequency (the frequency at which the sensitivity is reduced to 70% of the static sensitivity), is seen in another way: It depends on the number of samples one finds necessary to represent a complete sine wave period. If 10 is the approved number of samples for a complete period, the bandwidth is restricted to one-tenth of the writing frequency (in real-time applications). With respect to bandwidth considerations one must keep in mind that the result, the graphic record, is intended for visual inspection and that paper velocity is limited. For example, with a 100 Hz sine wave to be recorded, at high paper velocity, such as 100 mm s^{-1} , a full sine wave period covers only 1 mm, implying a poor recognizability. If the latter is essential, memory recording has to be used.

The ability to display additional information is complementary to the measuring performance. Recording

parameters (sensitivities, reference levels, time references, paper velocity) are, as already indicated in relation to accuracy, best reported in the form of calibration lines. Alphanumeric information for identification of signals and calibration lines, on the one hand, and data concerning the circumstances of the experiment (patient or subject name, date, experiment identification, stimulus type, etc.), on the other hand, are indispensable for off-line examination. They either must be manually added to the record or, preferably, must be added through the recorder itself. Digitally controlled recorders (using digital as well as analog transducers) may allow the operator to introduce the alphanumeric data together with tracings, provided the device is able to do it. In the case of digital transducers this is only a matter of appropriate hardware and software (to be supplied in the recorder) and/or interfacing with a computer. Whereas in digital transducers recording is essentially based on intensity control, in analog transducers this aspect is not necessarily provided.

Reading stored paper tracings by using waveform tracing techniques in order to be able to process data in digital format may be interesting for a number of studies. Nevertheless, attention has to be paid to the recording accuracy of older tracings and the experimental procedures used.

Graphic Quality

As mentioned in the introduction, the second of the two principal recording properties is the quality of graphics. The correctness of locating the point to be marked is part of the performance of a measuring instrument. In this section, the quality of the apparatus in producing a graphic hard copy is discussed. The clearness of the individual tracings is the first quality. Sharp and clean tracings on a bright background give the best impression. Vague or blurred lines and dirty background are not wanted. Furthermore, curves should be easily identified: Ease and speed in examination are thus improved and the risk of misinterpretation is considerably decreased. Although the use of different colors seems most appropriate, other techniques may be applied if color differences are technologically excluded, for example, writing intensity (gray scale) or even line thickness. Also, the impression of continuous lines can assist in curve examination; Adequate interpolation between sample points is thus a specific problem to be handled in the case of digital and discontinuous analog recorders (Fig. 2).

In the graphic process, two steps are involved. First, the act of writing (putting a visual mark on the paper at the located point). Second, the act of fixation (to maintain the mark for a long time). This evidently has to do with the aspects of visual inspection and permanence in the definition of graphic recorders. Graphic quality also relates to the aspect of permanence. For example, depending on the fixation process, photographic records may be affected by environmental light in the form of a darkened background.

Attention should be paid to the fact that the flow of marking matter (ink, heat, light, etc.) is to be adapted to the writing velocity in order to have optimal line thickness and grayness. In analog transducers the writing velocity is determined by the paper velocity (in the x direction) and

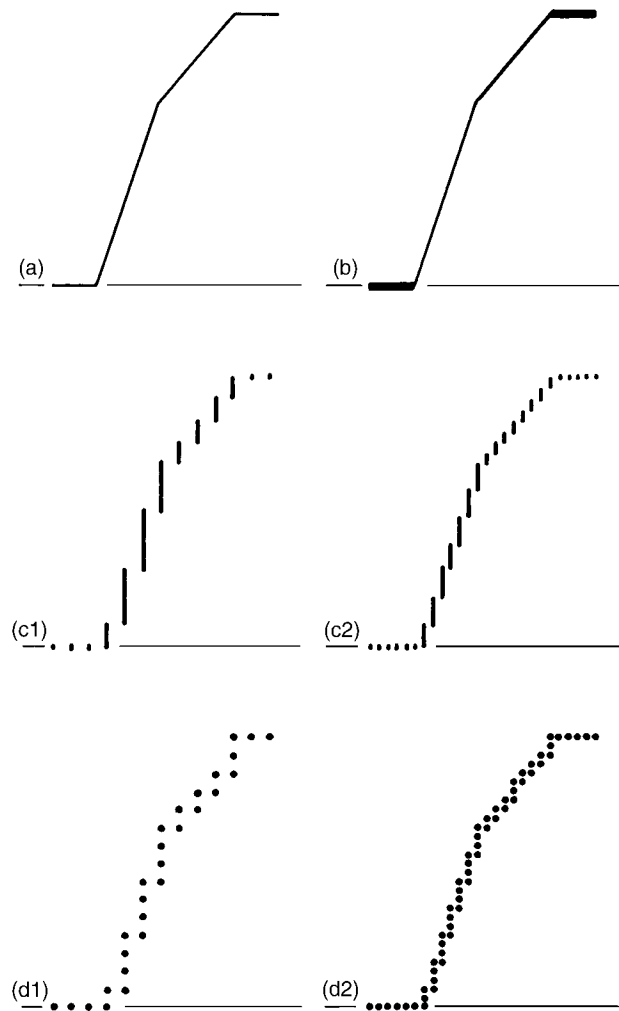


Figure 2. Graphic quality. (a) Artificial signal to be recorded. (b) Record from a pen-writing system, showing the effect of deflection velocity on line thickness. (c) Record from an analog scanning system. Interpolation is performed by vertical dashes drawn between previous and present signal values at each writing act. The effect of the writing frequency is visible in comparing (c1) and (c2) (doubled writing frequency). (d) Record from a digital system with similar interpolation. The effect of resolution is visible in comparing (d1) and (d2) (halved distance between writing points, i.e., one added to the number of bits, and doubled writing frequency).

the deflection velocity of the writing device (in the y direction). In digital recorders, only the paper velocity is involved. Adapting the marking flow is evidently a matter of intensity control (ink pressure, heating power, light intensity, etc.). In low cost apparatus, manual intervention is necessary. Adaptation might not even be possible. In more sophisticated apparatus, writing matter flow is automatically adjusted in relation to paper velocity. In the steady state, then only alterations in intensity and line thickness are discernible in the case of variations of deflection velocity (Fig. 2). This explains why, in typical ECG pen recordings, the baseline appears much heavier compared with the slopes in the QRS complex. One could say that this side phenomenon has a beneficial effect on waveform recognizability since information on the signal

time derivative is also included in the graph. Evidently, this statement may be subjective. Some apparatus are even equipped with an intensity control that continuously adapts the writing flow to the total writing velocity. The latter is possible only if the response of the writing system is fast enough. A good example is the thermal system with a pen-tip writing device. Because of the small size of the heating resistance in the tip of the pen, the thermal time constant can be reduced to a few milliseconds and fast heating flow adaptation is possible. In the thermal edge writing device, the resistance producing the heat is much larger. The thermal time constant is then ~ 1 s so that the latter may even result in a nonnegligible waste of paper when the instrument is started at high speed paper velocity.

When ink is used, the paper can be marked by a contacting device, a pen, or by a noncontacting device, an ink jet. Pens can be designed as capillary tubes connected to a pressurized ink reservoir. They can be fiber-tip pens or ballpoint pens (disposable pens) and ink can be supplied from a common container or from (disposable) cartridges connected to each pen. In the case of ink jets, there is no contact with the paper. Ink is squirted as a very thin jet from a fine nozzle directed toward the paper. Fixation occurs as the ink is absorbed and dries. Ordinary paper can be used, satisfying the needs of absorbency quality and surface smoothness (in the case of contacting pens). A disadvantage with ink systems is the possibility of ink stains and smears, as fixation (absorbency and drying) does not occur instantaneously. A practical problem is that pens may clog or "bleed."

The following methods require special paper types. These can be composed of an appropriate paper base coated with a special layer functional with respect to the writing process (e.g., thermochemical). In the case of the burning method, an electric current from a stylus tip in contact with the paper is passed through the electrically conductive paper to the metal support on which the paper is mounted. The current gives rise to paper burn and thus blackening. The density of the burned mark depends on the current magnitude. Thus, a gray scale effect can be obtained. As burning is irreversible, the graphics remain fixed. This method finds application in the previously mentioned sound spectrograph. Furthermore, marking can be executed by heat on thermosensitive paper. The writing effect can be originated by thermally melting off a white coating from the chart paper and thus exposing the black underlayer. In another process, the base paper layer is covered with a thermochemical layer that irreversibly changes color after exposure to heat. In thermal writing, the color is usually black, but other colors can be generated. The shelf life of thermal chart recordings can vary on paper, transducer, and user storage techniques; Thermal recordings may be degraded by exposure to heat, light, friction, solvents, and so on.

In both electrostatic and photographic methods, the writing act comprises two phases. First, a latent invisible image is transferred to the paper. Second, the image is developed and thus made visible. In the electrostatic method, paper coated with a special dielectric layer is locally charged (the latent image) with electrical charges.

These charges attract, in a second step, ink particles from a toner supply (fluid or dry) after the excess toner is removed only ink is left at the electrically charged sites. Fixation is then achieved by drying (in the case of toner fluid with ink particles in suspension) or heating. In the photographic method, which uses photosensitive paper (the most expensive type), a light beam (generally ultraviolet, UV) activates the photolayer, leaving a latent image. Graphics are then visualized and fixed according to the photographic process used by exposure to visible light (environmental light or an additional light source specially provided) or by heating. Dependent on the process used, long exposure to environmental light may degrade the quality of the record after the examination by darkening the background. A darkened background may also appear in the electrostatic writing process if the apparatus is not optimally adjusted (e.g., if not all of the excess toner is removed). The photographic method is best fit for the production of images, but it has also been used for tracings.

Some commonly used writing techniques have thus been described. Acts of marking and fixing are typical technological problems and this area continues to evolve. The method used in a specific apparatus depends on different factors. With respect to the application envisaged and the type of transducer used, specific writing techniques may appear advantageous, but the manufacturers' patents also play an important role in apparatus design. The paper cost is certainly not to be neglected (e.g., in long-term recordings). Electrostatic and photographic writing have lost interest in modern chart recorder design.

A remark has to be added with respect to graphic quality as well as precision. As already mentioned, recorded tracings should be optimally identifiable, especially when they cross each other on the record. The use of different colors is obviously beneficial. In this respect, the paper width and the usable range for one channel are important. Traces can be limited within separate ranges or can cover the complete paper width. Precision is maximal in the latter case, but there are problems involved. Some transducers, more specifically galvanometric devices, produce only limited deflections. Furthermore, there is a problem with contacting devices (pens). Since it is physically impossible that they cross each other, they are spatially shifted with respect to the abscissa of the paper, which corresponds to a desynchronization on the records. If the complete paper width is covered, precision is maximal, but identifiability is decreased, especially when the use of different colors is excluded as in those devices where the color is not determined by the writing element, but by the paper properties. The latter applies to all devices other than ink-writing devices.

Another remark concerns immediate visibility of real-time recorded tracings. Although the recording may be preferably in real time, the visibility may not have the same benefit: More specifically, tracings can be seen after a small time delay. This is the case if the writing device is positioned within the recorders housing; The time delay is then dependent on the distance to the visual part of the paper and, evidently, of the paper velocity. Digital transducers have this shortcoming: pen recorders do not, except in the multipen recorder with dislocated pens (allowing

overlapping graphs). In this case, a pen offset compensation may be used to synchronize recorded tracings, also causing a time delay.

In this Paragraph a number of technological aspects were just mentioned. Some of these concepts have lost interest with respect to new design, but may still be found in older apparatus.

ANALOG RECORDERS

An analog recorder has been defined as a recorder that applies an analog electromechanical transducing principle. The electrical magnitude is transduced into a translational value, that is, positioning of a movable part, such as a pen, toward the site to be marked. Although analog recorders are used mostly for signals in analog form, they may also be applied to digital signals and consequently to computer output, provided digital-to-analog conversion facilities and appropriate control access (paper velocity, intensity control) are available.

For the description of the transducer, analog system theory is applicable (1–3). Two typical measurement principles are commonly used in transducer design. In the direct method, the electrical magnitude is directly transduced into a displacement magnitude. In the feedback method, corresponding to the null detection technique, the actual position of the marking device is measured with an accurate sensor. The difference (the error) between the measured and the input value is, after appropriate conditioning, fed back to the transducer, which generates a movement tending to zero the position error. In the latter case, the accuracy of the transducing system depends on the quality of the position sensor. It is known that with this method higher accuracies and more stable transducing properties can be achieved. Furthermore, an analog transducer can be applied in the continuous mode or the discontinuous mode. In the first case, a single continuous tracing is generated by each transducer; The number of channels is equal to the number of transducers provided in the multichannel recorder. Production of an image is excluded. In the second case, the writing device is repeatedly swept over the complete paper width, setting dots or dashes at sites corresponding to the signal values (the scanning mode). A single transducer can thus handle a set of signals. It is obvious that what is gained by the ability to process different signals is lost with respect to the writing speed of the system.

Positioning of the writing device can be caused by a transducer that causes a rotation (the galvanometric transducer) or a translation (the translational servosystem). The translational servotransducer evidently makes use of a feedback mechanism. The galvanometric transducer, originally strictly built as a direct device, has also been designed according to the feedback principle.

Whereas two decades ago a number of analog transducer techniques were applied in apparatus design for biomedical signal recording, mostly for obtaining different bandwidths and image recordings (as for echocardiography), only a few of them are still used. The galvanometric recorder and, more importantly, the translational servorecorder.

Galvanometric Recorders

Galvanometric recorders make use of transducers that are essentially rotational transducers, making use of the d'Arsonval movement as applied in ordinary galvanometers (4). A positioning, more specifically a rotation over a certain angle, is obtained as a result of an equilibrium between two torques, the first an electromagnetic torque, proportional to current, the second a mechanical torque, proportional to positioning (a rotation angle). A coil wound in a rectangular form is suspended within the air gap between the two poles of a permanent magnet and a stationary iron core. A current flowing through the coil gives rise to an electromagnetic torque that tends to rotate the coil. To position at a specific angle θ , there must be a restoring torque, essentially proportional to the angle, for example, a torsional spring characterized by its rotational stiffness. The signal to be recorded is fed to a current amplifier that drives the coil. As such, there is direct proportionality between the angle and the signal. This principle corresponds to the direct transducing method.

In a discussion of dynamic behavior, not only stiffness, but also damping (viscous friction) and inertial moment (of the coil and the attached mechanical parts, e.g., the pen) must be taken into account. As a first approximation the device acts as a second-order linear system, characterized by a resonant frequency and a damping factor. The resonant frequency depends on the ratio of stiffness divided by inertia. The resonant frequency and the damping factor determine the bandwidth of the system (with respect to sine wave response) and the time delay, rise time, and overshoot (with respect to step input transient response). For frequencies sufficiently above the resonant frequency, the corresponding amplitudes are attenuated proportional to the square of the frequency. A second-order system can show resonant phenomena, depending on the value of the damping factor. If this factor is <1 , the response of the transducer to a step input shows a ringing effect, implying that an overshoot exceeding the steady-state value has occurred. The smaller the damping, the smaller the rise time, but the higher the overshoot. At critical damping (damping factor = 1) no overshoot is present. Such an overshoot is considered inconvenient as it gives an erroneous impression of the waveform (especially if sudden alterations in the signal occur). With the damping factor in the proximity of 0.7 a good compromise is obtained between the overshoot (4%) and the rise time (3) (Fig. 3). The useful bandwidth of the recorder is thus determined by the resonant frequency and thus by the ratio of stiffness to inertia. To obtain a sufficient damping factor, extra damping must be applied within the galvanometer (mechanical or electromagnetic). In the case of paper-contacting devices, the damping should exceed the effect of the static friction of the writing element (e.g., the pen) on the paper (as this effect is unreliable and consequently is not allowed to affect the measuring performance).

The feedback technique has also been used. According to the general idea, the actual angle is measured by a position-sensing device, delivering an output proportional to the angle, which, after amplification and waveform conditioning, is negatively fed back to the input of the

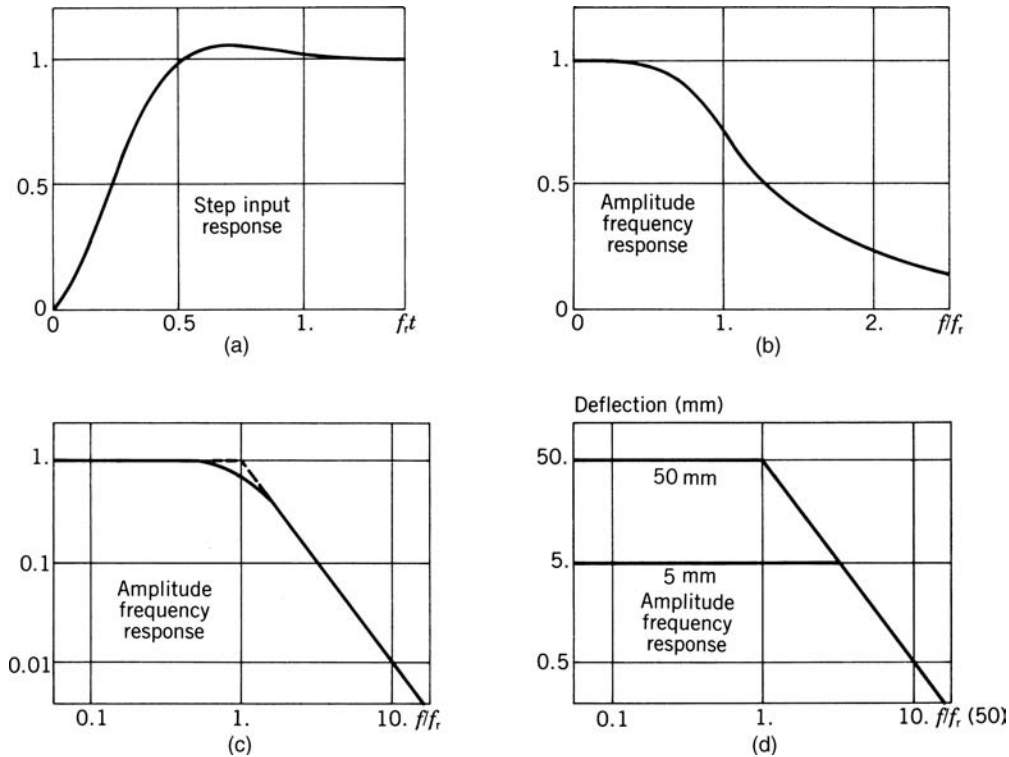


Figure 3. Characteristics describing dynamic galvanometric transducer responses. (a–c) Linear system behavior; Damping factor 0.7. (d) Nonlinear behavior. Coordinates are given in generalized form (time = $f_r t$; frequency = f/f_r) (f_r = the transducer resonant frequency). (a) Time response to a unit step input. Delay, rise time, and overshoot are visible. (b) Amplitude frequency response (sine waves) in linear coordinates. (c) Amplitude frequency response in logarithmic coordinates. (d) Typical amplitude frequency responses (50 and 5 mm deflection) for a nonlinearly behaving system. The bandwidth appears to depend on the tracing amplitude (displayed: inversely proportional to the square root of the amplitude).

current amplifier driving the galvanometer. In case the position signal simply undergoes amplification, the effect of the negative feedback is a restoring torque proportional to the angle, thus resembling the effect of a rotational spring in an apparent stiffness. An equilibrium is reached when the coil angle (as measured) corresponds to the value at the input of the system. At equilibrium, the current through the coil equals zero (provided one does not take into account the effect of an additional mechanical spring). This is advantageous with respect to the current amplifier.

A real galvanometer usually does not feature constant linear system properties, such as stiffness (mechanical or apparent, as generated by feedback) and damping, independent on deflection, deflection velocity, and acceleration. Particularly with the feedback system, nonlinear effects can appear as a result of current saturation. In a linear system, the bandwidth is independent of the signal amplitude; In a real galvanometer, the same usually does not apply. For larger sine wave amplitudes, the useful bandwidth appears decreased and sine wave distortion occurs at frequencies in the vicinity of the apparent resonant frequency. A bandwidth inversely proportional to the sine wave amplitude to the power of m , with the exponent m approximately between 0.5 and 1, is obtained (Fig. 3).

Apart from nonlinearity in dynamic behavior, a problem with respect to static linearity arises. The galvanometric recorder essentially generates a rotation, not a translation, as one would expect for a graphic recorder. If a pen is connected to the coil of the galvanometer, the tip of the pen rotates with a radius equal to the pen arm length. Such a recording is curvilinear. It has been used on paper with curvilinear coordinate lines. Rectilinear recording is obtained with special techniques. Whereas in curvilinear recording mostly ink pens are used (a low cost solution), there are different methods for obtaining rectilinear records: pen methods (long-arm pens, knife edge recording, mechanical rectilinearization), ink jet method and light beam method) (Fig. 4). Curvilinear errors can be kept small if the radius is large, which is the case with long-arm pens. Knife edge recording is the simplest real rectilinear recording. The chart paper is pulled over a sharp edge, accurately directed according to the y -axis, and the writing stylus moves over and presses on it. The impression is thus made at the site of the edge and thus rectilinear. In this case, the marking method applied is thermal. It should be remarked that in this method a large part of the stylus has to be heated (as compared to the thermal point writing), giving rise to large thermal time constants (order of magnitude 1 s). This is inconvenient for high paper speeds as, when

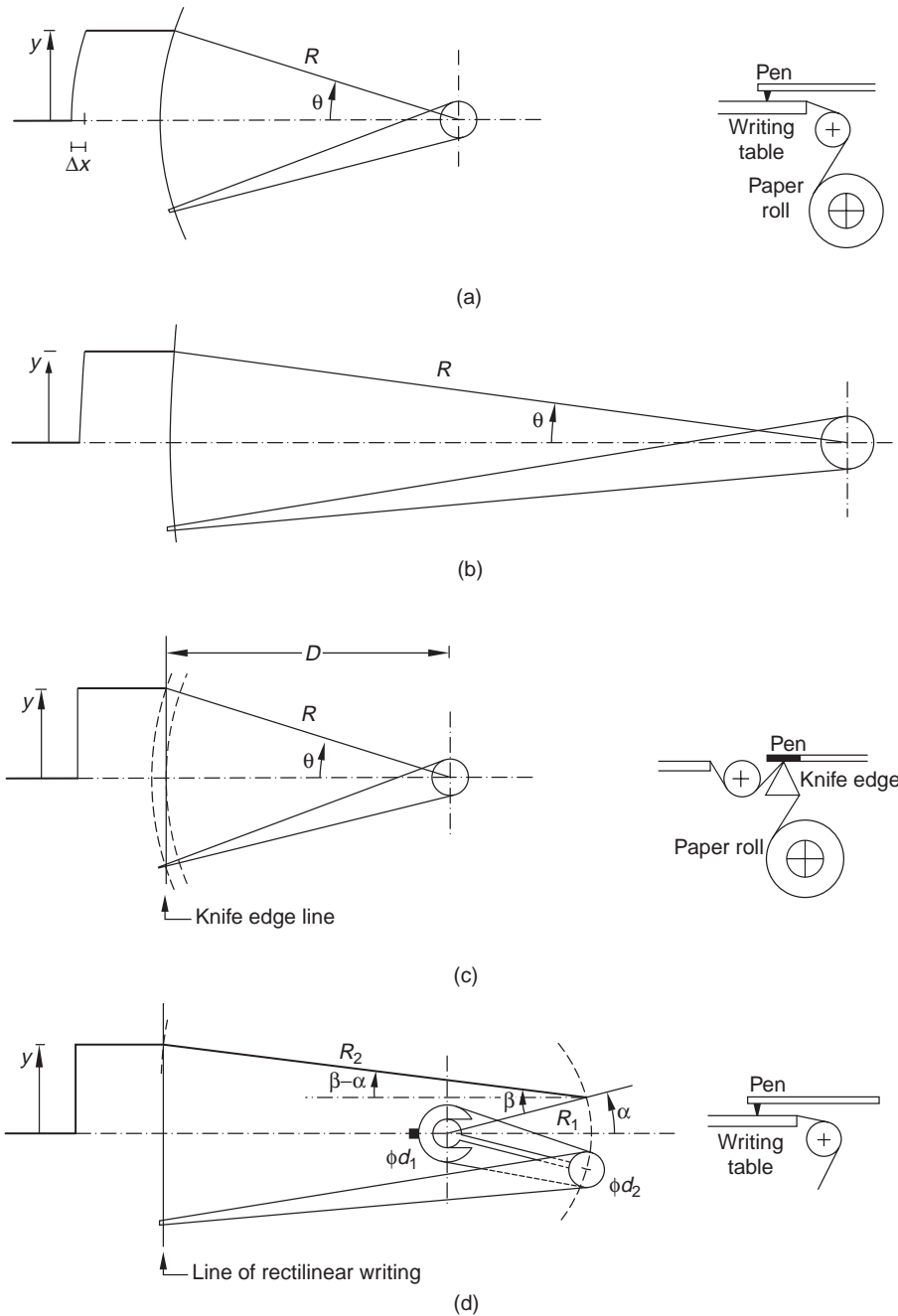


Figure 4. Galvanometric pen-recording assemblies. An artificial recording is shown corresponding to a zero level, abrupt transition to maximal positive, holding, and finally abrupt transition to maximal negative deflection. Errors resulting from dynamic behavior are not considered in this figure. (a) Curvilinear recording. (b) The long-arm pen recording is curvilinear but, as a consequence of restricted angle deflection, approximates rectilinear recording. (c) Knife edge recording. Rectilinear recording as the writing essentially occurs at the site of the knife edge, where the pen presses on the paper. For thermal writing, the black end of the pen represents the part to be heated (minimal length indicated by the dashed lines). (d) Mechanical linkage for rectilinearization. The pen arm (length R_2) is connected to a pulley (diameter d_2), which is allowed to rotate at an axis at the end of the galvanometer arm (length R_1). A metallic belt attached at a point of the pulley is also fixed at a stationary circular part (diameter d_1) at the galvanometer frame. As the galvanometer arm rotates, the pen arm is subjected to a corresponding rotation dependent on the diameter ratio: $\beta = (d_1/d_2)\alpha$. When $\beta = \alpha$, the pen arm is moved parallel to the zero position and the pen point moves in the same curvilinear direction as the end of the galvanometer arm. Thus β must be larger than α , and more specifically so that the rotation through the angle $(\beta - \alpha)$ the curvilinear abscissa error of the galvanometer arm end is corrected for. A power series expansion shows that this is obtained when $R_1/R_2 = [(d_1 - d_2)/d_2]^2$.

the instrument is started, a significant amount of thermo-sensitive paper may be wasted. Instead of a hot stylus, a carbon ribbon with an ordinary stylus pressing on it has been used. In this case, ordinary paper can be used, but an additional mechanical device linked with the paper velocity equipment for driving the carbon ribbon is necessary. Rectilinear recordings can also be generated via mechanical linkages that compensate for the normal curvilinear movement of the pen and convert the rotary motion of the coil into an (approximately) straight-line motion of the pen tip. In the case of ink jet recorders, the writing device does not make physical contact with the paper. In this case, a fine ink jet is produced by pumping ink through a nozzle connected to and thus rotated by the galvanometer toward

the paper. The recording is rectilinear at the intersecting line of the paper plane and the plane of the rotating ink jet. The same holds for the optical method, where a sharp light beam is reflected by a small mirror connected to the rotating coil of the galvanometer. In this case (expensive), photosensitive paper has to be used.

The principal difference in the abilities of these types of galvanometers is in the achievable bandwidth. Galvanometric pen-writing systems have to produce a considerable torque, as the pen, being pressed to the paper for graph production, must be moved easily without interference from the static friction in the recorder performance. This implies a large driving current and, for a defined angle, a large stiffness. This stiffness, together with the amount of

inertia, determines the resonant frequency and thus the bandwidth. A typical bandwidth value for a galvanometric pen system is 100 Hz. Theoretically, a higher bandwidth could be obtained for the same inertia, but this would imply a larger stiffness and thus a larger current, the latter being limited by its heating effect on the coil. As this paper-contacting problem is nonexistent in ink jet and light beam galvanometric recorders, current and stiffness can be lower. Moreover, the coil assembly can be made with such low inertia that, notwithstanding the lower stiffness, a much higher resonant frequency can be obtained. An order of magnitude for ink jets is 1000 Hz. For optical systems, 10 kHz has been reached.

The static friction in the case of paper-contacting writers give rise to a dead zone: A zone in which input voltage can be altered without causing any pen movement. It thus determines the resolution of the pen recorder. Hysteresis phenomena can consequently be observed. A curve recorded in continuously increasing coordinates will not exactly fit the same curve recorded in continuously decreasing coordinates.

The paper width covered by galvanometric pen writers is usually small (40–80 mm) as a result of the angle limitation (with respect to linearity) and pen arm length restriction (with respect to inertia and consequently bandwidth). Moreover, as overlap of tracings is physically impossible (the pens might strike each other) in multichannel recorders, a limited part of the paper is assigned to each channel.

As already mentioned, galvanometric recorders have lost interest. Pen recorders (e.g., edge recorders) can still be found on the market, but the types with higher bandwidth (ink jet and light beam) have been replaced by digital (memory) recorders.

Translational Servorecorders

In the translational servorecorder, the writing device (usually an ink pen) undergoes a real translation as it finds itself bound to mechanical straight-line guidance. The translation is generated by a motor and an appropriate mechanical linkage composed of a wheel and closed-loop wire system (Fig. 5). It is essentially a feedback method. A position sensor supplies a voltage directly proportional to the position of the pen, that is, the distance with respect to its zero position. In most apparatus, this sensor is a rectilinear potentiometer; This explains the use of the alternative nomenclature (potentiometric recorder) for this type. A further description of its function agrees with the general feedback principle. The input signal is compared with the pen position value and the amplified and conditioned difference voltage drives the servomotor, causing the pen to move (rectilinear) until the position value equilibrates with the input signal. The ink pens used are capillaries, fiber tips, or ballpoint types, generally supplied with ink cartridges. Evidently different colors can be used for optimal signal discrimination. In modern designs, digital servodevices are being used, with position reading by optical or ultrasonic means.

The static linearity of the recording is determined by the linearity of the position sensor. Noise can result from problems with the sliding contact on the potentiometer

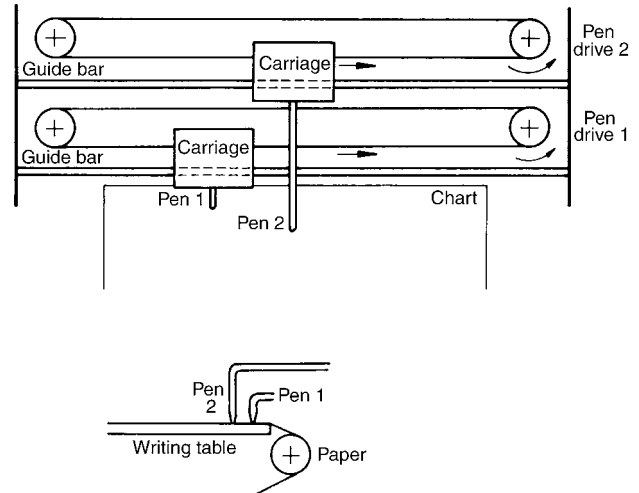


Figure 5. Translational servo t - y recorder. Two channels are shown. Pens (1 and 2) undergo real translations as they are connected to carriages that slide along guide bars. To be able to cover the complete paper width the pens are mechanically staggered with respect to the paper abscissa.

resistance. Hysteresis can be caused by the properties of the servosystem (e.g., static friction) and the mechanical backlash between the pen tip (the real location) and the wiper on the potentiometer (the location processed in the feedback loop). As for dynamic behavior, one can state that the servorecorder generally does not behave as a linear system (except for small amplitudes). It is characterized by a limited slewing velocity (order of magnitude: $0.5\text{--}2\text{ m}\cdot\text{s}^{-1}$) dependent on the type of motor and mechanical linkage. Also, the acceleration of the moving parts is subject to limitations (order of magnitude: $50\text{ m}\cdot\text{s}^{-2}$). If one of these limits is reached, the device ceases to act as a linear system and, as described for galvanometric recorders in the feedback mode, the bandwidth depends on the tracing amplitude (order of magnitude of the cutoff frequency: $1\text{--}5\text{ Hz}$). As such, these recorders are fit for slowly varying parameters (temperature, heart rate, mean blood pressure, respiration, etc.). They are generally applied for higher precision: The common paper width is 200 or 250 mm. A writing control is normally provided in the form of a pen lift (on-off functioning) by electromagnetic and/or by manual means.

Servorecorders are used as strip chart recorders (t - y recorders) and also as x - y recorders. In the latter case, two servosystems are assembled. The x system drives a long carriage, directed parallel to the y axis and covering the complete width of the paper in the y direction. Hereon, the second servosystem y , which eventually positions the pen, is mounted. Paper is fixed on the writing table (preferably by electrostatic means). Driven by computer output these x - y recorders can plot arbitrary graphs (plotters). When addition, alphanumeric information and coordinate lines can be added if the pen lift control is used. Some x - y recorders also have the built-in facility of t - y recording (with a ramp signal at the x input).

In the case of the t - y recorder, several channels (1, 2, 4, 8, 12 channels) can be assembled, all of them covering the

complete paper width, provided the pens are mechanically shifted with respect to their abscissa position (Fig. 5). This mechanical offset, evidently corresponding to a shift in time on the graph, may be inconvenient if values of different signals at a specific time instant have to be compared. To overcome this disadvantage, most servorecorders can be supplied with a pen position compensation unit. All channels, except one corresponding to the first positioned pen along the time axis, are digitized and stored in memory. Data are released, converted to analog, and supplied to their corresponding channels with a time delay equal to the distance from the pen to the first pen divided by the paper velocity. This compensation is evidently beneficial for examination afterward, but can be inconvenient at the time the experiment is executed as the information is plotted only after a time lag dependent on the paper velocity (except for the first pen).

Some modern recorders have data-acquisition facilities. Data (signals and instrument settings) are digitally processed and can be stored, for example, on floppy disk or on a memory card. Communication with computers can be done via standard interfacing (GP-IB, RS-232C, IEEE-488). In some apparatus, a display is provided allowing visualization of recorder settings. As slowly varying signals are envisaged, sampling frequencies used range from 100 to 400 Hz.

DIGITAL RECORDERS

A digital recorder has been defined as a recorder that uses a digital transducer. In this case, no moving parts are present and the transducer consists of a stationary straight-line array of equally spaced writing styli covering the complete chart width (Fig. 6). As such, the nomenclature array recorders can also be used. Progression along the x axis is essentially discontinuous. The paper is held stationary during the writing act. At a given x position, a set of writing points is activated. The paper is marked with dots at the sites in intimate contact with these points. The resolution with respect to the y axis depends on the density of writing points. The resolution in time (x axis) depends on the writing (printing) frequency (the inverse of the writing

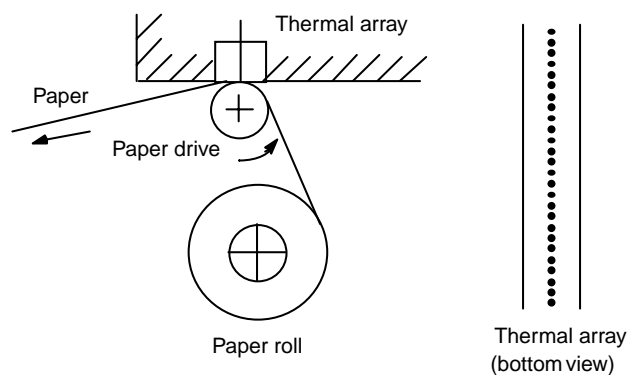


Figure 6. Thermal array recorder. The thermosensitive paper is pressed to and pulled over the thermal array by the paper driving roll.

interval) and on the paper speed. Signals are thus reproduced as discrete values at discrete times. Analog as well as digital signals can be processed. In the analog case, sampling and digitization are part of the recording process.

At present, the borderline between digital graphic recorders, digital oscilloscopes, and data-acquisition systems becomes unclear as a number of the latter instruments are also provided with graphic recording facilities with strip charts. Some digital recorder designs primarily focus on fast acquisition of data and storage in memory for reproduction afterward. These are indicated as memory recorders and generally their real-time properties do not match the fast graphical recording types. Digital graphic recorders have a number of properties regarding data-acquisition, monitoring, storage, and communication with computers.

The act of writing of the digital recorder is generally thermal on thermosensitive paper. Also, an electrostatic method has been used. The writing transducer is essentially a rectilinear array of equidistal writing points covering the total width of the paper. These writing styli consist of miniature electrical resistances that are activated, that is heated, by a current pulse of short duration. Typical resolutions are 8 or 12 dots·mm⁻¹ according to the y direction. Static linearity of the recorder thus depends on the quality of the construction of the array. Inconveniences resulting from mechanical moving parts are not present. There is no overshoot. Delay and rise time are equal to one writing interval. All processing occurs in digital form and recording is a matter of generating the correct addresses of writing points to be activated.

Evidently, the dynamic thermal properties of the resistances in close contact with the chart paper and the shape of the electric current pulse set a limit to the writing frequency. A typical value for the writing frequency is 1.6 kHz. In a memory recorder, it may be lower. Local heating of the thermosensitive paper results in a black dot with a good long-term stability, dependent on the paper quality. Evidently, thermal paper is sensitive to heat, but also care has to be taken with pressure, light, solvents, glue, and so on, in order to prevent deterioration of the graph. Generally, the heating pulse is controlled with respect to chart velocity in order to obtain an appropriate blackness and line thickness at different velocities. Nevertheless, at the highest velocities the print may become less black and sharp. Although possible within a limited range, blackness and thickness of tracings are seldom used for the purpose of trace identification.

Whereas the resolution according to the y axis is determined by the thermal dot array, the x -resolution is limited by the dynamic properties of the array, and thus by the paper velocity. At low velocities, generally 32 dots·mm⁻¹ (exceptionally 64 dots·mm⁻¹) are applied. At the highest velocities used (100 mm·s⁻¹, in some apparatus 200 and 500 mm·s⁻¹), the number of dots printed is determined by the writing frequency (e.g., 1600 Hz results in 8 dots·mm⁻¹ at 200 mm·s⁻¹).

Data acquisition is determined by the sampling frequency, expressed per channel recorded. According to the sampling theorem this should be at least twice the highest signal frequency in order to prevent aliasing

effects, but for obtaining a graph with minimal graphic quality at least 10 points are needed for a sine wave period. On the other hand, sampling frequency should be limited in order to prevent an excess of data stored in memory (if applicable). Sampling frequencies used are typical 100 or 200 kHz (exceptionally 500 kHz and 1 MHz). A basic component of the digital recorder is RAM memory in order to create high quality graphs. The sampling frequency is generally higher than the writing frequency. If the sample frequency equals the writing frequency, the recorder prints all dots between the former and the present value. In fact, curves composed only of points are difficult to examine and interpolation improves the impression of smoothness of the sampled curves. If the sample frequency is higher than the writing frequency, the recorder prints all the dots between the smallest and the largest values (including the former value within the writing interval). As such at each writing act vertical lines are drawn. In this way, fast transients, disregarded by the ordinary interpolation between former and present value, are also captured in the graph. Although the form of the transient cannot be interpreted, the graph provides evidence of its presence. A signal having a frequency larger than the writing frequency is consequently displayed as a continuous black band.

Typical digitization levels are 8 or 12 bits, exceptionally 16 bits. Widths of thermal arrays range from 100 to 400 mm. With a resolution of 12 dots-mm^{-1} , the largest number of dots full scale is then 4800. In case 16 bits are used, a signal with a large dc offset can be reproduced with excellent graphic quality as the offset can digitally be removed and the scale adapted.

Coordinate lines (with identification) can be generated and precalibrated paper is unnecessary. Errors resulting from paper shifting are thus eliminated. Tracings can cover the full width of the paper. Tracing identification in alphanumeric form and comment on experiments performed can be easily added through an internal or external keyboard. Signals can be calibrated and real physical values and units can be printed at the calibration lines. Simple mathematical calculations can be performed on signals in real time and results recorded in virtual channels. In most performant recorders a screen, mostly color LCD, is provided, allowing display of tracings in different colors and settings of recording parameters. Monitoring signals certainly has advantages. In real-time recording signals are not directly visible as a result of a delay corresponding to the distance between the internal thermal array and the visible part of the paper, but can be observed without delay on the monitor. Recording parameters (gain, offset, ...) can be set using the monitor and thus avoiding paper spoiling. Signals stored in memory can be viewed before recording on paper. Cursors on the display can be used for reproduction of selected parts of signals. Some recorders have a built-in display (up to 18 in., 45.72 cm). Others can be connected with an external display.

Digital recorders, especially memory recorders, are able to store recorded data and apparatus setting. Besides RAM for fast storage, recorders may have a built-in hard disk, floppy, magneto-optical or ZIP disk drive and a slot for a memory card. Connections may be provided for external

memory media. Signals can thus be reproduced from memory, processed and recorded on chart. Time compression or expansion is obviously possible. As data can be sampled at a high rate, stored in memory, and then recorded as they are released at a reduced rate, a bandwidth higher than that determined by the writing frequency can be achieved. In this way, fast transients that cannot be handled on-line can be accurately reproduced. Data capture and/or graphic recording (with an appropriate speed) can be controlled by trigger functions (external or derived from recorded signals); Pre- and posttrigger may be chosen. Via memory x - y plots can be obtained. When appropriate software is available, FFT plots can be made.

Digital recorders can have analog and logic channels. Typical numbers of analog channels are 4, 8, 16 (and exceptionally 32 and 64). General monopolar or differential amplifiers or signal conditioners with analog antialiasing filters can be plugged-in, as well as special purpose amplifiers for biomedical purposes. Furthermore interfaces can be provided for computer connection (RS-232C, IEEE-488, GPIB, Ethernet, USB, SCSI) or external hardware. Software can be provided for control, data transfer and conversion of data to formats for popular data analysis software programs.

EVALUATION

Table 1 shows a list of manufacturers of recorders, website addresses and typical products. Analogue recorders with a rotating pen arm, analog recorders with a translating pen and digital thermal array recorders. Evidently, the list is incomplete and continuously changing. For extensive practical details on recorders, signal conditioning modules, displays, data storage and processing, computer interfacing, and so on, the manufacturers' data sheets should be consulted.

Analog recorders are characterized by their specific technological limitations: the bandwidth, restricted mostly by the inertia of the moving parts, imposing constraints on the frequency content of the signal; the paper width covered by a trace, the number of channels, and the trace overlap (including pen dislocation); the stability of the recording parameters (including shifting of calibrated paper) and the difficulties of adding coordinate lines and alphanumeric identification to the records. The use of auxiliary devices (timing marker, printing head, pen position compensation, etc.) has been shown to overcome some of the limitations. The bandwidth is considered one of the most important limitations and thus the signal's spectral content determines the choice of the recorder type. Galvanometric pen recorder types (100 Hz bandwidth) have lost interest but some types can still be purchased. A number of mechanisms have been applied in commercially available devices, but in many applications analog types have been replaced by digital types or PC set-ups. Translational servorecorders have kept their position in their specific domain; They can be used for slowly varying signals with a signal spectrum up to 5 Hz (body temperature, heart rate, mean blood pressure, laboratory applications, etc.).

Table 1. List of Manufacturers of Graphic Recorders with Websites and Products^a

Manufacturer	Rotating Pen	Translating Pen	Thermal Array
Astro-Med, Inc. http://www.astromed.com	Model 7 (ink)		Dash 18 Dash 2EZ Dash 8X Everest
Western Graphtec, Inc http://www.westerngraphtec.com	WR3310 (tp) WR7200 (tp)	WX3000/ WX4000 (dig) (xy/ty)	WR300 WR1000 WR8500 DMS1000
Hugo Sachs Elektronik – Harvard Apparatus GmbH http://www.hugo-sachs.de	Mark VII-c (tp)	R-60	
LDS Test and Measurement LLC http://www.gouldmedical.com			TA11 TA240 TA6000 WindoGraf
Yokogawa Electric Corporation http://www.yokogawa.com/daq/daq-products.htm		3057 LR4100E/ LR4200E/ LR8100E/ LR12000E (dig) 3023/3024 (xy/ty) 3025 (xy/ty)	OR100E/ OR300E DL708E/ DL716 (ds) DL750 (ds)
Hioki USA Corporation http://www.hiokiusa.com			8826 (m) 8835-01 (m) 8841 (m) 8842 (m)
Kipp&Zonen http://www.kippzonen.com		BD11/12 SE 102/122 SE 110/111/112 SE 124 BD300 (dig) SE790 (xy/ty)	SE 520/540 SE 570
Soltec http://www.solteccorp.com		MCR 560 DCR 520 (dig) DCR 540 (dig)	TA200-938 TA200-3304 TA220-1200 TA220-3216/3208 TA220-3424 TA220-3608
Omega Engineering, Inc http://www.omega.com		142 156 555/585/595 640 RD45A/46A RD1101 RD1201/1202 RD2000 RD3720 (dig) RD6100 RD6110 RD6112 600A (xy/ty) 790/791(xy/ty) RD3020 (xy/ty)	
Linseis http://www.linseis.net		L120/200/250 L6514II L7005II LY14100II (xy/ty) LY15100II (xy/ty)	

^aRecorders with analogue transducers with rotating pen and with translating pen (servo) and recorders with digital transducer (thermal dot array). ink: inkpen; tp: thermal pen; ds: digitals scope with chart recorder; m: memory recorder; xy: x-y recorder; xy/ty: x-y and t-y recorder; dig: digital signal processing.

Digital recorders are characterized by the typical features of sampling and digitization in waveform reproduction. Although the writing frequency is limited by the thermal time constant of the thermal dots, with respect to bandwidth, it does not impose a constraint on the signal frequency spectrum, because, by the use of facilities inherent to digital apparatus (including a high sampling frequency), the problem can be solved by off-line recording. Signals that vary too fast for on-line recording can be stored in memory and reproduced at a speed the recorder is able to handle. Moreover by the specific writing method where vertical dotted lines are printed varying from the minimal to the maximal value sampled within the writing interval, fast transients, disregarded by ordinary interpolation, are also captured in the graph. Although the form of the transient cannot be interpreted, the graph provides evidence of its presence. In digital recording, the addition of coordinate lines and alphanumeric information to the record presents no difficulty for the transducer, as there is no essential difference compared to recording signal tracings. As connections to computing devices are possible, processed data can also be recorded, reducing the work of visual inspection of the record. Also, in some translational servorecorders signals are processed completely digitally, but in this equipment typical transducer limitations still exist and the domain of applications (i.e., the slow varying signals) remains the same.

As digital recorders are generally sophisticated, one has to pay the price for the flexibility provided. In addition to low price, analog recording has the advantage that different-colored inks can be used, which is important in multi-channel recorders with overlapping (full range) curves. Available digital recorders allow only one color, thereby reducing identifiability. As the digital writing array is inside the apparatus there is a small latency between writing and appearing of the tracings. In overlapping multipen devices pens are physically dislocated and application of the pen offset compensation also creates a time delay. In digital recorders, displays are added to provide immediate visibility and easy setting of the recorder parameters. Control of the analog recorder is generally limited and accordingly simple to perform. Digital recorders provide more facilities, implying the need of training and experience to install instrument settings for the application envisaged.

Analog and digital systems thus have their specific limitations and benefits. The choice of the equipment for a specific application is a matter of performance, operating flexibility, and price. Measurements where immediate visualization of the tracings on a long strip of paper is not required and data are to be digitally analyzed and stored, can be performed by PC with virtual instrument software and printer. Nowadays the majority of data collection is accomplished using digital PC approaches. As previously stated, besides the measuring properties, the graphic quality is extremely important, as it assists visual examination of the records. In this case the general rule is applicable. Before one chooses an apparatus, one should see it in operation, that is, making graphic representations of the data.

BIBLIOGRAPHY

1. Olsen WH. Basic concepts in instrumentation. In: Webster JG, editor. *Medical Instrumentation: Application and Design*. 3rd edition. New York: John Wiley & Sons; 1998. p 1–43.
2. Bentley JP. *Principles of measurement systems*. 3rd ed. Longman House: Longman Group Limited; 1995.
3. Sydenham PH. Static and dynamic characteristics of instrumentation. In: Webster JG, editor. *The Measurement, Instrumentation and Sensors Handbook*. Boca Raton (FL): CRC Press; 1999. p 3/1–3/22.
4. Miller A et al., editors. *Electronic Measurements and Instrumentation*. New York: Mc Graw-Hill; 1975. p 427–479.
5. Bell DA. *Electronic Instrumentation and Measurements*. 2nd ed. Englewood Cliffs (NJ): Prentice-Hall; 1994.

See also ELECTROCARDIOGRAPHY, COMPUTERS IN; ELECTROENCEPHALOGRAPHY; ELECTROMYOGRAPHY; PHONOCARDIOGRAPHY.

RECORDS MANAGEMENT SYSTEM. See MEDICAL RECORDS, COMPUTERS IN.

REGULATIONS FOR MEDICAL DEVICES. See CODES AND REGULATIONS: MEDICAL DEVICES.

REGULATIONS FOR RADIATION. See CODES AND REGULATIONS: RADIATION.

REHABILITATION AND MUSCLE TESTING

W.K. DURFEE
P.A. IAIZZO
University of Minnesota
Minneapolis, Minnesota

INTRODUCTION

There is a growing need in clinical medicine to validate the quantitative outcomes of an applied therapy. In addition, the measurement of muscle function is an essential component of many neurological and physical exams. Muscle strength is correlated to function, work productivity, and general quality of life. Muscle function becomes compromised: (1) as we age, (2) when associated with a skeletal impairment, and/or (3) as a secondary consequence of many disease processes. Therefore, assessing muscle function is an important clinical skill that is routinely used by neurologists, orthopedists, general practitioners, anesthesiologists, and occupational and physical therapists. Evaluation of muscle strength is used for differential diagnosis, to determine if an impairment or disability is present, to decide if a patient qualifies for treatment, and or to track the effectiveness of a treatment.

In a research setting, the measurement of muscle function is used to further our understanding of the normal and potentially impaired neuromuscular system in human and/or animal experiments. In such research, muscle function can be assessed at the intact individual level (*in vivo*), in chronic and acute animal models (*in situ*), within isolated

muscle strips or even within single myofibrils (*in vitro*), and/or at the molecular–biochemical level. In this article, only whole muscle testing (*in vivo* and *in situ*) is discussed.

There are several components of muscle performance. The American Physical Therapy Association uses various definitions to explain the characteristics of muscle function (1). Muscle performance is the capacity of a muscle to do work. Muscle strength is the force exerted by a muscle or group of muscles to overcome a resistance in one maximal effort. Instantaneous muscle power is the mechanical power produced by the muscle (muscle force times muscle velocity). Muscle endurance is the ability to contract a muscle repeatedly over time. Of these performance indicators, muscle strength is the one most commonly measured when assessing the muscle function of intact humans.

In assessing muscle strength, the conditions under which the muscle contracts must be specified so that the muscle test data can be interpreted properly. The following conditions are relevant: “Isometric contraction”: the muscle contracts while at a fixed length; “Isotonic contraction”: the muscle contracts while working against a fixed load, for example, a hanging weight; “Isokinetic contraction”: the muscle contracts while moving at a constant velocity; (generally, isokinetic contractions are only possible with the limb strapped into a special machine that imposes the constant velocity condition); “Eccentric contraction”: the muscle contracts against a load that is greater than the force produced by the muscle so that the muscle lengthens while contracting; and “Concentric contraction”: the muscle contracts against a load that is less than the force produced by the muscle so that the muscle shortens while contracting.

Isometric muscle tests are the most common as they are the simplest to perform and reproduce and, because the test conditions are well defined, they are the most appropriate for comparing results within a population. Two considerations are important when testing muscle under isometric conditions. First, because muscle force varies with muscle length, the length of the muscle must be specified when planning and reporting a muscle test. For example the manual muscle test has strict and well-defined rules for the subject’s posture and joint positions that must be followed if one is to make clinical decisions based on the test (2).

Second, all isometric muscle tests of intact human muscle are conducted with the limb either held in a fixed position by the examiner, or with the limb fixed to a brace or jig (see the Stimulated Muscle Force Assessment section). While these methods hold the limb in a fixed position, the muscle will not be strictly isometric because of tendon stretch. The mismatch between limb condition and muscle condition only causes problems when trying to infer details about muscle dynamics, such as rise time or contraction speed from externally measured forces. Even if the whole muscle could be fixed at proximal and distal ends, during a twitch, the distance between z lines in the myofibril will shorten, which means the sarcomeres are shortening due to internal muscle elasticity. This is why the length tension and dynamic properties of whole muscle deviate somewhat from those of the isolated sarcomere. Nevertheless, length



Figure 1. Muscles wrap around joints. Muscle force is related to external force produced by a limb through skeletal geometry of joints and attachment points.

tension or ankle–angle/isometric–torque analyses can be done *in vivo* (3,4).

Testing of intact human muscle requires that muscle output be measured external to the body and, as a result, muscle force is never measured directly. As shown in Fig. 1, muscles wrap around joints and attach to limbs at the proximal and distal ends. There is a kinematic relationship between the measured force and the actual muscle force that depends on the details of muscle attachment and varies with joint angle. Therefore, to ultimately solve the kinematic relationship, one will require information about muscle attachment location, the geometry of the joint, and the joint angle. Such geometric information can be readily obtained from a magnetic resonance imaging (MRI) scan or a more generic geometry can be assumed, for example, obtained dimensions gathered from cadaver studies (5,6).

While often reported as a force, external testing of muscles more correctly should be reported as a torque. Figure 2 illustrates how force varies with location of the resistive load along the limb, while torque does not. Reporting muscle strength as torque about a joint eliminates this difficulty. If force is reported, the distance between the

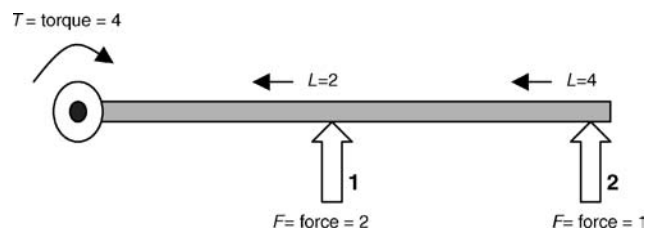


Figure 2. A torque of 4 is produced about a joint. It takes an opposing force of 2 to balance the torque if the opposing force is applied at arrow 1. If the opposing force is applied at arrow 2 it only takes a force of 1 to balance the torque. Thus, the perceived torque, and therefore the scoring of muscle strength, depends on where along the limb the examiner places his/her hand.

joint and the resistive load point should be measured to permit conversion to torque.

External measurement of torque about a limb joint means that all of the forces acting on that joint are measured, and that the contribution of the muscle or muscle group under study cannot be easily separated out. In other words, there are confounding forces generated by synergistic muscles. For example, when testing foot plantar flexion to determine gastrocnemius strength, the soleus may also be contributing to the measured torque about the ankle. Yet, another complicating factor may be undesired activations of antagonist muscles. One example is when you “flex your arm muscles”. In general, the resulting torque from the biceps and triceps are in balance, the arm does not move, and no external torque will be measured even though the muscles are contracting actively.

MANUAL MUSCLE TEST

The simplest and most common method of assessing muscle strength is the manual muscle test (MMT). Manual muscle testing is a procedure for evaluating strength and function of an individual muscle or a muscle group in which the patient voluntarily contracts the muscle against gravity load or manual resistance (2,7). It is quick, efficient, and easy to learn, however, it requires total cooperation from the patient and learned response levels by the assessor.

The procedures for conducting the MMT have been standardized to assure, as much as possible, that results from the test will be reliable (2,7,8). The specific muscle or muscle group must be determined and the examiner must be aware of, and control for, common substitution patterns where the patient voluntarily or involuntarily uses a different muscle to compensate for a weak muscle being tested.

To conduct a MMT, the patient is positioned in a posture appropriate for the muscle being tested, which generally entails isolating the muscle and positioning so that the muscle works against gravity (Fig. 3). The body part proximal to the joint acted on by the muscle is stabilized. A screening test is performed by asking the patient to move the body part through the full available range of motion



Figure 3. Manual muscle test of the iliopsoas.

Table 1. Manual Muscle Test Scores^a

Score	Description
0	No palpable or observable muscle contraction
1	Palpable or observable contraction, but no motion
1+	Moves limb without gravity loading less than one-half available ROM ^b
2–	Moves without gravity loading more than one-half ROM ^b
2	Moves without gravity loading over the full ROM ^b
2+	Moves against gravity less than one-half ROM ^b
3–	Moves against gravity greater than one-half ROM ^b
3	Moves against gravity less over the full ROM ^b
3+	Moves against gravity and moderate resistance less than one-half ROM ^b
4–	Moves against gravity and moderate resistance more than one-half ROM ^b
4	Moves against gravity and moderate resistance over the full ROM ^b
5	Moves against gravity and maximal resistance over the full ROM ^b

^aAdapted from Ref. 2

^bROM = range of motion.

(ROM). The main test is then performed either unloaded, against a gravity load, or against manual resistance, and a grade is assigned to indicate the relative muscle strength.

Manual grading of muscle strength is based on palpation or observation of muscle contraction, ability to move the limb through its available ROM against or without gravity, and ability to move the limb through its ROM against manual resistance by the examiner. Manual resistance is applied by the examiner using one hand with the other hand stabilizing the joint. Exact locations for applying resistive force are specified and must be followed exactly to obtain accurate MMT results (2). A slow, repeatable velocity is used to take the limb through its ROM, applying a resistive force just under the force that stops motion. The instructions to the patient are, use all of your strength to move the limb as far as possible against the resistance. For weaker muscles that can move the limb, but not against gravity, the patient is repositioned so that the motion is done in the horizontal plane with no gravity.

Grades are assigned on a 0–5 scale with ± modifiers (1 = trace score, 2 = poor, 3 = fair, 4 = good, 5 = normal) (Table 1). Grades > 1 demonstrate motion, and grades >3 are against manual resistance. Other comparable scoring scales exist (7).

As noted above, importantly, the assignment of scores is based on clinical judgment and the experience of the examiner. The amount of resistance (moderate, maximal) applied by the examiner is also based on clinical experience and is adjusted to match the muscle being tested as well as the patient’s age, gender and/or body type.

A common alternative to motion-based MMT is the isometric MMT in which the limb is held in a fixed position while the examiner gradually applies an increasing resistance force. The instructions to the patient are, Don’t let me move you. The amount of force it takes to “break” the patient is used to assign a score. Scoring norms for isometric MMT are provided in Table 2. While the MMT is the most widely used method to assess muscle function, its reliability and accuracy can be questionable (9,10). The MMT

Table 2. Grading of Isometric Manual Muscle Test^a

Score	Description
3	Maintains position against gravity
3+	Maintains position against gravity and minimal resistance
4-	Maintains position against gravity and less than moderate resistance
4	Maintains position against gravity and moderate resistance
5	Maintains position against gravity and maximal resistance

^aAdapted from Ref. 2.

scores are least accurate for higher force levels (9,11,12). Interrater reliability for MMT is not high, suggesting that the same examiner should perform multiple tests on one subject or across subjects (2). While not entirely accurate, MMT scores do correlate well with results from handheld dynamometers (13), implying that both are valid measures of muscle strength. However, as explained in a later section, all tests based on voluntary activation of a muscle are prone to artifact because of patient motivation and examiner encouragement.

APPARATUS

The appeal of the MMT is that it can be performed simply with the patient, an examiner, and a bench or table. This makes it ideal for the routine clinical environment where specialized equipment is unavailable and time is short. It is also suited for situations in which testing must be

performed away from the clinic, for example, in nursing homes, rural areas, or remote emergency settings.

When greater accuracy of results is needed, instruments are available that provide precise readouts of the resistive force the muscle works against (14). One example is a handheld dynamometer, such as the one shown in Fig. 4. This instrument can be sandwiched between the examiner’s hand and the patient’s limb, and provides a “readout” of force. The interrater reliability for handheld dynamometers is good when used with a standard procedure (15–17), as is the test-retest reliability (13).

Other products have been developed for specific tests of muscle strength, for example, the hand dynamometer and pinch grip devices shown in Fig. 5. These are easy to use and common for diagnostic tests of the hand. Despite their quantitative nature, readings between different types and brands of dynamometers can vary (18,19).

Computer-controlled dynamometers offer a variety of loading conditions for muscle testing and for strengthening treatments (20,21) (Fig. 6). Along with isometric and isotonic loading, dynamometer machines provide isokinetic conditions in which the muscle group acts against a computer-controlled resistance that moves the limb at a constant angular velocity.

ADVANCED MUSCLE ASSESSMENT METHODS

Measuring Muscle Dynamics

Muscle is a complex actuator whose external properties of force and motion result from the action of thousands of muscle fibers which, in turn, result from the action of

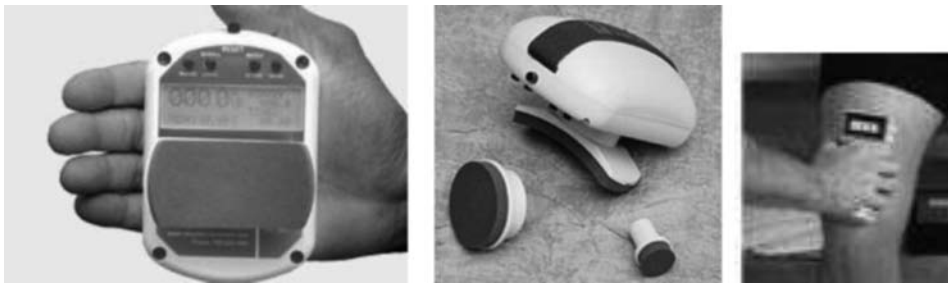


Figure 4. Handheld dynamometer. Pictured is the Lafayette manual muscle test system from Lafayette instrument company (Lafayette, IN).



Figure 5. The Jamar hand dynamometer is pictured on the left (NexGen Ergonomics, Quebec, Canada), and the B & L Pinch Gauge is shown on the right (B & L Engineering, Tustin, CA).



Figure 6. Biodesx dynamometer for computer-controlled muscle testing (Biodesx Medical Systems, Shirley, NY).

millions of structural and active proteins whose interaction is triggered by biochemical events. While most muscle testing focuses on the overall strength of a muscle or muscle group, more sophisticated assessment can be useful for in-depth examination of muscle function, including its dynamic, kinematic and fatigue properties.

The approach used to measure muscle function in more detail involves developing a mathematical model of muscle activity and then using experiments to identify the parameters of the model. Overviews of these methods are provided in Zajac and Winters (22), Durfee (23), Zahalak (24), Crago (25), and Kearney and Kirsch (26). The modeler must first choose the appropriate complexity of the mathematical model. The optimum choice is a model that is sufficiently complex to reveal the behavior of interest, but not so complex that parameters cannot be identified. Generally, Hill-type input-output models (27,28) are a good balance, as they capture key force-velocity, force-length, and activation dynamics at a whole muscle level (Fig. 7).

Model parameters can be identified one at a time, using the approach followed by Hill (27), or all at once using modern system identification techniques (23,26). Electrical activation of the muscle is a particularly convenient means for excitation because, unlike voluntary activation, there is control over the input, an essential component for an effective system identification method. Testing can be done

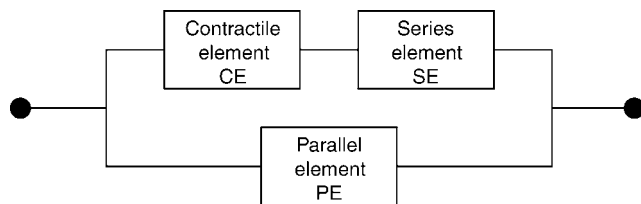


Figure 7. Hill muscle model. The contractile element (CE) contains the active element with dynamics, force-velocity and force-length properties. The series element (SE) is the inherent internal elastic elements, and the parallel element (PE) represents passive connective tissue.

under isometric conditions for determining recruitment and twitch dynamic characteristics, or under arbitrary loading to find active and passive force-length and force-velocity properties.

Identification of muscle properties is most easily accomplished using isolated muscle in acute animal model studies. Here the muscle is unencumbered by joint attachments and extraneous passive tissue. Muscle tendon can be directly attached to a force sensor and placed in a computer-controlled servo mechanism to apply known length and velocity trajectories, all while being stimulated. For example, the isometric recruitment curve, the relationship between muscle force and stimulus strength, can be identified using either point-at-a-time or swept amplitude methods, the latter being efficient in implementation (29). Using the model shown in Fig. 8, active and passive force-length and force-velocity properties can be estimated using brief bouts of controlled, random length perturbations, and then verified through additional trials where both stimulation and length are varied randomly (23,30) (Fig. 9). Simultaneous identification of active and passive muscle properties for intact human muscles is more challenging and represents an ongoing area of research (26).

Electromyogram

Contracting skeletal muscle emits an electrical signal, the electromyogram (EMG). Electrical recording of the EMG using needle or surface electrodes is an important diagnostic indicator used in clinical neurology to diagnose neuromuscular disorders including peripheral neuropathies, neuromuscular junction diseases, and muscular dystrophies. The EMG is also used in research as an estimator of muscle activity for biomechanics and motor control experiments. The reader is referred to Merletti and Parker (31) and Basmajian and DeLuca (32) for a comprehensive discussion of surface and needle EMG used in research applications, and to Preston and Shapiro (33), Kimura (34), and Gnatz (35) for an introduction to clinical EMG. Nevertheless, these assessment approaches require volitional activation of the patient’s musculature under investigation.

STIMULATED MUSCLE FORCE ASSESSMENT

As described above, most devices used clinically to quantify force and increase objectivity still rely on voluntary effort, which can be problematic. Pain, corticospinal tract lesions, systemic illness, and inconsistent motivation can significantly affect voluntarily activated muscle force. In addition, some neurologically impaired patients have difficulty maintaining a constant velocity of limb movement, and some very weak patients are unable to complete a full range of motion in voluntary force assessment tasks (36,37). As a specific example, monitoring muscle function in patients confined to the intensive care unit is a difficult challenge. Often such patients are on potent pain medications (e.g., morphine) and/or are sedated, or may have significant alterations in levels of consciousness due to critical illness (38,39). Thus, it can be extremely difficult to ask such patients to provide reproducible voluntary

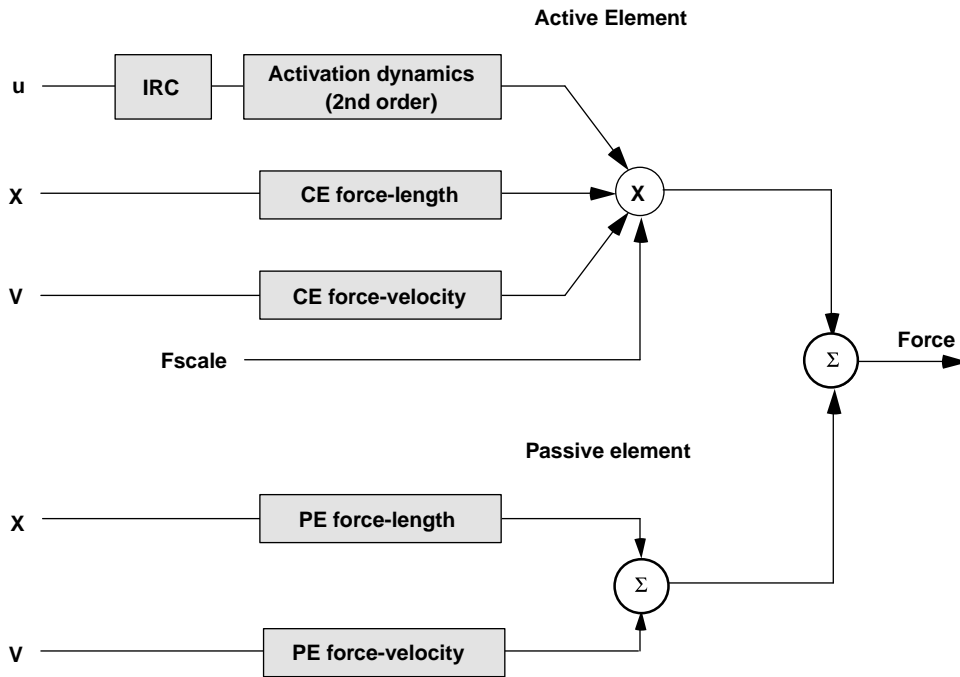


Figure 8. A model that can be used for muscle property identification. The active element has recruitment and twitch dynamics that multiplicatively combine with active force-length and force-velocity properties and sum with passive force-length and force-velocity properties to produce overall muscle force. (For details, see Refs. 23 and 30). IRC = isometric recruitment curve; CE = contractile element; PE = parallel element.

efforts. Even when cooperation is good, most measures of force assessment are qualitative, similar to using a hand-held unit for testing neuromuscular blockade.

Stimulated muscle force assessment is a versatile approach for quantitative involuntary muscle torque in humans. A muscle is activated by noninvasive nerve or motor point stimulation. A rigid apparatus is used to secure the appropriate portion of the subject's body in a predetermined position that confines movement to a specific direction, for example, ankle dorsiflexion, thumb adduction, arm flexion, or neck flexion (3,4,40,41) (Figs. 10 and 11). The innervating nerves or the motor points of the muscle are stimulated using surface electrodes, with either a single

stimulus to generate a twitch contraction or with short trains of stimuli to produce tetanic contractions (e.g., 5 ms interpulse intervals) (3,4). Incorporated strain gauges are used to measure isometric torque and, via acquisition software, all data are immediately displayed and on-line analyses are performed. Various parameters of the obtained isometric contractions are measured, for example, time between stimulus and torque onset, peak rate of torque development, time to peak torque, half-relaxation time, and other observed changes (Fig. 12; Table 3).

Such information is predicted to correlate with underlying physiological conditions and/or the presence of a myopathic or neuropathic disorder. To date, the average torque generated by healthy control subjects varies by <5% with repeated testing for contractions elicited from the various muscle groups studied (4,40,41). Thus, this assessment approach has potential utility in a number of research arenas, both clinical and nonclinical. Specifically, it has added clinical value in diagnosing a neuromuscular disorder, tracking weakness due to disease progression, and/or quantitatively evaluating the efficacy of a therapy (37,42–45). Compared to current assessment methods, we consider that monitoring isometric muscle torque generated by stimulation improves objectivity, reliability, and quantitative capabilities, and increases the type of patients that can be studied, including those under sedation (39) (Fig. 10). Stimulated muscle force assessment may be of particular utility in studying patients with a known underlying genetic disorder, for it could then provide important information as to genotype-phenotype associations (37).

The general configuration of the measurement system consists of the following main components: a stabilizing device that holds either the subject's arm or leg in a defined position; a force transducer that detects the evoked torque produced by a specific muscle group; hardware devices for nerve stimulation, signal amplification,

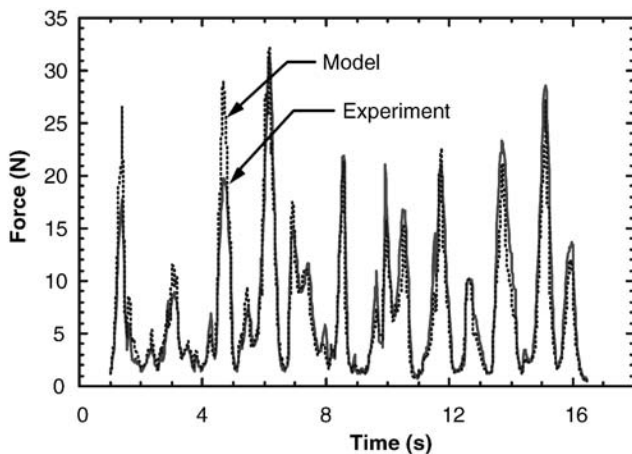
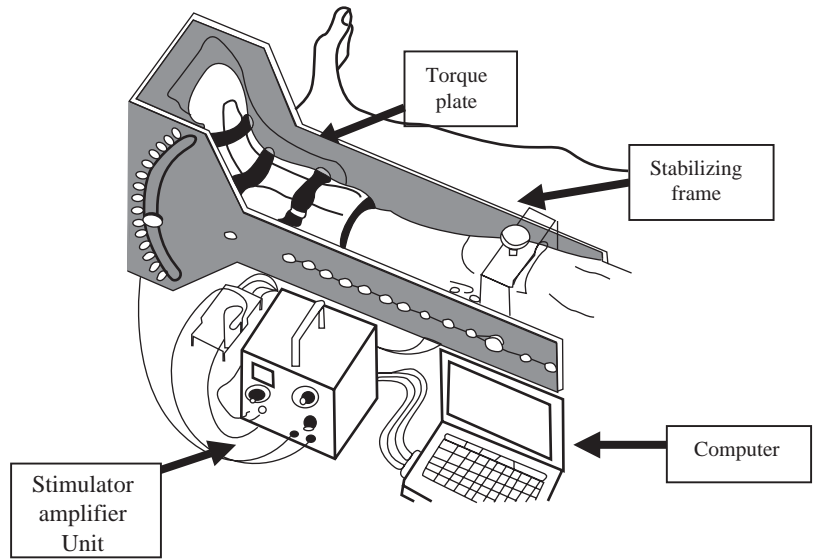


Figure 9. Results from an isolated muscle experiment where muscle active and passive properties were identified, then the resulting model was verified against experiment data. Data were generated while the muscle underwent simultaneous, random, computercontrolled stimulation and length perturbations (30).

Figure 10. Muscle force assessment system to determine involuntary isometric torque of the human dorsiflexor muscles. It is comprised of the following main components: (1) a stabilizing frame with knee supports; (2) the torque plate with mounted boot to fix the foot which can be rotated between -40° and 40° ; (3) a strain gauge system (Wheatstone bridge circuit) that detects the evoked torque; (4) a stimulator–amplifier unit that can supply variable stimulus pulse amplitudes and pulse durations and can amplify the voltage changes from the Wheatstone bridge circuit; and (5) a computer with data acquisition hardware and software for recording, analyzing and displaying all signals. (Modified from Ref. 39.)



and signal conditioning; a computer for stimulus delivery; and data acquisition software for recording, analyzing, and displaying all signals simultaneously (torque, EMG, applied stimulus) (Figs. 10–12).

The stabilizing device system currently used to study the dorsiflexors is a modification of a previously described apparatus (3). This device can be configured to maintain the subject's leg in a stable position while allowing access

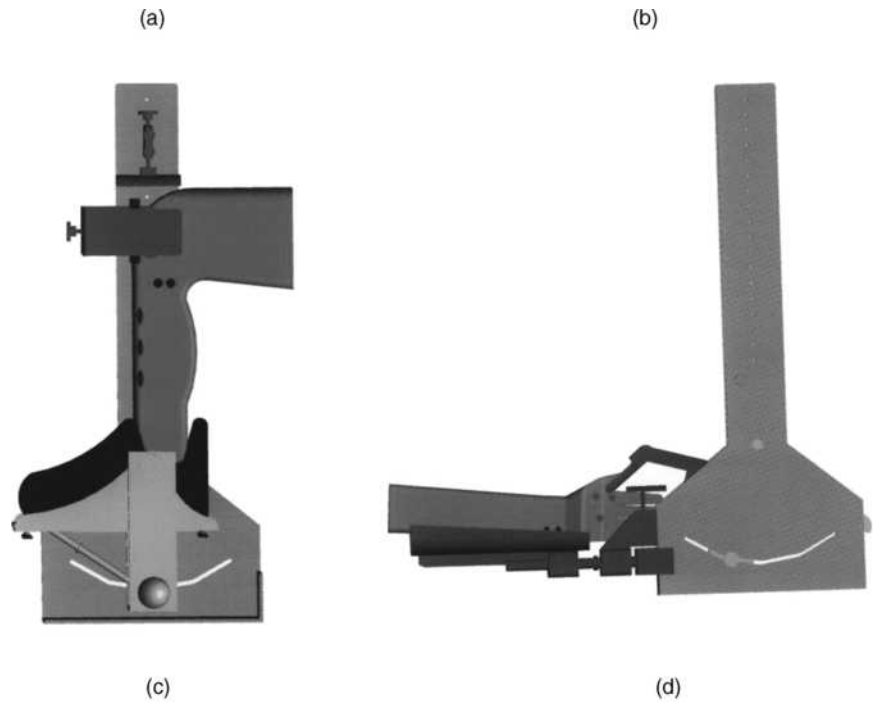


Figure 11. Various applications of stimulated muscle force assessment: (a) dorsiflexor muscles in a seated individual with stimulation of the common peroneal nerve lateral to the fibular head; (b) adductor pollicis muscle following ulnar nerve stimulation; (c) activated biceps force with motor point stimulation; and (d) head stabilizing/force system to study sternocleidomastoid muscle function following motor point stimulation. (See also Refs. 4, 40, and 41.)



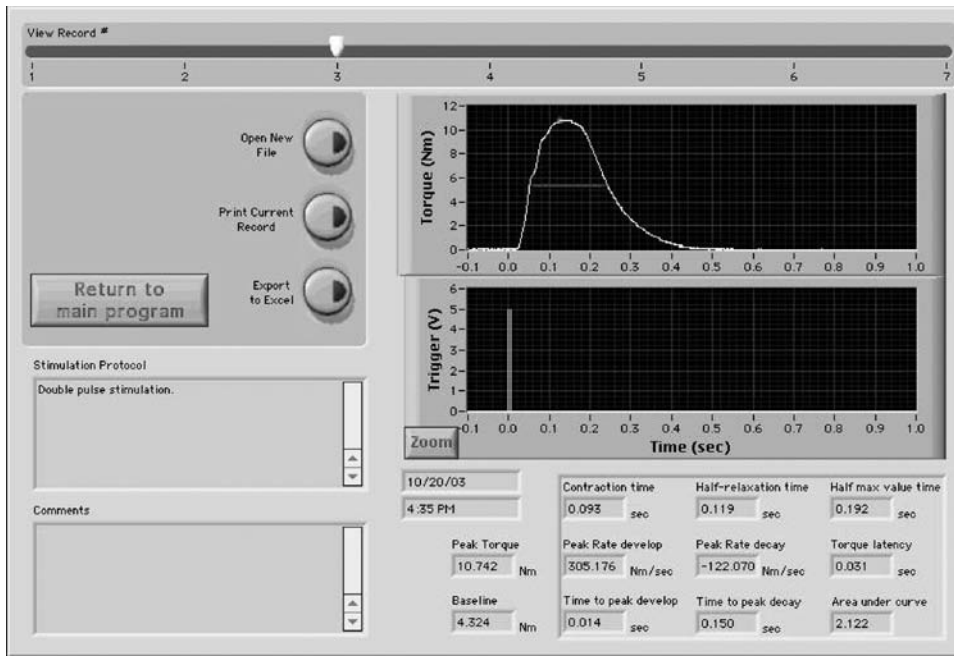


Figure 12. An example of a typical data display available to the investigator during subsequent off-line analyses. Graphically displayed are the muscle torque waveform and the stimulus administered (a double pulse with a 5 ms interpulse interval). Numerically displayed are various contractile parameters. Time 0 is the time of stimulation. The horizontal line indicates the time when half of the peak torque has been generated. The display shown is the torque generated by the dorsiflexor muscles of a normal, healthy subject.

for stimulation of the common peroneal nerve lateral to the fibular head (Fig. 11a). The torque about the ankle joint, produced by the dorsiflexor muscles (i.e., primarily generated by the tibialis anterior with contributions from the peroneus tertius and extensor digitorum muscles), is then quantified. One or two padded adjustable clamps can be used to maintain stability of the leg (knee slightly flexed in a supine position or flexed at 90° while seated). Modified in-line skate boots of varying sizes are affixed to the torque plate and adapted for either the right or left foot. The foot and ankle can be rotated within a 40° range while secured in the skate boot. This device can also be used for subjects in a supine position, in which case the support frame is secured to the upper leg proximal to the knee (Fig. 10) (39,45). To emphasize the extreme versatility of this methodology, note that a specialized version of this device was constructed and used to study dorsiflexor torques in hibernating black bears, *Ursus Americanus*, in the Rocky Mountains (46).

Briefly, the arm and hand stabilizing apparatus, used to measure muscle torque of the adductor pollicis, is easily attached to the main stabilizing frame (Fig. 11b). Using

straps, the forearm can be secured to the arm stabilizing unit, which can be adjusted for varying arm lengths. The digits (2–5) are placed in the hand well, and the thumb is secured to the constructed thumb bar attached to the torque plate. Shown in Fig. 11c is the configuration that is used to record force generated by the biceps muscle. As for the aforementioned muscles, force can be produced by peripheral nerve stimulation or by voluntary effort. In addition, we have successfully employed motor point stimulation of the muscle itself with large surface electrodes (40). The forces of the isometric muscle contractions are obtained as changes in torque applied to the instrument torque plate. Finally, we recently reported the optimization of an approach to study forces in the sternocleidomastoid muscle in the anterior neck (Fig. 11d), and plan to use these methodologies to study the effect of therapy in patients with cervical dystonia.

To date, this assessment approach has been used to study patients with a wide variety of disorders including: amyotrophic lateral sclerosis, Brody's disease, chronic inflammatory demyelinating polyneuropathy, malignant hyperthermia, muscular dystrophy, myotonia, periodic

Table 3. Contractile Parameters that Can Easily be Quantified

Parameter	Units ^a	Definition
Peak torque	N·m	Maximum amount of torque developed
Contraction time	s	Time from onset of torque to time of peak torque (e.g., calculated at 90% of peak)
Half-relaxation time	s	Time from peak torque to time when torque decays to half of peak torque
Peak rate of development	N·m/s	Maximum rate of torque development
Peak rate of decay	N·m/s	Maximum rate of torque decay
Time to peak development	s	Time from onset of torque to the peak rate of development
Time to peak decay	s	Time from peak rate of development to peak rate of decay
Half-maximal duration	s	Time when the generated torque is maintained at a level of half of the peak torque
Latency to onset	s	Time from the stimulus to the onset of torque development

^aN·m = newton meters.

paralysis, and nerve conduction blocks. The new insights to be gained by employing this approach in a variety of healthcare situations will further our clinical understanding of the underlying pathophysiologies, and provide us an accurate means to determine clinical outcomes. Recently, we employed this approach to evaluate athletes with potential overtraining syndrome (47), thus the applications for these methodologies could be considered quite limitless.

SUMMARY

The assessment of a patient's muscle strength is one of the most important vital functions that is typically monitored. Specifically, strength assessment is necessary for determining distribution of weakness, disease progression, and/or treatment efficacy. The particular assessment approach will be, in part, dictated by the clinical circumstance or severity of illness. Several assessment techniques and tools are currently available to the healthcare provider and/or researcher, yet each has its unique attributes. Nevertheless, as outcomes-based medical practice becomes the norm, the need for quantitative outcomes assessment of muscle strength will become even more important.

BIBLIOGRAPHY

- American Physical Therapy Association. Guide To Physical Therapy Practice. Alexandria, Virginia: American Physical Therapy Association; 1997.
- Clarkson HM. Musculoskeletal Assessment: Joint Range of Motion and Manual Muscle Strength. Philadelphia: Lippincott Williams & Wilkins; 2000.
- Quinlan JG, Iuzzo PA, Gronert GA, Lambert EH. Ankle dorsiflexor twitch properties in malignant hyperthermia. *Muscle Nerve* 1989;12:119–125.
- Brass TJ, Loushin MK, Day JW, Iuzzo PA. An improved method for muscle force assessment in neuromuscular disease. *J Med Eng Technol* 1996;20:67–74.
- Winter DA. Biomechanics of Human Movement. New York: John Wiley & Sons; 1979.
- Yamaguchi GA. A survey of human musculotendon actuator parameters. Winters JM, Woo S, editors. *Multiple Muscle Systems: Biomechanics and Movement Organization* 1990; New York: Springer-Verlag.
- Neistadt M, Crepeau E. Willard and Spackman's Occupational Therapy. Philadelphia: Lippincott Williams & Wilkins; 1998.
- Daniels L, Worthingham C. Muscle Testing: Techniques of Manual Examination. Philadelphia: W.B. Saunders; 1995.
- Iddings DM, Smith LK, Spencer WA. Muscle testing. 2. Reliability in clinical use. *Phys Ther Rev* 1961;41:249–256.
- Schwartz SM, Cohen ME, Herbison GJ, Shah A. Relationship between two measures of upper extremity strength: Manual muscle test compared to hand-held myometry. *Arch Phys Med Rehabil* 1992;73:1063–1068.
- Bohannon RW. Manual muscle test scores and dynamometer test scores of knee extension strength. *Arch Phys Med Rehabil* 1986;67:390–392.
- Knepler C, Bohannon RW. Subjectivity of forces associated with manual-muscle test grades of 3 +, 4 –, and 4. *Percept Mot Skills* 1998;87:1123–1128.
- Bohannon RW. Test-retest reliability of hand-held dynamometry during a single session of strength assessment. *Phys Ther* 1986;66:206–209.
- Amundsen L. Muscle Strength Testing: Instrumented and Non-Instrumented Systems. New York: Churchill Livingstone; 1990.
- Bohannon RW, Andrews AW. Interrater reliability of hand-held dynamometry. *Phys Ther* 1987;67:931–933.
- Horvat M, Croce R, Roswal G. Intratester reliability of the Nicholas Manual Muscle Tester on individuals with intellectual disabilities by a tester having minimal experience. *Arch Phys Med Rehabil* 1994;75:808–811.
- Dunn J, Iversen M. Interrater reliability of knee muscle forces obtained by hand-held dynamometer from elderly subjects with degenerative back pain. *J Geriatr Phys Ther* 2003; 26:23–29.
- Click Fenter P, Bellew JW, Pitts T, Kay R. A comparison of 3 hand-held dynamometers used to measure hip abduction strength. *J Strength Cond Res* 2003;17:531–535.
- Massy-Westropp N, et al. Measuring grip strength in normal adults: reference ranges and a comparison of electronic and hydraulic instruments. *J Hand Surg [Am]* 2004;29:514–519.
- Cabri JM. Isokinetic strength aspects of human joints and muscles. *Crit Rev Biomed Eng* 1991;19:231–259.
- Dvir Z. Isokinetics: Muscle Testing, Interpretation and Clinical Applications. New York: Churchill Livingstone; 1995.
- Zajac F, Winters JM. Modeling musculoskeletal movement systems. Winters JM, Woo S, editors. *Multiple Muscle Systems: Biomechanics and Movement Organizations*. New York: Springer-Verlag; 1990.
- Durfee WK. Model identification in neural prosthesis systems. In: Stein RB, Peckham PH, Popovic D, editors. *Neural Prostheses: Replacing Motor Function After Disease or Disability*. New York: Oxford University Press; 1992.
- Zahalak G. An overview of muscle modeling. In: Stein RB, Peckham PH, Popovic D, editors. *Neural Prostheses: Replacing Motor Function After Disease or Disability*. New York: Oxford University Press; 1992.
- Crago P. Creating neuromusculoskeletal models. In: Winters JM, Crago P, editors. *Biomechanics and Neural Control of Posture and Movement*. New York: Springer-Verlag; 2000.
- Kearney R, Kirsch R. System identification and neuromuscular modeling. In: Winters JM, Crago P, editors. *Biomechanics and Neural Control of Posture and Movement*. New York: Springer-Verlag; 2000.
- Hill AV. The heat of shortening and the dynamic constants of muscle. *Proc R Soc Lond [Biol]* 1938;126:136–195.
- Winters JM. Hill-based muscle models: A systems engineering perspective. In: Winters JM, Woo S, editors. *Multiple Muscle Systems: Biomechanics and Movement Organizations*. New York: Springer-Verlag; 1990.
- Durfee WK, MacLean KE. Methods for estimating isometric recruitment curves of electrically stimulated muscle. *IEEE Trans Biomed Eng* 1989;36:654–667.
- Durfee WK, Palmer KI. Estimation of force-activation, force-length, and force-velocity properties in isolated, electrically stimulated muscle. *IEEE Trans Biomed Eng* 1994;41:205–216.
- Merletti R, Parker P. *Electromyography: Physiology, Engineering, and Noninvasive Applications*. Hoboken: John Wiley & Sons; 2004.
- Basmajian JV, DeLuca CJ. *Muscles Alive: Their Function Revealed by Electromyography*. Baltimore, MD: Williams & Wilkins; 1985.
- Preston DC, Shapiro BE. *Electromyography and Neuromuscular Disorders: Clinical-Electrophysiologic Correlations*. Boston: Butterworth-Heinemann; 1998.
- Kimura J. *Electrodiagnosis in Diseases of Nerve and Muscle: Principles and Practice*. New York: Oxford University Press; 2001.
- Gnatz SM. *EMG Basics*. Austin, TX: Greenleaf Book Group; 2001.

36. Fillyaw MJ, Tandan R, Bradley WG. Serial evaluation of neuromuscular function in management of chronic inflammatory demyelinating polyneuropathy. *Phys Ther* 1987;67:1708–1711.
37. Day JW, et al. Force assessment in periodic paralysis after electrical muscle stimulation. *Mayo Clin Proc* 2002;77:232–240.
38. Jackson AC, Gilbert JJ, Young GB, Bolton CF. The encephalopathy of sepsis. *Can J Neurol Sci* 1985;12:303–307.
39. Ginz HF, Zorzato F, Iaizzo PA, Urwyler A. Effect of three anaesthetic techniques on isometric skeletal muscle strength. *Br J Anaesth* 2004;92:367–372.
40. Hong J, Iaizzo PA. Force assessment of the stimulated arm flexors: quantification of contractile properties. *J Med Eng Technol* 2002;26:28–35.
41. Hong J, Falkenberg JH, Iaizzo PA. Stimulated muscle force assessment of the sternocleidomastoid muscle in humans. *J Med Eng Technol* 2005;29:82–89.
42. Quinlan JG, Iaizzo PA, Gronert GA, Lambert EH. Twitch responses in a myopathy with impaired relaxation but no myotonia. *Muscle Nerve* 1990;13:326–329.
43. Quinlan JG, Wedel DJ, Iaizzo PA. Multiple-pulse stimulation and dantrolene in malignant hyperthermia. *Muscle Nerve* 1990;13:904–908.
44. Schulte-Mattler WJ, et al. Increased metabolic muscle fatigue is caused by some but not all mitochondrial mutations. *Arch Neurol* 2003;60:50–58.
45. Ginz HF, et al. Decreased isometric skeletal muscle force in critically ill patients. *Swiss Med Weekly* 2005; (in press).
46. Harlow HJ, Lohuis T, Beck TD, Iaizzo PA. Muscle strength in overwintering bears. *Nature (London)* 2001;409:997.
47. Nelson MF, Day SL, Sandell EN, Iaizzo PA. Quantitative analyses of dorsiflexor forces in marathon runners—A pilot study. *J Orthop Sports Phys Ther* (in review).

See also BIOMECHANICS OF EXERCISE FITNESS; ELECTROMYOGRAPHY; JOINTS, BIOMECHANICS OF; STRAIN GAGE.

REHABILITATION, COMPUTERS IN COGNITIVE

BEATRIZ C. ABREU
 GARY SEALE
 RICHARD O. TEMPLE
 ARCHANA P. SANGOLE
 Transitional Learning Center
 at Galveston
 Galveston, Texas

INTRODUCTION

The aim of this chapter is to provide an overview of current issues involving the use of computers in cognitive rehabilitation. The chapter begins with a brief historical review of computer use with a variety of disabilities including brain injury, learning disability, psychiatric disorders, and dementias. It continues to address selected research findings on the use of virtual reality for rehabilitation of impairments in attention, memory, and functional daily living activities. Finally, the chapter ends with conclusions and ethical reflections on using computers in research and direct care practice.

Impairments in cognitive function frequently occur as a result of acquired brain injury (i.e., trauma, cerebrovascular

accidents, anoxic encephalopathy, meningitis), specific learning disabilities, mental illness, Alzheimer's disease and other causes of dementia, and as a result of the natural aging process. These cognitive impairments can include, but are not necessarily limited to, decreased attention/concentration, memory, problem-solving and decision-making, planning, and sequencing. These impairments can negatively impact learning and skill acquisition, interfere with the ability to engage in everyday activities, preclude participation in social engagements, and hinder quality of life.

Cicerone et al. (1) and Giaquinto and Fiore (2) have defined cognitive rehabilitation as the systematic application of interventions to remediate or compensate for cognitive deficits and improve abilities in daily living skills and problem-solving. Cognitive rehabilitation interventions teach an individual to appropriately attend to, select, understand, remember relevant information, and apply the information appropriately in order to engage in meaningful daily activities and solve problems that occur in our complex society. Successfully completing daily activities and solving novel problems supports participation in meaningful societal roles such as breadwinner, husband, wife, and parent. Cognitive rehabilitation is generally carried out as part of a service delivery system that is interdisciplinary in nature (3). Services delivered are tailored to individual needs, are relevant to the person receiving the services, and bring about change that impacts daily functioning (1,2).

Over the past three decades, the personal computer has been employed as a tool to deliver interventions to remediate or compensate for cognitive deficits. Computers offer a number of advantages over traditional methods of cognitive remediation (i.e., flexibility, data collection, accessibility, portability, and cost), and, while not a treatment approach in and of themselves, computers can be a powerful tool to enhance the efforts of educators and clinicians.

COGNITIVE TRAINING FOR PERSONS WITH BRAIN INJURY

Computer-assisted cognitive retraining (CACR) for persons with acquired brain injury (ABI) began with the use of standard video games as an adjunct to traditional approaches to address deficits in attention/concentration, visual scanning, information processing speed, and divided attention (4). Lynch (4) reported that the initial use of video games for cognitive retraining was appealing to both clinicians and ABI survivors, as the games were inexpensive and widely available, and were interesting and motivating to the user. However, despite improvements on neuropsychological measures of basic cognitive skills reported in early single subject and small group pre-post design experiments using CACR (4,5), little carry over into everyday activities occurred. Other limitations existed as well, such as the inability to modify computer games for individual use and score user performance in a consistent and meaningful way (6). For these reasons, computer games gave way to educational programs that were developed for drills and practice of basic academic skills (i.e., vocabulary, math skills, and simple problem-solving/decision-making). However, the commercially produced educational software was

not without limitations, primarily the inability to easily modify the program to meet specific needs of an individual user or clinician. By the mid-to-late 1980s, specific computer software was developed for CACR with the ABI population. Some software programs addressed a number of cognitive skills (such as attention, memory, and sequencing) in a "package" (2,7). These programs allowed the clinician to vary levels of task complexity and individualize treatment by adjusting the speed of stimulus presentation, the delivery of cues/prompts, establishing reinforcement schedules, and so on. Results of studies using CACR to address specific deficits in attention/concentration (8,9), memory (10), visual processing (11), and visual scanning (12) also appeared in the literature.

Today, computers are used as assistive devices to help persons with cognitive deficits complete essential daily activities such as remembering appointments and items on a to-do list, and to overcome specific limitations such as difficulty speaking (voice synthesizer). Virtual environments also allow persons to practice skills in a safe, simulated environment (13,14). Weiss et al. (13) used a PC-based VR system to train stroke patients with unilateral spatial neglect to practice crossing a typical city street safely. Zhang et al. (14) developed a PC-based virtual kitchen allowing persons with acquired brain injury to practice cooking a meal. The computer program provided prompts as needed to assist the user to sequence and complete the task correctly. The use of virtual reality in cognitive rehabilitation is discussed in greater depth later on in the chapter.

COGNITIVE TRAINING FOR STUDENTS WITH LEARNING DISABILITIES

Personal computers appeared in the classroom in the late 1970s and early 1980s (15). Computer-assisted instruction (CAI) presented information in the three primary modalities: drill and practice, simulation, and tutorials. Drill and practice programs provided a single question or problem and elicited a response from the student. The student's answers were met with feedback from the computer program, followed by another question or problem. Simulation programs were more complex, requiring the student to process information on more than one level simultaneously and engage in a decision-making process. The computer program provided cues and feedback to assist the student in reaching the correct answer. Tutorial programs simulated the traditional style of classroom education delivered by most teachers. The information or content material was presented to the student. The computer program asked questions of the student, provided cues, prompts, and feedback as needed until the material was mastered. For some teachers and administrators, computers were thought to be the answer to problems associated with traditional approaches to instruction, particularly for educating challenging students. Teachers struggled with providing appropriate levels of instruction for students with a variety of learning styles, aptitudes, and in some cases, disabilities. With PCs in the classroom, gifted and talented students would be able to receive additional or more

challenging assignments, while ensuring additional drill and practice and self-paced learning for students who required more individualized assistance. Those who embraced early computer technology in the classroom experienced the PC as inexpensive, portable, and a way to provide challenging but nonthreatening instruction. Some viewed the computer as a "fashion statement," whereas others were afraid of the technology and resisted its use in the classroom. The greatest limitation of early computer technology for the classroom was memory capacity. Also, with pressure from parents and the booming PC industry, CAI programs for the classroom were introduced before being adequately assessed. Finally, it was difficult to measure the effectiveness of CAI as compared with traditional methods of instruction due to the number of variables that must be controlled in the classroom setting.

Over the past two decades, computer usage has increased, primarily due to the increased memory capacity available in today's computers. Special applications have been developed for the special education populations and for students with specific learning disabilities. Multimedia and hypermedia (i.e., presentation of information in text, graphics, sound, animation, and video) are now used with special populations to enhance writing skills (16) and mathematics and problem-solving skills (17). PCs are now used to overcome physical disabilities and language difficulties (i.e., problems understanding or using spoken or written language) of students participating in special education curriculums and continue to be used for drill and practice of basic academic skills. Research is mixed with regard to the impact of computer-assisted instruction on student's academic performance, primarily due to research design flaws in two critical areas: (1) inclusion of adequate controls and (2) holding instructional variables constant. However, a common theme emerging from most published studies is that technology cannot take the place of good teacher instruction. Use of computers in educational settings must include effective instruction from teachers, proper social organization of the classroom, and meaningful assignments.

COGNITIVE TRAINING FOR PERSONS WITH PSYCHIATRIC DISORDERS

Early use of computers in psychiatry and psychotherapy borrowed successes of the technology in educational settings and rehabilitation of persons with acquired brain injury (18,19). Mentally ill patients often demonstrate cognitive deficits similar to individuals with learning disabilities and traumatic brain injury, including decreased attention/concentration, memory, planning and problem-solving, and judgment/decision-making (18,19).

In the area of psychiatry, reports of computer-assisted interventions began to appear in the literature in the late 1980s. Computers were used as interviewing and assessment devices (e.g., diagnostic interviews) and for self-administered rating scales for depression and other mental illnesses (20–22). Computers were thought to have some advantage over a professional conducting a clinical interview, as some psychiatric patients were more willing to

disclose sensitive information to a computer rather than to a person (22).

Later, studies appeared in the psychiatric literature using computers to treat specific cognitive deficits, such as decreased attention and psychomotor speed in persons with schizophrenia (23,24).

Greater memory capacity and improved graphics allowed the development of multimedia presentations for patient education and specific data collection. In one study, Morss et al. (25) used a multimedia computer program to assess and evaluate the side effects of antipsychotic medication in persons with schizophrenia.

Finally, computers have been used in long-term psychiatric settings to teach high level vocational skills and to remediate educational disabilities. Brief (26) reported use of computers to teach advanced computer applications (such as database development for accounting and inventory, desktop printing and publishing, installing and upgrading software, and teaching staff word processing and spreadsheet skills), and to remediate deficits in mathematical abilities, reading comprehension, and vocabulary in persons with chronic mental illness. Brief (26) and Perr et al. (19) have cited numerous advantages of computer use with this population. Persons with chronic mental illness can learn to use computers and appear motivated to use the technology. Computers can be more engaging and provide for self-paced learning, making computer use more attractive than traditional classroom settings. Computer programs can be easily modified or individualized. Many patients demonstrate enhanced self-esteem as they master specific skills or become productive. Some other benefits of computer use with this population included increased attention/concentration and decreased frustration.

Computer technology has also been applied to the use of psychotherapy. Like psychiatry, computers have been used to aid in diagnostics, patient education and computer-assisted instruction, and cognitive rehabilitation (27). Computers have also been used in some forms of psychotherapy. With the advent of the PC, reports of the application of computer technology immediately began to appear in the literature in the late 1970s and early 1980s. Computers were used to desensitize anxious test-takers (28), in the treatment of agoraphobia (29), in the treatment of obesity (30), and in the treatment of sexual dysfunction in individuals and couples (31). Later, applications appeared in the field of behavioral health using computers for promoting smoking cessation and substance use/abuse (32).

In addition to numerous self-help applications, computers have been used in brief psychotherapy for presenting straightforward information or feedback (33).

Studies using computers in psychotherapy have shown that the technology is widely accepted and most people find computers to be a reliable source of information (databases, Internet, etc.). Similar to reports in the psychiatric literature, some persons find it easier to share sensitive information with a computer as opposed to a person. Rialle et al. (27) cite a number of advantages of computer-mediated psychotherapy services. First, regarding ethical considerations, it is highly unlikely that exploitation, abuse, or boundary issues will occur in a relationship between

a computer and persons receiving computer-assisted psychotherapy services. With continued increases in health-care costs, computers can be cost-effective and provide greater access to growing demands for mental health services. It is unlikely that computers, at least in the near future, will replace human psychotherapists given the complexity of the interaction that takes place in the course of intensive and long-term psychotherapy. The relationship between the person and the therapist is where the work of psychotherapy occurs, and computers cannot replace the warmth, empathy, and genuineness responsible for change in the therapeutic relationship.

COGNITIVE TRAINING FOR PERSONS WITH DEMENTIA

Reports of computer-assisted instruction and cognitive rehabilitation with the elderly have appeared in the nursing (34), geriatric (35), and psychology (36) literature over the past 15 years. PCs have been used in the treatment of age-related cognitive decline (37), Alzheimer's disease (AD), and other dementias (38). Computers have also been used for instruction, entertainment, and socialization of otherwise healthy elderly people without self-reported cognitive decline (39).

In the treatment of age-related cognitive decline, Gunther et al. (37) demonstrated that a 14 week computer-assisted cognitive rehabilitation program resulted in improved memory, information processing speed, learning, and interference tendency for 19 elderly participants who showed age-related cognitive decline without dementia. Follow-up five months after the completion of the cognitive rehabilitation program showed that information processing speed, learning, and interference tendency were maintained. The study also listed a number of advantages of computer use with this population, including the computer's value to motivate the elderly to learn, the computer's ability to directly measure success, its flexibility, and the ability to provide immediate and totally value-free feedback.

Computer-based interactive programs have been used to treat mild to moderate AD. Hoffman et al. (38) reported results from a pilot study of 10 AD patients. Although no evidence of general cognitive improvement or transfer of skills to real-life settings was noted, most participants in the study showed increased speed of computer use, required less assistance to use the computer program, and 8 of 10 made fewer mistakes. Mahendra (40) pointed out that many of the principles that facilitate learning in patients with AD or vascular dementia are easily incorporated into computer programs. For instance, repetition, active involvement in learning (i.e., interactive programs), cueing, feedback, and reinforcement of correct responses can easily be built into computer-assisted cognitive rehabilitation programs.

In a review article, Hendrix (39) reported on a number of studies that revealed computer usage by otherwise healthy elderly individuals resulted in improved self-esteem (i.e., from a sense of accomplishment or productivity), increased attention to task, and greater social interaction. Computer usage also provided entertainment and mental stimulation

in the form of games, puzzles, and the like. Many of the sensory (i.e., visual and hearing) and motor deficits associated with aging were overcome by the use of a PC. For example, increasing font size to at least 18 made text more readable. External speakers (for amplification) or visual indicators on the screen compensated for poor hearing. Touch screens, voice-activated typing programs, and keyboard guards were among other modifications that allowed the elderly to use computers for entertainment, research, and contact with friends and family.

Finally, computers and computer programs have recently been developed as a screening tool to identify mild cognitive impairment in early dementia patients (41). Future trends in the use of computers with the elderly will be aimed at preventing cognitive loss. "Geroprophylaxis" (37) or preventive therapies for the elderly may incorporate the use of computers in everyday activities, and probably at a younger age (for example, just after retirement) in an effort to prevent cognitive decline.

Although all these implementations present the conventional use of computers in cognitive rehabilitation, which have followed ever since the advent of computers in rehabilitation, VR technology, a more recent computer-based intervention in cognitive training, is increasing gaining attention. The technology offers a safe appendage to clinical interventions and is therefore just beginning to gain a therapeutic appeal in the area of rehabilitation. The following section discusses a few studies that have implemented the VR technology in cognitive rehabilitation.

VIRTUAL REALITY AS A COMPUTER-GENERATED TECHNOLOGY FOR REHABILITATION

VR provides a natural and intuitive interaction with a simulated environment because of the enhanced kinesthetic feedback gained by using various haptic devices. It provides the capability to display a 3D computer-generated environment, which allow individuals to interact and become immersed in the simulation as if they were in a naturalistic setting (42,43). Immersion or presence is achieved by a variety of projection systems ranging from basic flat-screen systems to projection walls and rooms known as CAVES (www.evl.uic.edu/pape/CAVE/). These projection systems produce virtual or mixed environments where real and simulated representations of objects and people can be used for evaluation and training of individual skills (44).

Specialized devices such as head-mounted displays (HMDs) combined with tracking systems, earphones, gesture-sensing gloves, and haptic feedback also facilitate the sense of immersion in the virtual environment (45). However, the VR headsets may produce deficits of binocular function after a period as short as 10 min (46,47). HMDs contribute to ocular discomfort, headaches, and motion sickness (48). Many factors may contribute to the symptoms when using HMDs, including the weight and fitting of the HMDs, the postural demands of the equipment, low illumination and spatial resolution, as well as the sensory conflict between the visual, vestibular, and nonvestibular proprioceptive system (48). Although some studies report

minimal risk or rapid dissipation of side effects when using VR technology, additional research is needed to determine the duration and severity of the symptoms (45,48).

Computer graphics techniques are used to create a virtual setting, complete with images and sound, that corresponds to what the user would experience in a real environment. The VR headset and a tracking system sense the position of the user's head and communicate the information to the computer that uses this spatial information to immerse and orient the user in the virtual setting.

Therefore, the user can navigate and interact with objects in the virtual environment using other VR devices such as data gloves, joy sticks, or even the natural hand (45,49). In other words, these collections of 3D computer-generated images can generate a continuum of high fidelity virtual environments. The VR devices and displays range from "virtual world objects" to "mixed reality" in which the real world and the virtual world objects are presented together within a single display (49,50).

The VR technology seems to offer many opportunities for evaluation and training of healthy and disabled populations. Successful reports include the U.S. National Aeronautics and Space Administration's use of VR training for astronauts to repair the Hubble telescope in space (51) and use of VR in the rehabilitation of people with intellectual disabilities (52).

VR FOR COGNITIVE TRAINING

This section will focus on the potential of VR for cognitive training. Persons who suffer from central nervous system damage may experience profound and pervasive difficulties with cognition. The beneficial aspects and limitations of VR for cognitive impairments will be discussed from the impairment and functional disability perspective. This model represents the continuum of the health-care services at the body and functional levels. VR training in cognitive rehabilitation has been divided into attention training and memory training.

VR FOR ATTENTION ASSESSMENT AND TRAINING

The brain allows individuals to constantly scan the environment for stimuli. Arousal, orientation, and focus are regulated in the brain. The reticular activating systems, the superior colliculus and the parietal cortex, and the lateral pulvinar nucleus in the thalamus are active processors of attention. All these brain structures in connection with the frontal lobes allow the individual to establish and maintain stimulus control and observe features in the environment. Attention is a multicomponent behavior and can be impaired in many neurological conditions. VR has been used for attention training. Some of the theories about the perception of reality in the virtual experience are, in fact, attributed to the three dimensions of attention (53). The three attention dimensions are (1) the focus of attention between presence and absence of reality; (2) the locus of attention between the virtual and physical world; and (3) the "sensus" of attention between arousal and the users internal physiological responses (53).

In general, attention is the ability to focus on critical aspects of the stimulus in the environment. VR technology can provide a controlled stimulus environment in which cognitive distractions can be presented, monitored, manipulated, and recorded using various levels of attention, including (1) **focused attention** (to perceive and respond to specific information/stimuli), (2) **sustained attention** (to maintain consistent concentration or vigilance on performing a task), (3) **selective attention** (to avoid distractions or competing stimuli), (4) **alternating attention** (to shift or alternate the focus of attention between tasks), and (5) **divided attention** (to respond to multiple stimuli or to give two or more responses simultaneously) (45,54). Several studies have supported the potential use of VR for the assessment and training of attention skills. In 2001, Rizzo et al. (55) reported on a virtual classroom environment, and in 2002, Lengenfelder et al. (42) reported on a driving course.

Divided attention specifically requires the ability to respond to multiple tasks at the same time or give two responses simultaneously. Lengenfelder et al. (42) used a VR driving course environment displayed on a computer screen to study divided attention of VR drivers with and without traumatic brain injury (TBI). The task required driving while identifying a four-digit number appearing in the same or random locations on the vehicle's windshield. The preliminary results showed no differences in relative speed between VR drivers with and without TBI on any of the four attention conditions used but rather the rate of the stimulus presentation seemed to influence the driving performance. In addition, the VR drivers with TBI showed a greater number of errors on the secondary task (number recall) performed while driving. Spearman's correlations between the VR performance of divided attention and neuropsychological measures of divided attention indicated that the more errors that were made during the VR divided attention task, the lower the number of correct responses on neuropsychological measures of divided attention. The findings suggest that VR may provide a way to measure divided attention and its impact on driving.

Other researchers have designed VR classrooms to examine children with attention deficit hyperactivity disorder (ADHD) during the control and manipulation of visual and auditory distracters (41,56). The preliminary results showed that children with ADHD had significantly more omission errors in the distracting conditions than the children without ADHD (56). Cho et al. (41) developed and studied two virtual cognitive training courses, VR classroom environment and a comparison computer cognitive training. The VR head-mounted display was used to validate the possibility of attention enhancement on 30 teenagers who had been isolated in a reformatory facility. The participants were assigned into three groups: VR group, nonVR (cognitive training) group, and control group (no special treatment). The interventions took eight sessions over two weeks. The results showed that the VR group was the most improved in attention training. These studies support the use of VR for attention impairment training. VR also has potential for memory impairment training.

VR FOR MEMORY ASSESSMENT AND TRAINING

An important feature of cognitive rehabilitation is neuropsychological assessment in order to determine the areas of impairment and function to use for training and compensation (45). Conventional standardized memory assessments have been criticized for lacking in ecological validity (57). VR may be able to increase the ecological validity that many conventional standardized memory assessments are lacking (57). VR can be designed for assessment and training of memory impairments after disability. Memory is a multifaceted process of brain function. Although many types of memory exist that activate a complex network of structures in the nervous system, memory requires attention to the information, encoding and maintaining information in short-term memory, followed by storing information in long-term memory, and finally, consolidating the information for retrieval as needed (54). At the cellular level, memory can be conceived as a specific neuronal association pattern that remains as a permanent pattern in the brain after the original information stimulus has ceased to exist (58).

Many brain structures are involved in memory function. Many areas are located anatomically beneath the cingulate cortex, and include the thalamus, fornix, mammillary bodies, hippocampus, amygdala, basal forebrain, and prefrontal cortex (59).

Some exploratory studies indicate that VR has potential for memory remediation in people with memory impairments. It has been found that VR can promote procedural learning and transfer to improved real-world performance (57). Burgess et al. (60) investigated four types of recognition memory in a VR town environment. Thirteen healthy young male volunteers explored the VR town until they felt confident they could find their way around. They were then asked four types of forced choice recognition memory questions about person, place, object, and the width of the object. Event-related functional magnetic resonance imaging (efMRI) was performed while participants answered the questions. The results revealed that no significant difference in performance existed between the memory for person and the memory for place. The performance on the memory for object was significantly better. By combining fMRI with VR technology, the investigators were able to investigate spatial memory in simulated life-like events. The results suggested that the retrieval of VR spatial events (place, location) activated medial temporal, parietal, and prefrontal systems in the brain.

In another study, Hoffman et al. (61) explored remembering real, virtual, and false memories using a Virtual-Real Memory Characteristic Questionnaire (VRMCQ). As people can differentiate between memories of real and imagined events, VR environments offer a new source of memories for events. The authors explored how accurately people can distinguish reality from VR in memory via a source identification test. Participants were exposed to real and virtual objects and, one week later, took an identification test in which they determined whether the object had been real, virtual, or new. They then rated the qualities and associated with each memory using the (VRMCQ). Clear differences in the qualities associated with real

and VR objects were found. The real objects were more likely than VR objects to be associated with perceptual cues (61,62).

A wide variety of exploratory studies have been reviewed by Brooks and Rose (57) indicating that VR can enable a more comprehensive and controlled assessment of prospective memory than is possible with paper-pencil standardized tests. Brooks and colleagues further investigated the differences between active and passive participation in the nonimpaired participants.

Phobias have been treated using VR to desensitize patients to heights (acrophobias) (63), public speaking (agoraphobia) (64), small spaces (claustrophobia) (65), and flying (66).

VR FOR ACTIVITIES OF DAILY LIVING ASSESSMENT AND TRAINING

VR Kitchen Environments

Effective VR assessment and training of activities of daily living have been reported in a variety of areas including a virtual kitchen (14,67–70). Our research team developed an HMD virtual kitchen environment designed to assess persons with brain injury on their ability to perform a 30 step meal preparation task (68). The prototype was tested using a sample of 30 persons with traumatic brain injury. The pilot results showed the stability of performance estimated using intraclass correlation coefficient (ICCs) of 0.73. When three items with low variance were removed, the ICC improved to 0.81. Little evidence of vestibular optical side effects was noted in the subjects tested. In 2001, our research team used the same virtual kitchen environment to assess selected cognitive functions of 30 patients with brain injury and 30 volunteers without brain injury to process and sequence information (70). The results showed that persons with brain injuries consistently demonstrated a significant decrease in the ability to process information identify logical sequencing and complete the overall assessment compared with volunteers without brain injury. The response speed was also significantly different. Our team's VR third pilot was designed to test the stability and validity of the information collected in the VR environment from 54 consecutive patients with TBI (14). The subjects completed meal preparation both in a virtual kitchen and in an actual kitchen twice over a three week period. The results showed an ICC value of 0.76. Construct validity of the VR environment was demonstrated. Multiple regression analysis revealed that the VR kitchen test was a good predictor for the actual kitchen assessment. Our investigations demonstrated adequate reliability and validity of the VR system as a method of assessment in persons with brain injury.

Exercising in VR environments offers the potential for gains in cognitive and motor functions (71–73). Refer to Fig. 1 for examples of a VR software program that has been used for cognition and motor skills training. Grealy et al. (71) supported the impact of exercise and VR on the cognitive rehabilitation of persons with traumatic brain injury in a study that provided a four week VR intervention and

compared it with control patients of similar age, severity, and time post injury. They found the patients performed significantly better than controls on the tests of psychomotor speed, and on verbal and visual learning. Significant improvements were observed following a single session of VR exercise.

VR Mobility Environments

VR environments have been developed to train aspects of mobility including street crossing, wheelchair mobility, and use of public transportation and driving. Strickland et al. (74) found that children with autism were able to accept and wear VR helmets. McComas et al. (75) also investigated the effectiveness of VR for pedestrian safety in healthy children. Their results showed that significant change in performance occurred after three trials with VR intervention. Children learned safe street crossing in a desktop VR program. Their crossing skills were transferred to real behaviors in the suburban school group but not in the urban school group. The investigators provided no explanation for the difference noted.

VR environments have been developed for detection of driving impairment in persons with cognitive disability in persons with TBI and persons with driving and flying phobias (76,77). As stated before, impairments in divided attention has an impact on driving skills; therefore, attention training is promoted for driving skills training.

CONCLUSION

The fast data recording and processing capabilities have stimulated the application of computers in cognitive rehabilitation since the mid-1970s. It was first used in the treatment of persons with traumatic brain injury and learning disability. The technology was later applied to persons with mental illness and, most recently, in the treatment of adults with Alzheimer's and other dementias. In all applications, computers have been found to be interesting and motivating to the users, an efficient stimulus presentation, and an excellent data collection system. Today's computers are portable, fast, and software can be customized to meet specific needs of an individual user, teacher, or therapist.

A growing body of research continues to accumulate investigating the potential of computer technology to improve impairments, functional performance, and societal participation in persons with disabilities. It is clear that some evidence exists to support computer-based interventions as an adjunct to clinician-guided treatment (78). However, sole reliance on repeated exposure and practice on computer-based tasks without some involvement and intervention by the therapists is not recommended for cognitive rehabilitation (78).

Computer technology can provide valuable opportunities to health-care providers, educators, and researchers. Computer technology training is not a simple or single solution, and although it can promote knowledge and solutions, we also need to address other implications for cognitive rehabilitation. For example, access to computer technology is needed for personal assistance and to provide



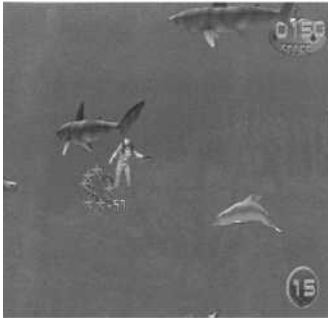
Snowboarding



Drums



Soccer



Diving



Beach Volleyball



Birds & Balls



Juggler



Coconutz

Figure 1. Examples of virtual reality environments. Image of the IREX™ system used with permission of GestureTek Inc™ - World Leaders in Gesture Recognition Technology www.gesturetek.com.

access to education, employment, and social opportunities. However, because of the expense that limits technology access because of constraints in institutions budgets, personal income limits, and funding inadequacies and constraints, only a limited number of persons with disabilities can access computer technology (79).

Finally, ethical considerations of computer technology, including VR, require mentioning. Several areas need to be considered, including (1) the patient or user must share the control and responsibility of the computer technology experience with the therapist, educator, or researcher; (2) careful care and monitoring is required for patients with certain psychopathology for potential user difficulties that may be encountered through the computer technology

experience; (3) research participants must not be deceived into believing that they are experiencing real-life events; and (4) the patient or user must not be deprived of real-life experiences.

The application of computer technology is promising. However, substantial work needs to be conducted to identify and improve best practices through this medium. An important question is whether the improvement in the individual who used computer technology training is transient or sustained. Cognitive rehabilitation needs to promote generalization to everyday functioning and everyday activities. We need more rigorous research activities geared toward establishing a body of evidence that supports the effectiveness of computer-based technology.

BIBLIOGRAPHY

1. Cicerone KD, Dahlberg C, Kalmar K, Langenbahn DM, Malec JF, Berquist TF, Felicetti T, Giacino JT, Harley JP, Harrington DE, Herzog J, Kneipp S, Laatsch L, Morse PA. Evidence-based cognitive rehabilitation: Recommendations for clinical practice. *Arch Phys Med Rehabil* 2000;81:1596–1615.
2. Giaquinto S, Fiore M. THINKable, a computerized cognitive remediation: First results. *Acta Neurol (Napoli)* 1992;14(4–6):547–560.
3. Harley JP, Allen C, Braciszewski TL, Cicerone KD, Dahlberg C, Evans S, Foto M, Gordon WA, Harrington D, Levin W, Malec JF, Millis S, Morris J, Muir C, Richert J, Salazar E, Schiavone DA, Smigelski JS. Guidelines for cognitive rehabilitation. *NeuroRehabilitation* 1992;2(3):62–67.
4. Lynch B. Historical review of computer-assisted cognitive retraining. *J Head Trauma Rehabil* 2002;17(5):446–457.
5. Larose S, Gagnon S, Ferland C, Pepin M. Psychology of computers: XIV. Cognitive rehabilitation through computer games. *Percept Mot Skills* 1989;69(3 Pt. 1):851–858.
6. Kurlychek RT, Levin W. Computers in the cognitive rehabilitation of brain-injured persons. *Crit Rev Med Inform* 1987;1(3):241–257.
7. Harley JP. Software review: “COGREHAB” cognitive rehabilitation programs. *Comput Psychiatr Psychol* 1984;6:15.
8. Niemann H, Ruff RM, Baser CA. Computer-assisted attention retraining in head-injured individuals: A controlled efficacy study of an outpatient program. *J Consult Clin Psychol* 1990;58(6):811–817.
9. Gray JM, Robertson I. Remediation of attentional difficulties following brain injury: Three experimental case studies. *Brain Inj* 1989;3(2):163–170.
10. Glisky E, Schacter D. Long-term retention of computer learning by patients with memory disorders. *Neuropsychologia* 1988;26:173–178.
11. Drette DK, Hinojosa J. The effects of a compensatory intervention on processing deficits of adults with acquired brain injuries. *Occup Ther J Res* 1999;19(4):224–239.
12. Ross FL. The use of computers in occupational therapy for visual-scanning training. *Am J Occup Ther* 1992;46(4):314–322.
13. Weiss PL, Naveh Y, Katz N. Design and testing of a virtual environment to train stroke patients with unilateral spatial neglect to cross a street safely. *Occup Ther Int* 2003;10(1):39–55.
14. Zhang L, Abreu BA, Seale GS, Masel BE, Christiansen CH, Ottenbacher KJ. A virtual reality environment for evaluation of a daily living skill in brain injury rehabilitation: Reliability and validity. *Arch Phys Med Rehabil* 2003;84:1118–1124.
15. Kolich EM. Microcomputer technology with the learning disabled: A review of the literature. *J Learn Disabil* 1985;18(7):428–431.
16. MacArthur CA. Using technology to enhance the writing process of students with learning disabilities. *J Learn Disabil* 1996;29(4):344–354.
17. Babbitt BC, Miller SP. Using hypermedia to improve mathematics problem-solving skills of students with learning disabilities. *J Learn Disabil* 1996;29(4):391–401.
18. Burda PC, Starkey TW, Dominguez F, Vera V. Computer-assisted cognitive rehabilitation of chronic psychiatric patients. *Comput Hum Behav* 1994;10(3):359–368.
19. Perr A, White S, Rekoutis PA. Assistive technology and computer-based intervention in psychiatric settings. *Technol Special Interest Section Quart* 2002;12(2):1–4.
20. Erdman HP, Greist JH, Gustafson DH, Taves JE, Klein MH. Suicide risk prediction by computer interview: A prospective study. *J Clin Psychiatry* 1987;48:464–467.
21. Ancil RJ, Rogers D, Carr AC. Comparison of computerized self-rating scales for depression with conventional observer ratings. *Acta Psychiatr Scand* 1983;71:315–317.
22. Erdman HP, Klein MH, Greist JH. Direct patient computer interviewing. *J Consult Clin Psychol* 1985;53:760–773.
23. Hermanutz M, Gestrich J. Computer-assisted attention training in schizophrenics: A comparative study. *Eur Arch Psychiatry Clin Neurosci* 1991;240(4–5):282–287.
24. Benedict RHB, Harris AE. Remediation of attention deficits in chronic schizophrenic patients: A preliminary study. *Br J Clin Psychol* 1989;28:187–188.
25. Morss SE, Lenert LA, Faustman WO. The side effects of antipsychotic drugs and patients’ quality of life: Patient education and preference assessment with computers and multimedia. *Proc Annu Symp Comput Appl Med Care* 1993;17–21.
26. Brief R. Personal computers in psychiatric rehabilitation: A new approach to skills training. *Hosp Commun Psychiatry* 1994;45(3):257–260.
27. Rialle V, Stip E, O’Connor K. Computer-mediated psychotherapy: Ethical issues and difficulty with implementation. *Humane Med* 1994;10(3):185–192.
28. Biglan A, Willwock C, Wilk S. The feasibility of a computer-controlled program for the treatment of test anxiety. *J Ther Exp Psychiatry* 1979;10:47–49.
29. Ghosh A, Marks IM. Self-treatment of agoraphobia by exposure. *Behav Ther* 1987;18:3–16.
30. Wylie-Rosett J, Swenckionis C, Ginsberg M, Cimino C, Wassertheil-Smoller S, Caban A, Segal-Isaacson CJ, Martin T, Lewis J. Computerized weight loss intervention optimizes staff time: The clinical and cost results of a controlled clinical trial conducted in a managed care setting. *J Am Diet Assoc* 2001;101(10):1155–1164.
31. Reitman R. The use of small computers in self-help sex therapy. In: Schwartz M, editor. *Using Computers in Clinical Practice*. New York: Haworth Press; 1984. p 363–380.
32. McDaniel AM, Hutchison S, Casper GR, Ford RT, Stratton R, Rembusch M. Usability testing and outcomes of an interactive computer program to promote smoking cessation in low-income women. *AMIA 2002 Annual Symposium Proceedings*. 2002. 509–513.
33. Gould RL. The Therapeutic Learning Program (TLP): A computer-assisted short-term treatment program. *Comput Psychiatry Psychol* 1986;8(3):7–12.
34. Taira F. Computer use by adults with disabilities. *Rehabil Nurs* 1994;19(2):84–86, 95.
35. Weisman S. Computer games for the frail elderly. *Gerontologist* 1983;23(4):361–363.
36. Schreiber M, Schweizer A, Lutz K, Kalveram KT, Jancke L. Potential of an interactive computer-based training in the rehabilitation of dementia: An initial study. *Neuropsycholog Rehab* 1999;9(2):155–167.
37. Gunther VK, Schafer P, Holzner BJ, Kemmler W. Long-term improvements in cognitive performance through computer-assisted cognitive training: A pilot study in a residential home for older people. *Aging Ment Health* 2003;7(3):200–206.
38. Hofmann M, Hock C, Kuhler A, Muller-Spahn F. Interactive computer-based cognitive retraining in patients with Alzheimer’s disease. *J Psychiatr Res* 1996;30(6):493–501.
39. Hendrix C. Computer use among elderly people. *Comput Nurs* 2000;18(2):62–71.
40. Mahendra N. Direct interventions for improving the performance of individuals with Alzheimer’s disease. *Semin Speech Lang* 2001;22(4):291–303.
41. Cho, Yang J, Kim SY, Yang DW, Park M, Chey J. The validity and reliability of a computerized dementia screening test developed in Korea. *J Neurol Sci* 2002;203–204:109–114.
42. Lengenfelder J, Schultheis MT, Al-Shihabi T, DeLuca J, Mourant R. Divided attention and driving: A pilot study using virtual reality. *J Head Trauma Rehabil* 2002;17(1):26–37.

43. Rizzo AA, Buckwalter JG. Virtual reality and cognitive assessment and rehabilitation: The state of the art. In: Riva G, editor. *Virtual Reality in Neuro-Psycho-Physiology: Cognitive, Clinical And Methodological Issues In Assessment And Rehabilitation*, vol. 44. Amsterdam: ISO Press; 1997. p 123–145.
44. Cosman PH, Cregan PC, Martin CJ, Cartmill JA. Virtual reality simulators: Current status in acquisition and assessment of surgical skills. *ANZ J Surg* 2002;72(1):30–34.
45. Schultheis MT, Himelstein J, Rizzo AA. Virtual reality and neuropsychology: Upgrading the current tools. *J Head Trauma Rehabil* 2002;17(5):378–394.
46. Mon-Williams M, Plooy A, Burgess-Limerick R, Wann J. Gaze angle: A position mechanism of visual stress in virtual reality. *Ergonomics* 1998;41(3):280–285.
47. Wann JP, Rushton S, Mon-Williams M. Natural problems for stereoscopic depth perception in virtual environments. *Vision Res* 1995;35(19):2731–2736.
48. Ames SL, Wolffsohn JS, McBrien NA. The development of a symptom questionnaire for assessing virtual reality viewing using a head-mounted display. *Optom Vis Sci* 2005;82(3):168–176.
49. Nothhelfer U. Landscape Architecture in the Reality-Virtuality. Paper presented at the Trends in GIS and Virtualization in Environmental Planning and Design, Anhalt University of Applied Sciences, 2002.
50. Milgram P, Kishino F. A taxonomy of mixed reality visual displays. *IEICE Trans Inform Syst (Special Issue on Networked Reality)* 1994;E77-D(12):1321–1329.
51. Loftin RB, Kenney PJ. Training the Hubble-space telescope flight team. *IEEE Comput Graph Appl* 1995;15(5):317.
52. Standen PJ, Brown DJ. Virtual reality in the rehabilitation of people with intellectual disabilities: Review. *CyberPsychol Behav* 2005;8(3):272–282.
53. Waterworth L, Waterworth JA. Focus, locus, and sensus: The three dimensions of virtual experience. *CyberPsychol Behav* 2001;4(2):203–213.
54. Malia KB, Bewick KC, Raymond MJ, Bennett TL. *Brain-wave-R: Cognition Strategies and Techniques for Brain Injury Rehabilitation: User's Guide and Introduction to Brain Injury*. Austin, TX: Pro-Ed; 2002.
55. Rizzo AA, Buckwalter JG, McGee JS, Bowerly T, Van der Zaag C, Neumann U, Thiebaut M, Kim L, Pair J, Chua C. Virtual environments for targeting cognitive processes: An overview of projects at the University of Southern California Integrated Media Systems Center. *Presence: Teleoperators Virtual Environ* 2001;10:359–374.
56. Rizzo A, Buckwalter JG, Bowerly T, Van der Zaag C, Humphrey L, Neumann U, Chua C, Kyriakakis C, van Rooyen A, Sisemore D. The virtual classroom: A virtual reality environment for the assessment and rehabilitation of attention deficits. *CyberPsychol Behav* 2000;3(3):483–499.
57. Brooks M, Rose FD. The use of virtual reality in memory rehabilitation: Current findings and future directions. *NeuroRehabilitation* 2003;18(2):147–157.
58. Carter R. *Mapping the Mind*. Los Angeles, CA: University of California Press; 1998.
59. Parenté R, Herrmann D, editors. *Retraining Cognition: Techniques and Applications*. Gaithersburg, MD: Aspen; 1996.
60. Burgess N, Maguire EA, Spiers HJ, O'Keefe J. A temporoparietal and prefrontal network for retrieving the spatial context of lifelike events. *Neuroimage* 2001;14(2):439–453.
61. Hoffman HG, Garcia-Palacios A, Thomas AK, Schmidt A. Virtual reality monitoring: Phenomenal characteristics of real, virtual, and false memories. *CyberPsychol Behav* 2001;4(5):565–572.
62. Johnson MK, Raye CL. Reality monitoring. *Psychol Rev* 1981;88:67–85.
63. Rothbaum O, Hodges LF, Kooper R, Opdyke D, Williford JS, North M. Effectiveness of computer-generated (virtual reality) graded exposure in the treatment of acrophobia. *Am J Psychiatry* 1995;152(4):626–628.
64. North MM, North SM, Coble JR. Virtual reality therapy: An effective treatment for psychological disorders. In: Riva G, editor. *Virtual Reality in Neuro-Psycho-Physiology: Cognitive, Clinical and Methodological Issues in Assessment and Rehabilitation*, vol 44. Amsterdam: ISO Press; 1997. p 59–70.
65. Bullinger AH, Roessler A, Muller-Spahn F. Three-dimensional virtual reality as a tool in cognitive-behavioral therapy of claustrophobic patients. *CyberPsychol Behav* 2000;3:387–392.
66. Kahan M, Tanzer J, Darvin D, Borer F. Virtual reality-assisted cognitive-behavioral treatment for fear of flying: Acute treatment and follow-up. *CyberPsychol Behav* 2000; 3(3):387–392.
67. Gourlay, Lun KC, Lee YN, Tay J. Virtual reality for relearning daily living skills. *Int J Med Inform* 2000;60(3):255–261.
68. Christiansen C, Abreu B, Ottenbacher K, Huffman K, Masel B, Culpepper R. Task performance in virtual environments used for cognitive rehabilitation after traumatic brain injury. *Arch Phys Med Rehabil* 1998;79(8):888–892.
69. Lee JH, Ku J, Cho W, Hahn WY, Kim IY, Lee S-M, Kang Y, Kim DY, Yu T, Wiederhold BK, Wiederhold MD, Kim SI. A virtual reality system for the assessment and rehabilitation of activities of daily living. *CyberPsychol Behav* 2003;6(4): 383–388.
70. Zhang L, Abreu BC, Masel B, Scheibel RS, Christiansen C, Huddleston N, Ottenbacher KJ. Virtual reality in the assessment of selected cognitive function after brain injury. *Am J Phys Med Rehabil* 2001;80(8):597–604.
71. Grealy MA, Johnson DA, Rushton SK. Improving cognitive function after brain injury: The use of exercise and virtual reality. *Arch Phys Med Rehabil* 1999;80(6):661–667.
72. You SH, Jang SH, Kim Y-H, Hallett M, Ahn SH, Kwon Y-H, Kim JH, Lee MY. Virtual reality-induced cortical reorganization and associated locomotor recovery in chronic stroke: An experimenter-blind randomized study. *Stroke* 2005;36:1166–1171.
73. Gaggioli A, Morganti F, Walker BA, Meneghini A, Alcañiz M, Lozano JA, Montesa J, Gil JA, Riva G. Training with computer-supported motor imagery in post-stroke rehabilitation. *CyberPsychol Behav* 2004;7(3):327–332.
74. Strickland D, Marcus LM, Mesibov GB, Hogan K. Brief report: Two case studies using virtual reality as a learning tool for autistic children. *J Autism Dev Disord* 1996;26(6):651–659.
75. McComas J, Mackay M, Pivik J. Effectiveness of virtual reality for teaching pedestrian safety. *CyberPsychol Behav* 2002;5(3): 185–190.
76. Wald J, Liu L, Hirsekorn L, Taylor S. The use of virtual reality in the assessment of driving performance in persons with brain injury. *Stud Health Technol Inform* 2000;70:365–367.
77. Wald J, Taylor S. Efficacy of virtual reality exposure therapy to treat driving phobia: A case report. *J Behav Ther Exp Psychiatry* 2000;31(3–4):249–257.
78. Cicerone KD, Dahlberg C, Malec JF, Langenbahn DM, Felicetti T, Kneipp S, Ellmo W, Kalmar K, Giacino JT, Harley P. Evidence-based cognitive rehabilitation: Updated review of the literature from 1998 through 2002. *Arch Phys Med Rehabil* 2005;86(8):1681–1692.
79. U.S. Department of Health and Human Services. *Delivering on the Promise: Report of Federal Agencies' Actions to Eliminate Barriers and Promote Community Integration*. Executive Order No. 13217, Washington, DC: U.S. Department of Health and Human Services; 2002.

See also COMMUNICATION DEVICES; COMMUNICATIVE DISORDERS, COMPUTER APPLICATIONS FOR; ENVIRONMENTAL CONTROL; HOME HEALTH CARE DEVICES.

REHABILITATION, ORTHOTICS IN

LOREN LATTA
University of Miami
Coral Gables, Florida

INTRODUCTION

Paralysis

Muscle weakness resulting from various forms of paralysis is the major type of disability treated with orthotic devices. Paralysis may be acute or chronic to varying degrees depending on the state of the disease causing paralysis. The degree of paralysis or paresis can change with time as recovery is accomplished through therapy or healing of the physiological cause. Thus, the orthotic need for support, control of joint motion, and so forth, may change with time for many of these patients. Orthotic devices in general are designed for long-term use because many of the designs were developed for chronic applications to polio patients in the late 1940s and 1950s. With many new applications and new materials, a variety of devices have been developed that are ideal for short-term applications to more acute conditions, for example, stroke, head injury, spinal cord injury, and fractures.

Neuromuscular Control

Patients who suffer loss of neuromuscular control in the form of spasticity, paralysis of isolated muscle groups, and/or recovery of neuromuscular function resulting from regeneration of neural tissue, which requires relearning of coordination, are often helped with orthotic devices. Although most of these applications are short term, a few are chronic, such as in cerebral palsy, established spinal cord injury, and head injury.

Deformity

Another common use of orthotics in a growing number of cases is for musculoskeletal deformity. The first type of deformity is related to mechanical instability of the limbs or spine. Mechanical instability can result from soft tissue or skeletal injury or from degenerative joint diseases that cause chronic and progressive instability of the musculoskeletal system. One form of soft-tissue injury relating to the instability of the skeleton is caused by surgical reconstruction of the joints. Therefore, many orthotic devices are used for stabilization postoperatively when the spine or upper or lower limb joints have been reconstructed. The surgery often causes necessary damage to the stabilizing soft-tissue structures that often adds to instability of prosthetic components that require bone healing for final stabilization. The second type of deformity is due to the growth or remodeling disturbances in the skeleton. Orthotic applications are both acute and chronic in problems relating to musculoskeletal deformity. Most chronic applications or orthotics are in progressive deformities resulting from degenerative joint diseases such as rheumatoid arthritis, osteoarthritis, hemophilia, and diabetes. Orthoses generally are applied in these instances to prevent progressive deformity and any resultant mechanical instability in the

skeleton that it may cause. In realigning limb mechanics, orthoses may also prevent some of the disabling pain of degenerative joint diseases. Acute applications of orthoses for deformities include the treatment of fractures and soft-tissue injuries about the joints and postoperative protection for fracture stabilization or joint reconstructive or arthrodesis procedures.

DEVICES

Orthoses have an almost infinite variety of materials, designs, and constructions. Many factors contribute to this variety: (1) performance criteria, (2) available materials, (3) skills of the orthotist, and (4) desires of the patient, surgeon, and therapist.

Performance Criteria

As mentioned, many devices are used in chronic applications with significant loading and require great

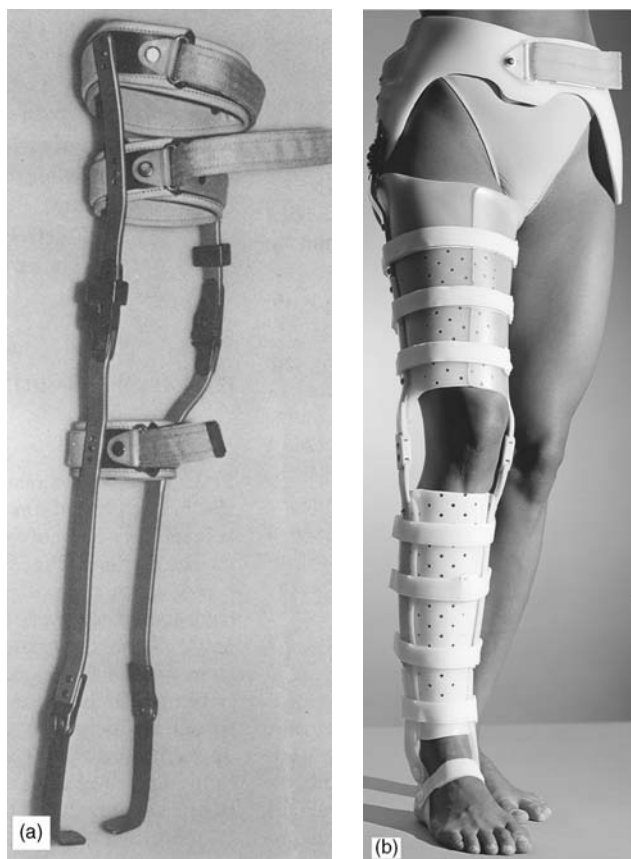


Figure 1. (a) This KAFO is designed for chronic use. It has metal uprights connected to metal bands covered with leather and padding and a “caliper” type of stirrup that attaches to the shank of an orthopedic shoe. All parts are custom fit from mostly prefabricated components. (b) The other KAFO is designed for short-term use and has a thermoplastic thigh and leg interface with the soft tissue, connected to a plastic knee hinge suspended with a plastic posterior shoe insert. All parts are prefabricated in standard sizes. Reprinted with permission from Mamed Orthopaedic Systems.

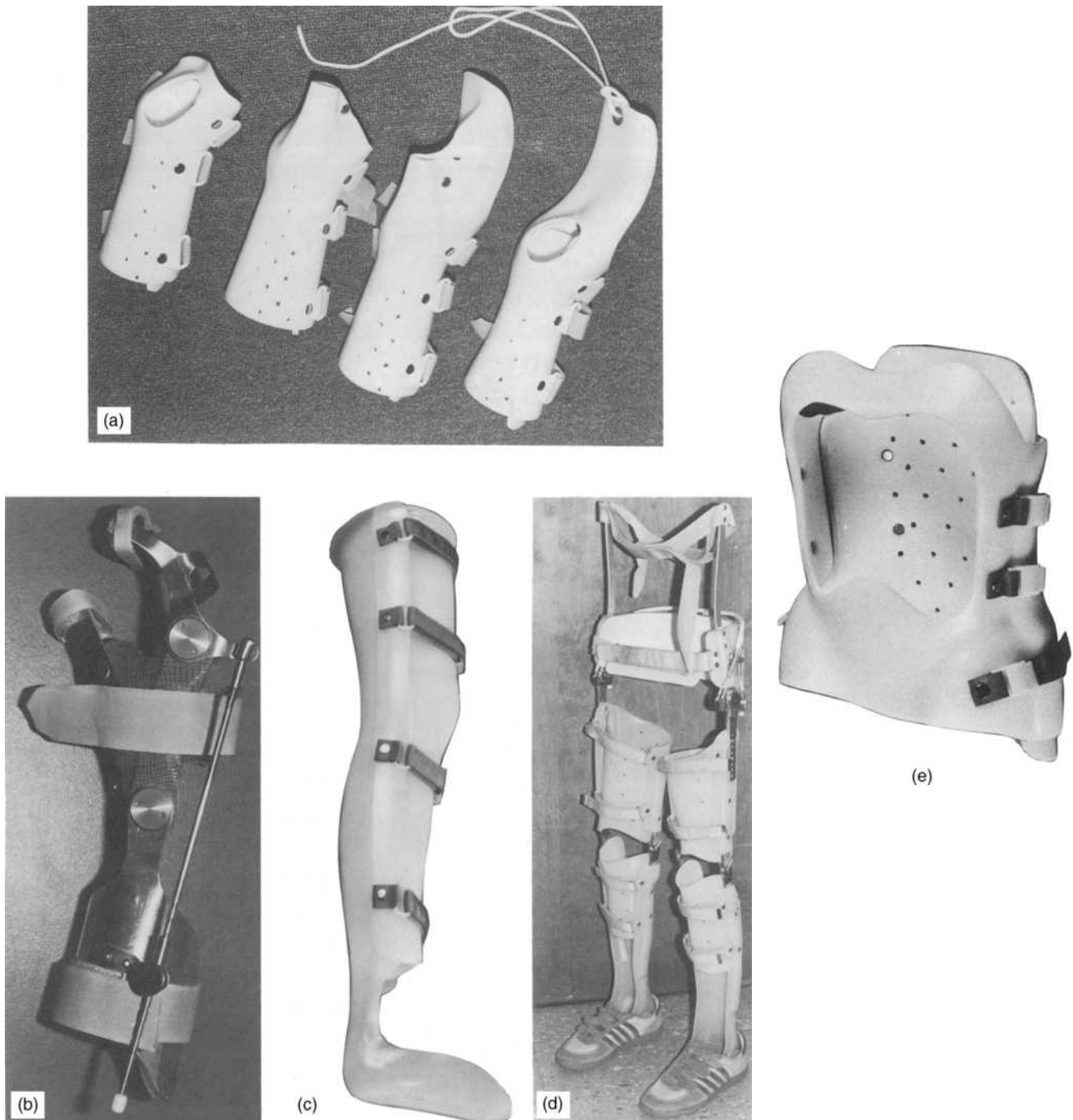


Figure 2. Splints that are used for short-term or long-term immobilization of joints are typically of a simple design and can be prefabricated as in (a) WHO (reprinted with permission from Maramed Orthopaedic Systems) or custom fabricated as in (c) KAFO or (e) TLSO. Orthoses for similar parts of the anatomy with much more complex control criteria have typically more complex designs as in the WHO in (b), which provides a tenodesis action to close the fingers with wrist extension, and the complex reciprocator, bilateral HKAFO in (d), which provides stability for the knee and ankle while providing assistance to hip flexion.

fatigue resistance. In other instances, the loading conditions are slight, but the application is long term; thus, the primary requirement is good compatibility with the skin. Applications for very short-term use under either heavy or light loading conditions are much

less demanding on the materials and design of the device (Figs. 1–3).

Many devices simply act as a splint to immobilize musculoskeletal structures for a short time (Figs. 2 and 7); the design for such a device is thus relatively simple. Other

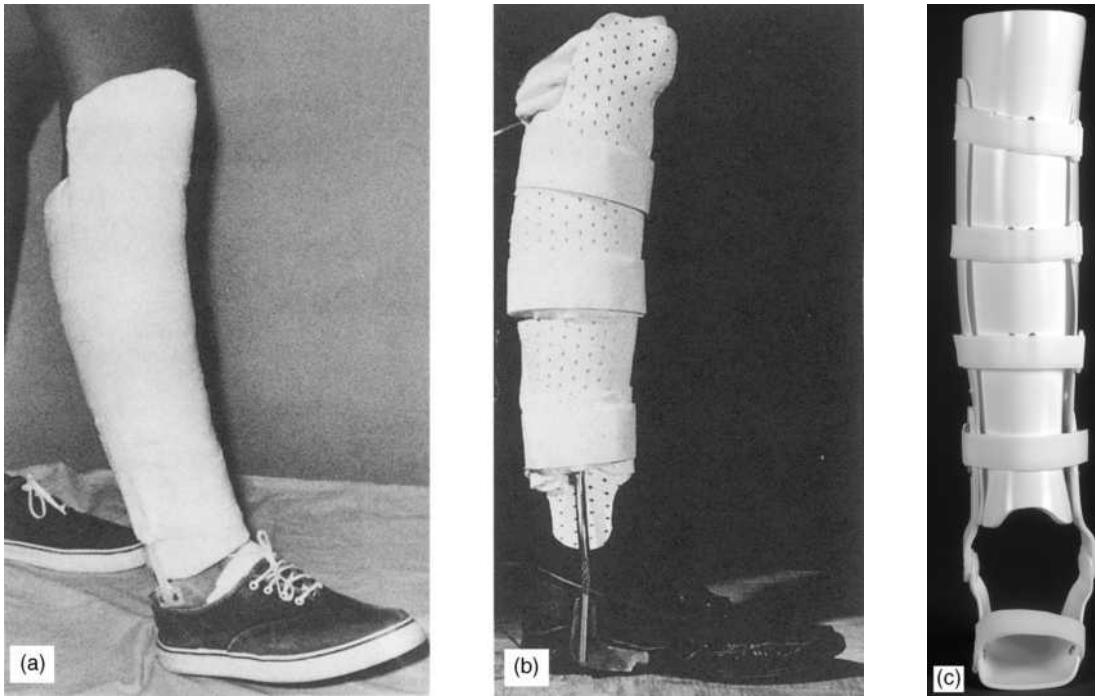


Figure 3. Many material options can accomplish the same performance criteria for a particular orthotic prescription. In this example, a tibial fracture orthosis (AFO) is constructed by three different techniques to accomplish the same end result. Plaster can be molded directly to the patient's limb and attached to a prefabricated ankle joint (a), an isoprene thermoplastic material can be molded directly to the patient's limb attached to a metal ankle joint incorporated into a stirrup permanently attached to the shoe (b), or prefabricated components can be applied with hand trimming and assembly (c). Reprinted with permission from Maramed Orthopaedic Systems.

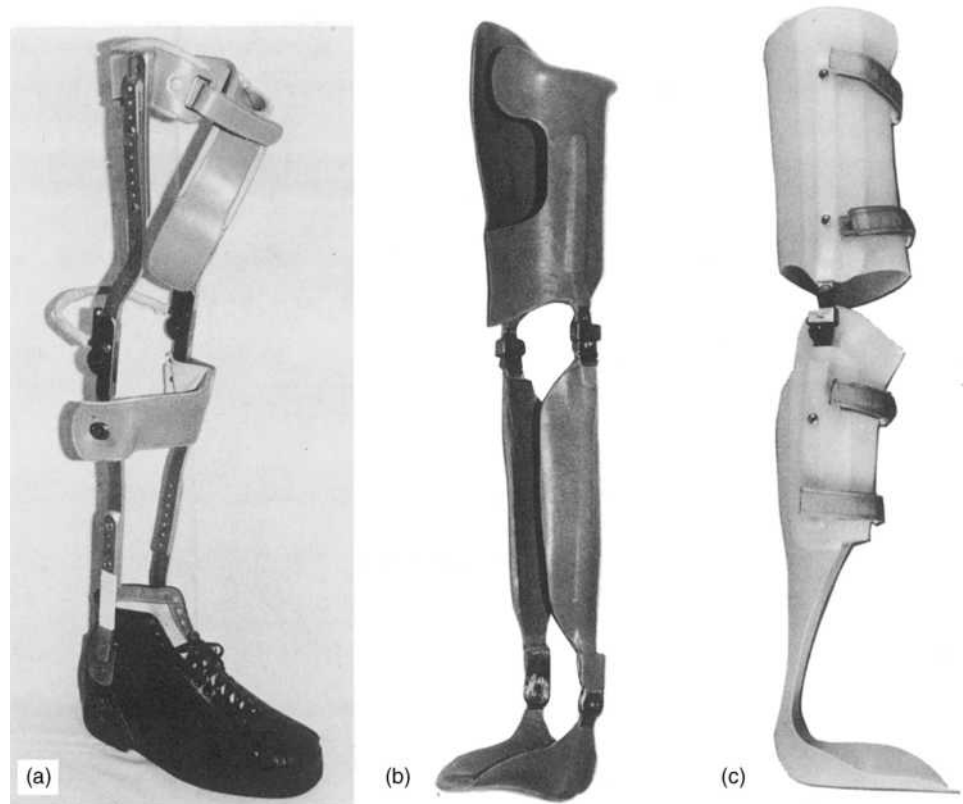


Figure 4. Typical fabrication techniques for orthoses include the use of metal uprights connected by metal bands with leather and padding for covering and skin interface (a), or custom laminated thermosetting plastic with fiber-glass or fabric reinforcement incorporating prefabricated joints (b), or custom molded thermoplastic sleeves incorporating prefabricated joints (c).

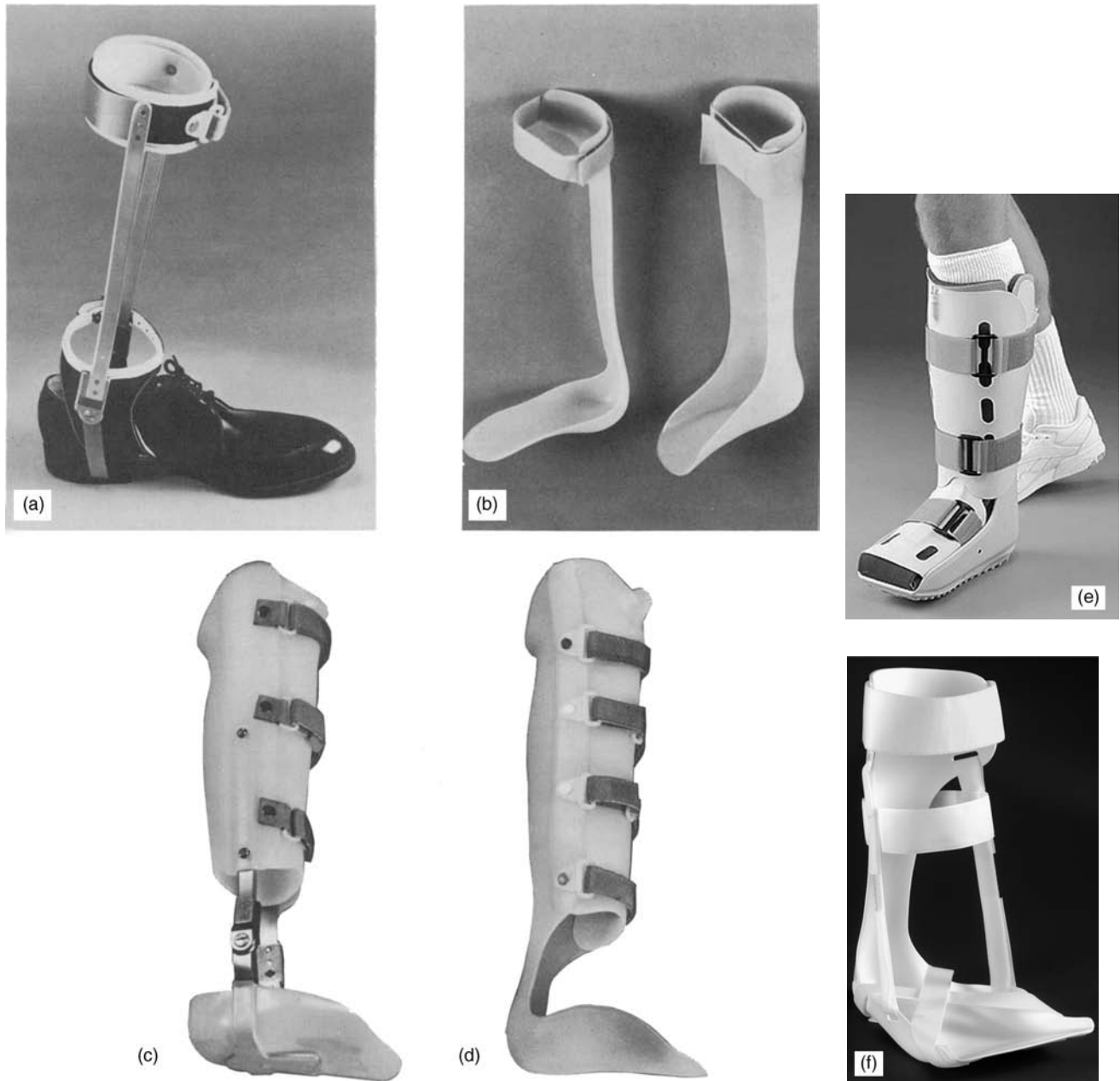


Figure 5. Various combinations of prefabricated and custom-fabricated components can be used to accomplish a particular prescription criterion as in this series of examples of AFOs. Prefabricated components are used for the stirrup connection to the shoe and the modular dorsiflexion assist ankle joint that are attached to custom-formed metal bands, and custom-fabricated leather cuffs and medial T-strap (a). These thermoplastic, posterior leaf AFOs (b) demonstrate totally prefabricated design (left, reprinted with permission from Maramed Orthopaedic Systems) for dorsiflexion assist, custom fabrication with complete limitation, or ankle motion (right). The patellar tendon bearing orthosis (PTB) is designed to transmit axial loads as well as to control ankle and subtalar motion. This example uses a custom-molded thermoplastic soft-tissue interface with the calf, incorporating metal uprights into modular, prefabricated limited motion-ankle joints connected to a stirrup custom formed from prefabricated components incorporated into a “UCB”-type shoe insert custom molded from a thermoplastic (c). A similar type of PTB orthosis uses totally custom-fabricated thermoplastic sleeves with a posterior leaf-limited-motion ankle control attached to a shoe insert (d). The CAM walker, which is totally prefabricated, provides a “roll over” sole to accommodate ambulation with an immobilized ankle, (e, reprinted with permission from Aircast, Inc.). This plantar fasciitis AFO is also completely prefabricated (f, reprinted with permission from Sky Medical, Inc.).

devices have complex prescription criteria if used for assisting, resisting, or holding motion of joints under a variety of temporal and spatial conditions; such designs might be complex (Fig. 2d).

Materials

The choice of material varies according to durability, strength, stiffness, skin compatibility, and fabrication requirements. For short-term applications, prefabricated devices that can be produced by sophisticated manufacturing techniques are often used. Thus, most any commonly available material may be used for short-term custom applications; materials that can be applied directly to the patient will often be used (Figs. 3 and 9–11). Four basic forms of such material exist: plaster, fabric reinforced, fiberglass reinforced, and solid sheets. Some are thermosetting materials activated by promoters or moisture; some are thermoplastics. Devices for long-term use tend to be custom fabricated from materials that are formed to measurements or molds of the patient. The materials are typically thermoplastic, laminated thermosetting materials with carbon fiber, fiberglass or fabric reinforcement, leather, fabric, aluminum, stainless steel, steel, and foam plastics and elastomers (Figs. 1–13).

Description of Conventional Devices

Tables 1–3 describe in a general nature the types of components that make up conventional orthoses and the functional controls and typical materials used. In general, the soft-tissue interfacing components tend to be custom fabricated, the structural members tend to be prefabricated, and the joint mechanisms tend to be prefabricated and modular. It is possible, however, to find prefabricated components and preassembled prefabricated devices in each of these areas. It is also common to find orthotic devices constructed completely from custom-fabricated components integrated together in a totally custom device for a given patient (Figs. 1–3, 5, 7, and 10).

Under functional controls, the word “feedback” is used to indicate that certain proprioceptive feedback signals may be incorporated into these components to give information to the patient regarding the load on the device and/or the position of the device in space. This is sometimes helpful to supplement loss of normal proprioception in the limb. It is possible for feedback mechanisms; EMG-, load-, OR microprocessor-activated locks; and/or resistive or assistive mechanisms at all of these joints to be applied; however, the tables simply reflect those applications that are used in relatively common practice in the field of orthotics. Applications not listed here, to the knowledge of the authors, are simply research

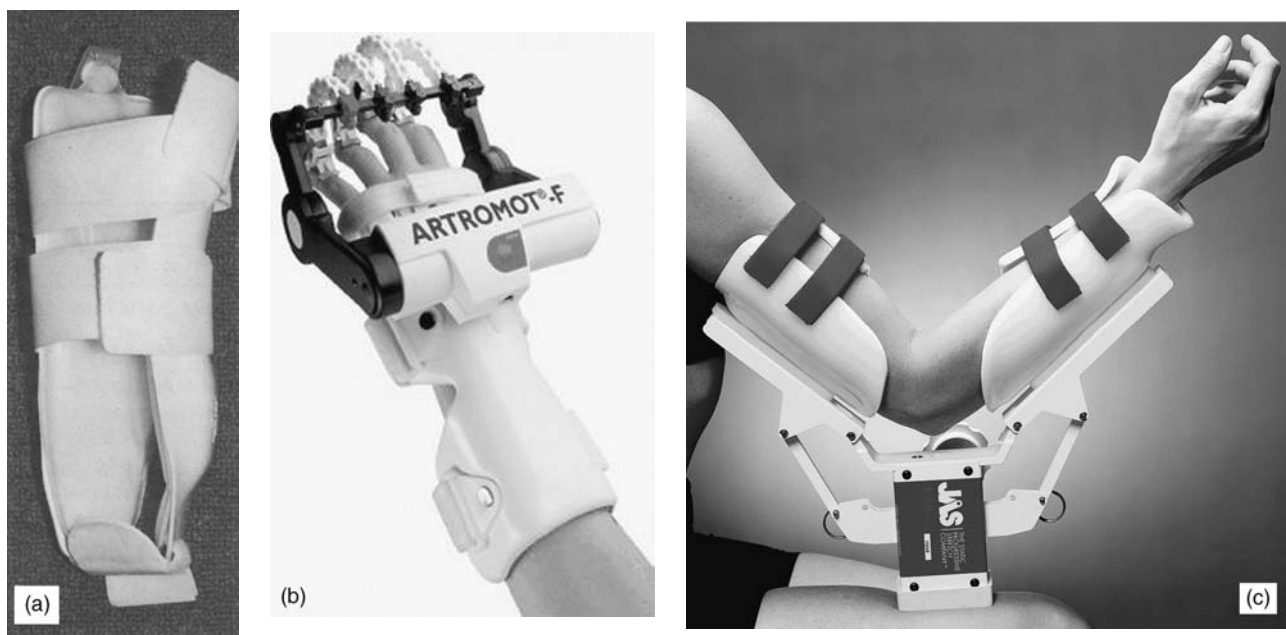


Figure 6. Many types of devices use components that are similar to classic orthoses, but at this time are considered by the authors not to be “classic” orthoses because they are not fabricated or applied by orthotists. An example is this acutely applied air splint for first aid or emergency, designed to control edema and limit motion at the ankle joint for short-term use (a). Orthotic-like devices are used to supplement various therapeutic regimens in the rehabilitation of patients post-injury and post-surgery. This example provides continuous passive motion to the joints through powered systems attached to orthotic-like components (b, reprinted with permission from Orthomed Medizintechnik, GmbH), and the other one provides a corrective force to gradually regain motion in joints with contractures, (c, reprinted with permission from Joint Active Systems, Inc.).

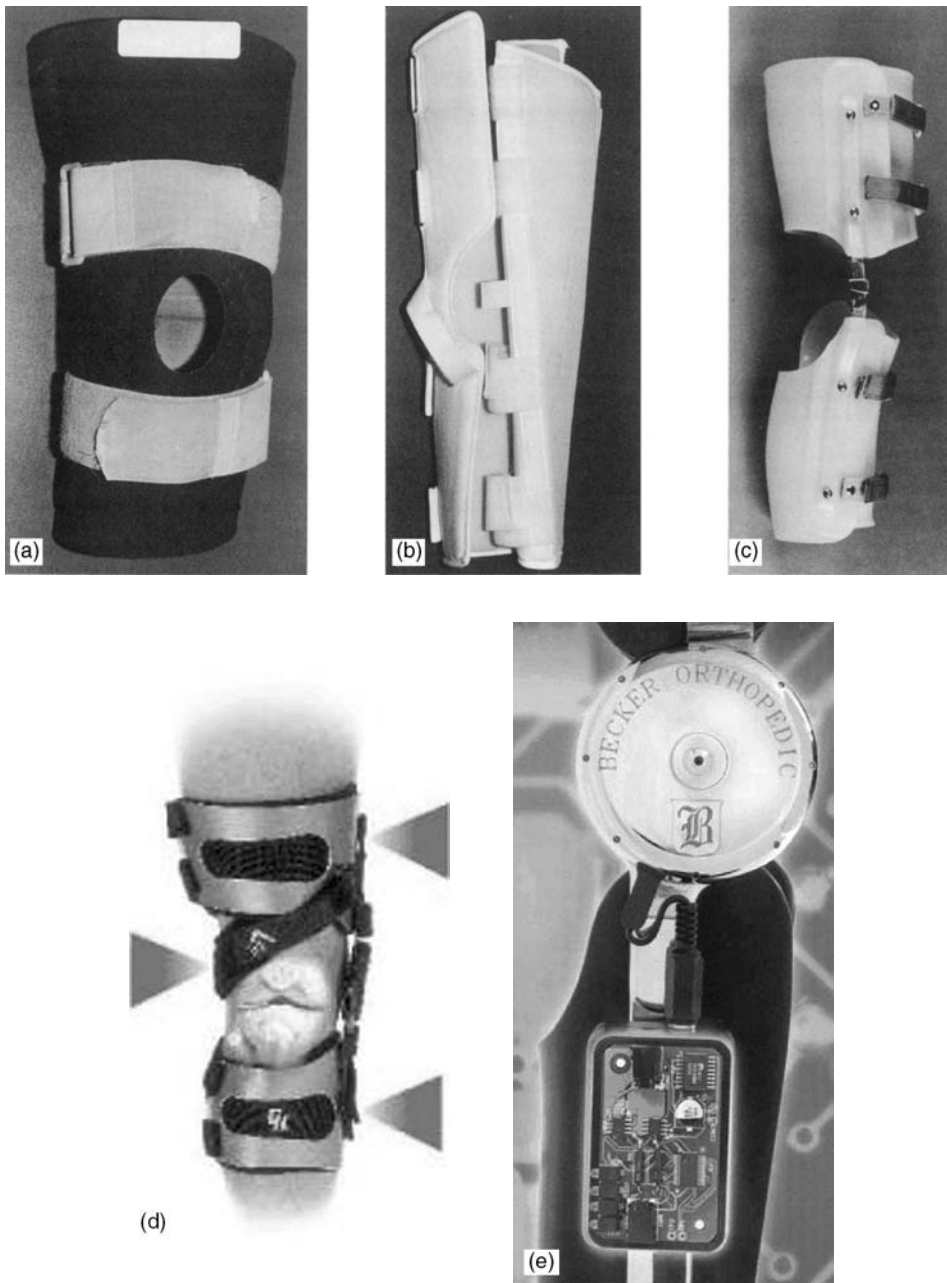


Figure 7. Most KOs are designed for the prevention or protection of knee injuries. Simple soft devices to limit knee motion (a), splints to immobilize the knee temporarily (b), custom orthoses with knee mechanisms to control motion and provide support to the knee (c), and orthoses that provide adjustable corrective forces or moments to the knee to alter knee mechanics (d), reprinted with permission from Generation II USA, Inc.), and even powered knee mechanisms with microprocessor controls, (e), reprinted with permission from Becker Orthopedic, Inc.).

applications not in common use. The description of typical design types and materials also reflects the opinion of the authors on the systems commonly used in orthotics today. Many sophisticated designs and materials are being used in research and may be commonly applied in the future.

Some conventional devices do not strictly meet the nomenclature system because they do not cross a joint or control a joint. They simply are a sleeve that compresses soft tissue to stabilize a fracture or protect a limb with soft-tissue injuries. These devices are only for short-time use until an injury can heal. Examples are shown in Fig. 14.

FUNCTIONAL EVALUATION

Evaluation of musculoskeletal function is the key to adequate prescription criteria or performance criteria, for conventional orthoses. Communication of this evaluation by the orthopedic surgeon to the orthotist is an important step in obtaining agreement on, and optimum use of, orthoses in the rehabilitation of the patient. To this end, the Committee on Prosthetics and Orthotics of the American Academy of Orthopedic Surgeons has devised the technical analysis form that is used for recording the functional evaluation of the patient, documenting the abilities and disabilities of the patient, and forming a

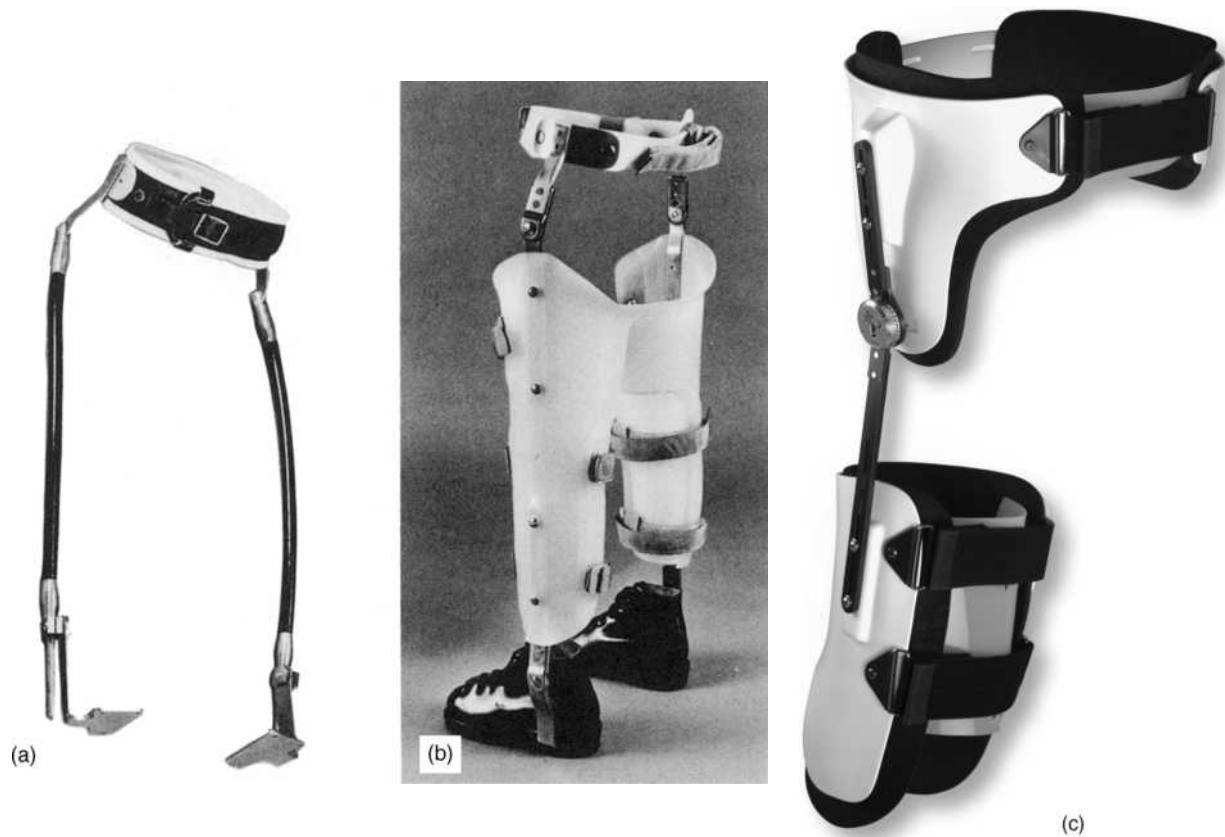


Figure 8. HKAFOs are generally used for the control of deformities. Axial rotation alignment of the lower limbs can be controlled with an HKAFO commonly called a twister brace (a), which uses a flexible cable to provide free motion at the knee and ankle in all degrees of freedom while applying a constant axial torque for rotation correction. An HKAFO for control of valgus at the knee provides three-point support through the soft-tissue interfacing sleeves with as little restriction to ankle and hip flexion and extension as possible (b). An HO is often used after total hip replacement to provide an abduction resistance to the hip during flexion to help prevent hip dislocation until the soft-tissue healing is completed (c, reprinted with permission from Sky Medical, Inc.).

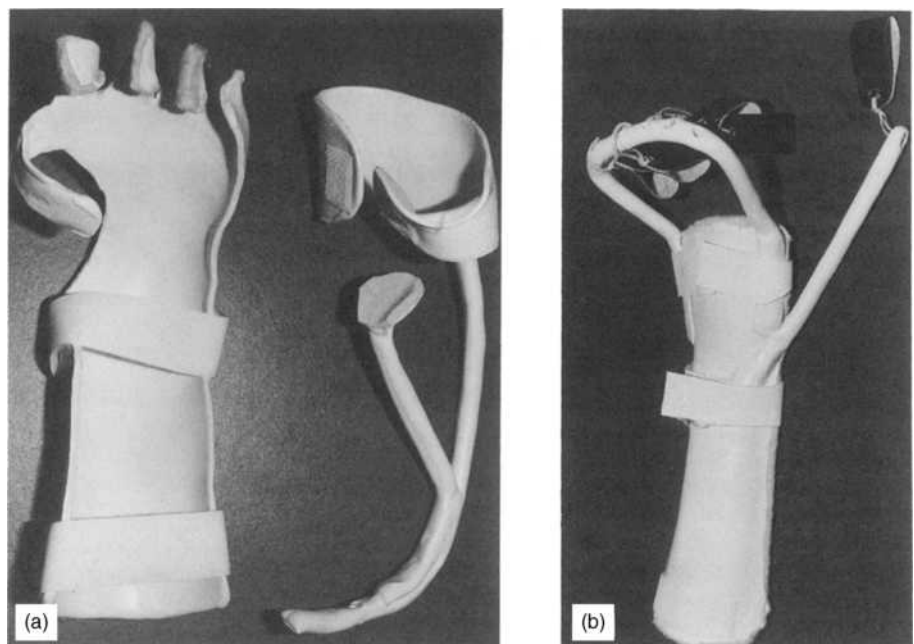


Figure 9. WHOs can be of static (a) or dynamic (b) variety to control, support, resist, or assist motion of the fingers, hand, and/or wrist. These examples are made of low-temperature thermoplastics that can be applied directly to the patient.

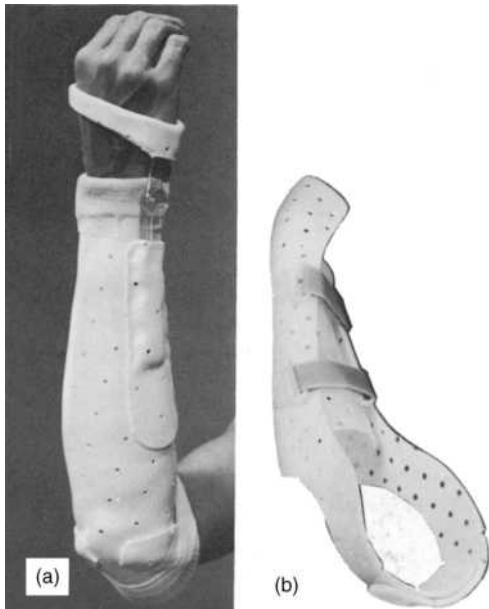


Figure 10. EWHOs are most typically of the custom-fabricated type (a), but they are also available in prefabricated forms (b, reprinted with permission from Maramed Orthopaedic Systems), as shown in these examples of devices for the treatment of Colles' fractures.

recommendation for the orthosis. In a uniform, simplified system, these details can be communicated to the orthotist for construction and fitting of the device and rehabilitation of the patient. Evaluation of each segment of the musculoskeletal system is divided into parts as described in Fig. 14, which shows a typical evaluation form. Note that the range of motion of each joint is indicated on the background of the normal range of motion for that joint and there are places for recording the static position of each joint and part of the skeleton and the deformities that exist in the skeletal structures. The first page of the technical analysis form contains didactic information on the impairments and

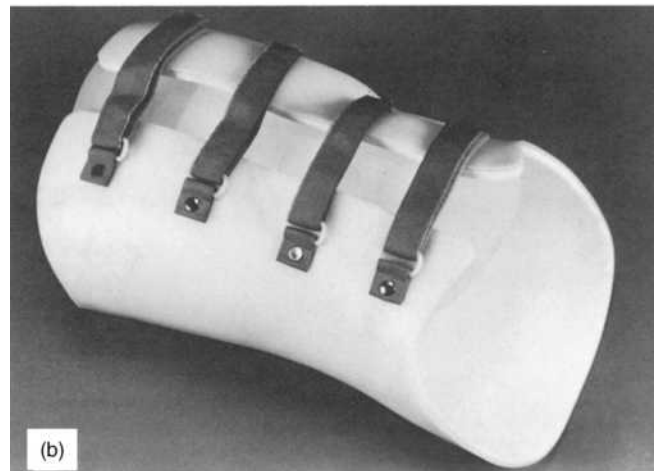
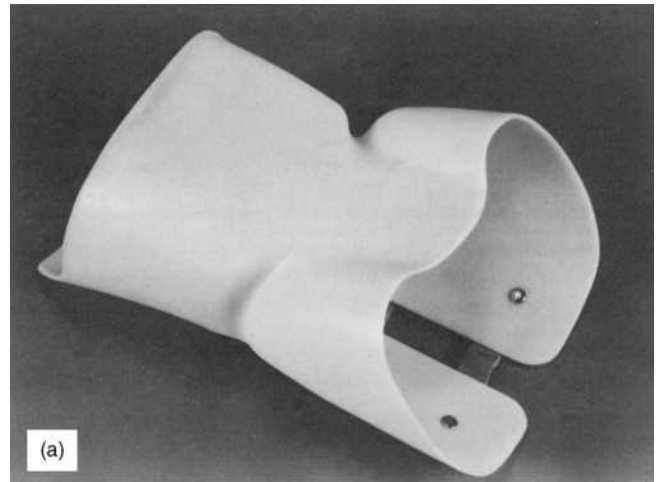


Figure 12. TLSOs are used in a wide variety of forms, most typically a custom-fabricated thermoplastic design for control of spinal deformities, such as this low-profile scoliosis TLSO for control of spinal curvature (a). This anterior closing TLSO is used for instability of the low back or low back pain (b).

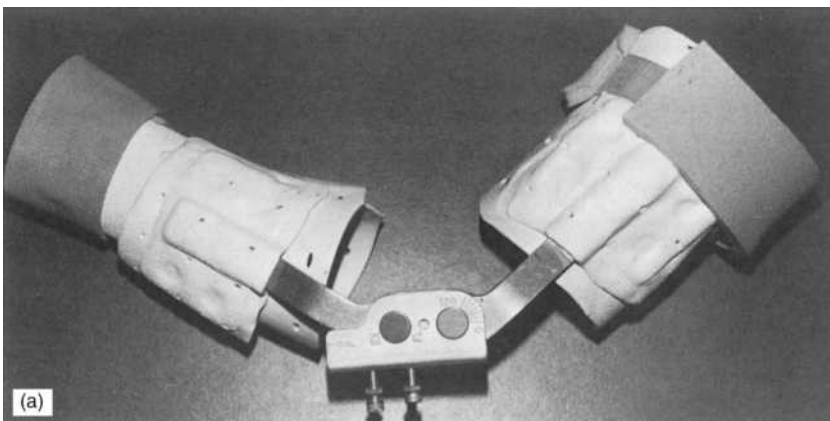


Figure 11. EOs typically are of the custom-fabricated type with prefabricated components for the control of elbow flexion and extension during rehabilitation post-injury (a) or elbow reconstruction, but many are totally prefabricated designs (b, reprinted with permission from Aircast, Inc.).



Figure 13. This CTLSO is designed to provide corrective forces for spinal curvature with thermoplastic skin interfacing components for the pelvis and chin and prefabricated, metal uprights (a) This halo orthosis is designed to protect the cervical spine after acute injury and after spinal fusion (b). The skull pins in the proximal portion of the orthosis are percutaneous components that anchor into the skull through the skin for positive control of head position.

a nomenclature system for their description on the physical evaluation form. A legend for description of fixed deformities, alterations of range of motion, and description of skeletal as well as muscular abnormalities is provided in Fig. 16 for use in the boxes in Fig. 15 for description of the physical examination. On the reverse side of this form, the treatment objectives are outlined and the orthotic recommendation includes a description of the controls for each part of the musculoskeletal system that are recommended as performance criteria for the orthosis. The orthotist is then given the choice of how to accomplish these recommendations in orthotic design and materials. The recommendation form is completed with standard nomenclature describing the control of the joints and musculoskeletal structures: F designates free motion, A assistance, R resistance, S stop, H hold, and L lock controls on motion. A “v” accompanying any of these designations will recommend a variable control of the type described for that degree of freedom (Fig. 17). The degrees of freedom are described across the top of the table for flexion, extension, abduction, adduction, internal and external rotation, and axial loading on the skeletal parts. Also note that a standard nomenclature system is developed to describe the joints that are encompassed by the orthosis. Thus, FO designates an orthosis that controls the foot; AFO describes an orthosis that controls the ankle and the foot, and so forth. In addition to the designations shown on the orthotic recommendation table, it is possible to have a device that simply

Table 1. Lower Limb Orthotic Devices

Component	Functional control		Typical Design Types	Typical Materials
	Passive	Active		
Soft-tissue interface	Alignment to structural members and joints	Feedback load/position	Wraparound or interlocking shells; adjustable pads and straps	Fabric, leather, hand laminates, thermoplastics, foam or elastomer polymers
Structural members	Alignment to joints and soft tissue interfaces, extension blocks	Feedback load/position	Modular uprights; reinforced shell; extension of joints and bands; single axis; 1 or 2 DOF, cam or drop lock	Metals, hand laminates, prepregs, thermoplastics
Hip joint	Lock, stop, free, assistive, resistive motions	Manual-activated lock	Single axis; 1 or 2 DOF; cam or drop lock	Metals, reinforced polymers, thermoplastics
Knee joint	Lock, stop, free, assistive, resistive motions, adjustable varus/valgus corrective force	Feedback position, load-, manual-, or EMG-activated lock or resistance	Single- or multiaxis; 1 DOF; variable stop; cam, link, or drop lock; friction resistance; elastic assistance; single lateral or posterior upright; medial-lateral upright	Metals, reinforced polymers, thermoplastics
Ankle joint	Lock, stop, free, assistive, resistive motions	Feedback position, load-, manual-, or EMG-activated lock or resistance	Single- or multiaxis; 1 DOF variable stop; cam; friction resistance; elastic assistance; single lateral or posterior upright; medial-lateral upright	Metals, reinforced polymers, thermoplastics
Foot	Alignment		Shoe insert: caliper or stirrup attached to shoe with shank	Metals, reinforced polymers, thermoplastics
FES or stimulation	Electrode placement	Microprocessor, load- or EMG-activated stimulus	Implantable: electrodes, self-contained power and control module, stimulator module RF to external power External: electrodes, stimulator, and controls	Silicone-coated silver-braided steel, silver-impregnated silicone, SS electrodes, epoxy- or silicone-encapsulated electronic components

^aFor examples, see Figs. 1–5, 7, and 8. DOF, degree(s) of freedom; FES, functional electrical stimulation.

Table 2. Upper Limb Orthotic Devices

Component	Functional control		Typical Design Types	Typical Materials
	Passive	Active		
Soft-tissue interface	Alignment to joints and structural members	F feedback load/position	Wraparound or interlocking shells; adjustable padding and straps	Fabric, leather, hand laminates, thermoplastics, foam or elastomer polymers
Structural-members	Alignment to soft tissue interfaces and joints, extension blocks	F feedback load/position	Modular reinforced shells; extension of joints and bands	Metals, hand laminates, prepregs, thermoplastics
Shoulder joint	Lock, stop, free, assistive, motions	Manual-or EMG-activated lock or assistance	Single axis; 1 or 2 DOF; electrical or pneumatic motor; manual or elastic assistance	Metals and reinforced polymers
Elbow joint	Lock, stop, free, assistive, motions	Manual-or EMG-activated lock or assistance	Single- or multiaxis; 1 DOF; electrical or pneumatic motor; manual or elastic assistance; drop or cam lock, extension or mechanical joint	Metals and reinforced polymers
Wrist joint	Lock, stop, free, assistive, motions	Manual-or EMG-activated lock or assistance	Single- or multiaxis; 1 or 2 DOF; electrical or pneumatic motor; manual or elastic assistance; drop or cam lock, extension or mechanical joint	Metals and reinforced polymers
Hand	Lock, stop, free, assistive, motions	Manual-or EMG-activated lock or assistance	Single- or multiaxis; 1 or more joints; 1 DOF; electrical or pneumatic motor; manual or elastic assistance; drop or cam lock, extension or mechanical joint stop in 1 or more DOF	Metals and reinforced polymers
FES orstimulation	Electrode placement	Microprocessor-, load- or EMG-activated stimulus	Implantable: electrodes, self contained power and control module, stimulator module RF to external power External: electrodes, stimulator, and controls	Silicone-coated silver-braided steel, silver-impregnated silicone, SS electrodes, epoxy- or silicone-encapsulated electronic components

^aFor examples, see Figs. 2 and 9–11. DOF, degree(s) of freedom.

Table 3. Spinal Orthotic Devices

Component	Functional control		Typical design types	Typical materials
	Passive	Active		
Soft-tissue interface	Alignment to structural members	F feedback load	Wraparound or interlocking shells; adjustable pads and straps	Fabric, leather, hand laminates, thermoplastics, foam or elastomer polymers
Structural members	Alignment to soft tissue interfaces	F feedback load	Upright adjustable superstructures; reinforced shells; percutaneous pins and belts	Metals, hand laminates, thermoplastics
Electrical stimulation	Electrode placement	Microprocessor-activated stimulus	Implantable: electrodes, self contained power and control module, stimulator module RF to external power External: electrodes, stimulator, and controls	Silicone-coated silver-braided steel, silver-impregnated silicone, SS electrodes, epoxy- or silicone-encapsulated electronic components

^aFor examples, see Figs. 2, 12, and 13.



Figure 14. Fracture orthoses that do not span a joint are used for selected humeral fractures (a) and selected isolated ulnar fractures and some soft-tissue injuries (b, reprinted with permission from Sky Medical, Inc.).

controls one joint and does not include all joints distal to it. For example, an orthosis that encompasses and controls only the knee joint is termed a “KO”. A similar nomenclature system is shown for the upper limb and for the spine in Figs. 18 and 19, which show the orthotic recommendation

and treatment objective portions of the technical analysis form for each of those applications. For examples of orthoses fitting each nomenclature descriptor, see the figures associated with each table. Many orthotic facilities and surgeons have created their own forms for orthotic

Figure 15. This portion of the technical analysis form for the lower limb demonstrates the graphic manner in which a complete passive and active evaluation of the extremity is recorded to document both normal and abnormal behavior. Reproduced by permission from the American Academy of Orthopaedic Surgeons, *Atlas of Orthotics*, 2nd ed., St. Louis, C. V. Mosby, 1985.

LEGEND		
	= Direction of Translatory Motion	
	= Abnormal Degree of Rotary Motion	
	= Fixed Position	
	= Fracture	
	Volitional Force (V)	Proprioception (P)
	N = Normal	N = Normal
	G = Good	I = Impaired
	F = Fair	A = Absent
	P = Poor	
	T = Trace	
	Z = Zero	
	Hypertonic Muscle (H)	D = Local Distension or Enlargement
	N = Normal	
	M = Mild	
	Mo = Moderate	
	S = Severe	
		= Pseudarthrosis
		= Absence of Segment

Treatment Objectives: Prevent/Correct Deformity Improve Ambulation
 Reduce Axial Load Fracture Treatment
 Protect Joint Other _____

ORTHOTIC RECOMMENDATION

LOWER LIMB	FLEX	EXT	ABD	ADD	ROTATION		AXIAL LOAD
					Int.	Ext.	
HKAO Hip							
KAO Thigh							
Knee							
AFO Leg							
Ankle	(Dorsi)	(Plantar)					
Subtalar					(Inver.)	(Ever.)	
FO Foot							
Midtarsal							
Met. phal.							

Figure 16. The legend for the technical analysis form uses standard descriptors for qualitative and quantitative documentation of the physical examination. Reproduced by permission from the American Academy of Orthopaedic Surgeons, *Atlas of Orthotics*, 2nd ed., St. Louis, C. V. Mosby, 1985.

Treatment Objectives: Prevent/Correct Deformity Improve Function
 Relieve Pain Other _____

ORTHOTIC RECOMMENDATION

UPPER LIMB	FLEX	EXT	ABD	ADD	ROTATION		AXIAL LOAD
					Int.	Ext.	
SEWHO Shoulder							
EWHO Humerus							
Elbow							
Forearm					(Pron.)	(Sup.)	
WHO Wrist			(RD)	(UD)			
HO Hand							
Fingers 2-5	MP						
	PIP						
	DIP						
Thumb	CM					(Opposition)	
	MP						
	IP						

Figure 17. The orthotic recommendation portion of the lower limb technical analysis form provides for description of the treatment objective as well as a prescription recommendation for control of musculoskeletal system by the orthosis. Reproduced by permission from the American Academy of Orthopaedic Surgeons, *Atlas of Orthotics*, 2nd ed., St. Louis, C. V. Mosby, 1985.

Treatment Objectives: Spinal Alignment Motion Control
 Axial Unloading Other _____

ORTHOTIC RECOMMENDATION

SPINE	FLEX	EXT	LATERAL FLEXION		ROTATION		AXIAL LOAD
			R	L	R	L	
CTLSO Cervical							
TLSO Thoracic							
LSO Lumbar							
(Lumbo sacral)							
SIO Sacroiliac							

Figure 18. The upper limb technical analysis form provides for description of treatment objectives and orthotic prescription with the standard nomenclature system. Reproduced by permission from the American Academy of Orthopaedic Surgeons, *Atlas of Orthotics*, 2nd ed., St. Louis, C. V. Mosby, 1985.

prescription, but they include similar features to those in the accepted standards shown here.

OUTCOME

Orthotic devices are designed to improve the function of persons with musculoskeletal disabilities. The goals of treatment are outlined on the technical analysis form for each application. The optimal outcome is obviously achievement of these goals. In many instances, achievement of these goals is related to the provision of functional independence for many persons who would otherwise be partially or totally dependent. Where total or partial functional independence is not feasible, the goal is to provide a better quality of life. When orthoses are applied acutely for temporary stabilization or temporary protection until healing or recovery from a disabling injury can occur, the goal is often an uneventful recovery. In many instances, the patient returns to employment or functional independence, or function is completely restored before the injury heals. After sufficient healing, the orthosis is usually discontinued. In some cases, orthotic care allows the patient to be discharged from the hospital or transferred to a less expensive support system earlier, reducing the cost of medical care. In most short-term applications of orthotic devices, patients are relatively compliant and cooperative and use the orthoses well. Most long-term applications of orthoses are well accepted by patients, but in a higher percentage of chronic (compared with acute) applications, patients use orthoses for limited activities or periods of time and choose to alter the original treatment goals. Such behavior is also not unusual for the users of many external prosthetic devices. Such patients often develop compensatory means of function (i.e., retraining of the contralateral limb) or use other assistive devices (like wheelchairs) to accomplish their personal goals.

For acute applications of orthoses, there is a strong trend toward the greater use of totally prefabricated systems. For chronic applications, very few prefabricated systems have proved to be adequate; the trend here is to use more thermoplastic materials. Patients find these materials more lightweight, cosmetic, and comfortable.

Another trend helping to facilitate these changes is the increased use of central fabrication facilities. Traditionally, orthotists have measured the patient, designed the orthosis, and completely fabricated the orthosis before fitting and training the patient in its use. Today, orthotists are being trained to devote more attention to the measurement, fitting, and training of patients and less attention to the fabrication of devices. This is because most types of devices can be fabricated in a factory with highly skilled technicians using the prescription criteria and the measurements of the orthotist. This trend helps to reduce costs; if the orthotist spends most of his or her time evaluating, fitting, and training the patients, and the technicians fabricate the devices, the orthotist can care for a much greater number of patients and the consistency of fabrication of the devices is improved considerably. Fabrication of a replacement device for a patient already under the care of an orthotist is often

simplified with the availability of a central fabrication facility.

FUTURE

Historically, developments in orthotics have followed developments in the field of external prosthetics. If this continues to hold true, one can anticipate that in the immediate future, orthotics will make increased use of proprioceptive feedback systems, composite materials, microprocessor controls for the dynamics of the orthosis, automation of production facilities and improvements in the design of skin interfacing components, and temporary use of standardized devices for patient training before fabrication and fitting. Also, because orthotic devices are increasingly used for many new applications and means for increasing the number of skilled orthotists needed to meet the demands are not available, there will probably be major increases in the use of prefabricated systems, central fabrication facilities, and/or simplified systems for specific applications that can be handled by other paramedical personnel. There is a growing trend toward greater use of functional electrical stimulation for even the most complex neuromuscular disabilities using microprocessor-controlled multichannel systems to control coordinated muscular activity in limbs with paralysis or severe paresis. Recent advances in electrode design, miniaturized microprocessor systems, and knowledge of musculoskeletal functions are continually producing breakthroughs in research. CAD/CAM methods for automated measurement, modification, and fabrication of custom devices are developing also. The orthotist will find continuing advances in orthotic systems for preventive medicine in sports and the work environment. The rehabilitation team, of which the orthotist is an important member, will include new specialists for biofeedback, rehabilitation engineering, and new branches of therapy and physicians and surgeons from specialties that previously were not involved in rehabilitation.

DEFINITIONS

1. *Ortho.* From the Greek *orthos*, meaning straight or to correct.
2. *Orthosis.* Orthopedic appliance used to straighten, correct, protect, support, or prevent musculoskeletal deformities.
3. *Orthotics (orthetics).* Field of knowledge relating to the use of orthoses to protect, restore, or improve musculoskeletal function. (Note: This field overlaps the field of mobility devices, including wheelchairs, crutches, and special vehicles, for transportation of persons with musculoskeletal disabilities. Mobility aids will not be discussed in this article.)
4. *Orthotist (Orthetist).* A person practicing or applying orthotics to individual patients.

5. *Custom versus prefabricated orthoses.* Custom-fabricated orthoses are devices that have components that are molded specifically to the contours of the musculoskeletal structures of a patient. This fabrication can be accomplished by making molds, tracings, or careful measurements of the patient's anatomy for custom shaping in the fabrication of the device. Prefabricated orthoses are made completely from prefabricated components that are fit to the patient in standard sizes. Some systems are preassembled, and some are not. Prefabricated components may be used in custom-fabricated devices along with custom-fabricated components to provide a custom orthosis.
6. *Modular component.* A component that can be assembled from prefabricated parts and can be disassembled and reassembled in different combinations.
7. *Nomenclature.* A standard set of abbreviations for an orthosis related to the anatomic parts that are to be controlled by the orthosis (see Figs. 14–18); for example, AFO is ankle-foot orthosis.
8. *Free motion.* No alteration or obstruction to normal range of motion of an anatomic joint.
9. *Assistance.* Application of an external force for the purpose of increasing the range, velocity, or force of motion of an anatomic joint.
10. *Resistance.* Application of an external force for the purpose of decreasing the velocity or force of motion of an anatomic joint.
11. *Stop.* Inclusion of a static unit to deter an undesired motion in a particular degree of freedom of an anatomic joint. Variable control parameter that can have multiple adjustments without making a structural change.
12. *Hold.* Elimination of all motion in a prescribed degree of freedom of an anatomic joint or anatomic structure.
13. *Lock.* A device that has an optional mode of holding a particular anatomic joint from motion and that can be released to allow motion when desired.
14. *Paralysis.* Loss or impairment of motor function (paresis, incomplete paralysis).
15. *Volitional force.* Voluntarily controlled muscle activity.
16. *Hypertonicity.* High resistance for muscle to passive stretching.
17. *EMG.* Electromyography is a measure of electrical potential changes caused by muscle contraction/relaxation.

See also CARTILAGE AND MENISCUS, PROPERTIES OF; JOINTS, BIOMECHANICS OF; LIGAMENT AND TENDON, PROPERTIES OF.

RESIN-BASED COMPOSITES

MARK CANNON
Northwestern University
Chicago, Illinois

INTRODUCTION

Composite dental restorative filling materials are synthetic resins that have had a revolutionary impact on dental practice. No other dental materials have stimulated such rapid change in an inarguably short period of time. The advent of resin-based composite occurred during the "Golden Age of Dentistry" and mirrored the overall effect that high technology had on society. Virtually everyone has been made aware of "bonding" or "cosmetic dentistry" by the mass media. The success of "bonding" or "cosmetic dentistry" depended on the development of esthetic dental restorative (filling) materials such as resin-based composites and to the advent of dental adhesives, which allowed for the successful placement of the new esthetic materials.

Composite resins have replaced the previous dental restorative materials for anterior teeth, silicate cements, and unfilled acrylic resins, because of superior physical properties. The overall improvement in physical properties has encouraged a great increase in the clinical use of composite resins. Composite resins are now used in virtually all aspects of dentistry. Their application was first limited to simple cavities in anterior teeth (incisors) or the repair of a broken tooth. However, composite resins are now used to cement orthodontic brackets onto teeth, seal pits and fissures in molars and premolars, splint periodontally or traumatically loosened teeth together for stability, repair not just broken teeth, but also porcelain restorations, revitalize and enhance the esthetic quality of discolored or misshapen teeth. Certainly, no other dental material systems offers such a broad range of applications or greater opportunities for the improvement of dental care. Researchers are working diligently to develop newer resins that possess all the necessary qualities required of an ideal dental restorative material, esthetic by nature and resistant to all the deleterious effects of the oral environment.

FABRICATION

Unfilled Acrylic Resins

The autopolymerizable, unfilled acrylic resin was one of the precursors to the development of composite resin materials. The composite resins were introduced to overcome the problems associated with the clinical use of the unfilled resins. The unfilled acrylic resins were supplied as powder and liquid preparations. The powder consisted of the polymer, polymethylmethacrylate and an initiator, benzoylperoxide (1). The monomer consisted mostly of methyl methacrylate, a cross-linking agent (ethylene dimethacrylate), a tertiary amine, and an inhibitor (methylhydroquinone). Although the unfilled resins were considered to be

Acronym	Organic compounds
Bis-GMA	Bisphenol- α -glycidyl methacrylate

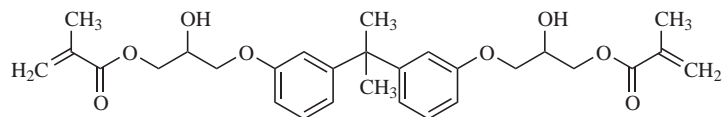


Figure 1. Bis-GMA.

polishable and esthetically acceptable, numerous problems were reported clinically. Pulpal irritation and sensitivity were noted and probably related to microleakage. The microleakage was undoubtedly due to the high degree of polymerization shrinkage and the higher coefficient of thermal expansion of unfilled resins to tooth structure. The expansion and contraction of the unfilled resin restoration would “percolate” deleterious salivary contents leading to pulpal sensitivity, marginal discoloration and secondary caries.

Resin Based Composite

Dentistry has long recognized the need for an esthetic anterior restorative material. The unfilled acrylic resins and the silicate cements were generally considered to be clinical failures and inadequate for many situations commonly seen in dental patients (veneering of discolored teeth, repair of badly fractured teeth, etc.) The composite principle, using filler particles with a proper index of refraction, thermal expansion coefficient similar to enamel and a resin adhesion capability was advocated in 1953 by Paffenbarger et al. (2) The term composite may be defined as a three-dimensional (3D) combination of at least two different materials with a distinct interface (3). A composite resin restorative material consists of three phases: the organic phase (matrix), the interfacial phase (coupling agents), and the dispersed phase (filler particles).

Matrix Phase

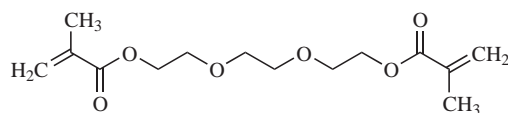
One of the main components of all composites is the addition reaction product of bis(4-hydroxyphenol), dimethylmethane, and glycidylmethacrylate known as “Bis-GMA” (4). Bis(4-hydroxyphenol) is an oligimer with a fairly high molecular weight. Figure 1 shows its generalized structural formula.

Other aromatic dimethacrylates are also used in composite resin materials (5). These oligimers include Bis-MA (2,2-bis[4-(2-methacryloyloxyphenol)] propane), Bis-EMA (2,2-bis [4-(3-methacryloyloxyphenol)] propane) and Bis-PMA (2,2-bis[4-(3-methacryloyloxypropoxy)] phenol] propane. To decrease the high viscosity of the Bis-GMA resin systems, low viscosity liquids, such as TEGDMA (triethyleneglycol dimethacrylate) and EGDMA (ethyleneglycol dimethacrylate) are used. Figure 2 shows the structural formula for TEGDMA.

Inhibitors are necessary to prevent the premature polymerization of the dimethacrylate oligimers and viscosity controllers. An example of an inhibitor would be BHT (2,4,6- tritertiarybutylphenol). Autopolymerizable compo-

site resins (paste mixed with paste or powder mixed with liquid two component systems) utilized a thermochemical initiator, most commonly, benzoylperoxide. The decomposition of benzoyl peroxide results in free radicals that initiate polymerization. The initiator would be present in only one portion of the two-component system. The other portion would contain the accelerator, such as, a tertiary aromatic amine. When the two components of an autopolymerizable composite resin are mixed, the tertiary aromatic amine (e.g., *N,N*-dimethyl *p*-toluidine) interacts with benzoyl peroxide to produce free radicals necessary to initiate the matrix polymerization (6). The main disadvantage of auto-cure resin-based composites was the inability of the clinician to control the setting time. Once mixed, the material would irreversibly begin to polymerize whether the clinician was totally prepared or not.

The ability to initiate the polymerization reaction when most desired has been achieved by the introduction of photochemically initiated composites. The first photochemically initiated composite resins used in dentistry required ultraviolet (UV) radiation (7). The UV radiation source employed a high pressure mercury arc and provided a simple mechanical shutter to mask off the radiation when not in use. The UV radiation was emitted through a light guide to expose the composite resin. The effective wavelength was between 364 and to 367 nm. An organic compound that generates free radicals when exposed to 365-nm wavelength electromagnetic radiation, such as, benzoin alkyl ether, was added to composite resins in place of the thermochemical initiator (benzoyl peroxide). Visible light initiated composite resins depend on a diketone (e.g., camphoroquinone) and an organic amine (e.g., *N,N*-dimethylaminoethylmethacrylate) to produce free radicals that result with polymerization initiation. The diketone absorbs electromagnetic radiation in the 420–450 nm wavelength range. The unit that provides the electromagnetic radiation (dental curing light unit) usually consists of a light source, a filter that selects the range of transmitted wavelength and a light tube that directs the light beam to the composite. The units generally emit wavelength in the 400–550 nm range. There is some concern over the potential of damage to the eyes of dental operators (8). Indeed, health concerns



TEGDMA Triethyleneglycol dimethacrylate

Figure 2. TEGDMA.

over the use of UV radiation for polymerization initiation initially encouraged the investigation into using “visible” light instead.

Dispersed Phase

Inorganic filler particles were added to dental resins as early as 1951. It was a number of years, however, before composite materials were generally utilized by dentists. The research of Bowen (4) improved the mechanical properties of filled resins. Bowen treated silica powder with 1.0% aqueous solution of tris(2-methoxyethoxy) vinyl silane to which sodium hydroxide was added and the resultant slurry dried at 125°C. Peroxide was added in an acetone solution and the solvent evaporated. This treatment resulted with an organophilic silica powder. The first dental composites utilized fillers, such as E-glass fibers, synthetic calcium phosphate, fused silica, soda-lime glass beads, and other glass formulations (9). The commonly used filler particles currently used as reinforcing materials are quartz, colloidal silica, lithium aluminum silicate, and silica glasses containing either strontium or barium (10). Quartz was the most successful of the commercial fillers due to its index of refraction and inert nature, however, its hardness and large particle size made polishing difficult resulting with the transition to softer glass fillers (11). In addition to being radiopaque, the softer glass fillers facilitate easier polishing of the set composite resin and the production of fine filler particles for incorporation into the matrix. There is, however, a potential toxicity problem with barium glasses and the radiopaque glasses may be susceptible to degradation in water (12). Chemical processes may also be used to synthesize filler particles, such as the colloidal silicon dioxide pyrogenic particles (13). The colloidal silicon dioxide particles may be fabricated by burning silicon tetrachloride in the presence of an hydrogen and oxygen gas mixture. Particles ranging from 0.007 to 0.14 μm result. The small size of the colloidal silicon dioxide filler particle, when incorporated in the dispersed phase, allows the polishing of a resin-based composite to a smooth finish. However, the small filler particle size prevents high filler loading. This tends to decrease the filler fraction of resin-based composites manufactured with a colloidal silicon dioxide dispersed phase (Table 1).

Hydroxyapatite and amorphous calcium phosphate filler particles have been advocated for resin-based composites (15,16). The reported advantage of these fillers is the potential for remineralizing adjacent tooth structure.

Filler-Matrix Interface

Transfer of stress from the dispersed-phase filler particles through the ductile matrix phase should occur to improve mechanical properties. The bond between the dispersed filler phase and the organic matrix may be achieved by two different methods, chemically or mechanically. The mechanical retention of the dispersed phase is achieved by sintering particles or fibers together, or by etching away a phase of a glass filler particle leaving a porous surface. Monomer may then flow into the porous surface creating a mechanical interlocking. Chemical bonding to the dispersed

phase may be achieved by using coupling agents, such as an epoxy silane (17). Two of the silane agents are γ -glycidoxypropyltrimethoxysilane and γ -methacryloxypropyltrimethoxysilane.

Silane coupling may involve hydrolysis of the methoxy groups with bound surface water on the dispersed phase-filler particle or with aluminol or silanol groups of the dispersed phase filler particle. During polymerization of the composite resin, the unsaturated carbon double bonds are available to react with the matrix. *In vitro* tests suggest, that a general rule, as filler volume is increased, wear is reduced regardless of filler treatment (18).

PHYSICAL AND MECHANICAL PROPERTIES

Physical Properties

Wettability partially determines the marginal microleakage and surface staining of a composite. Wettability may be determined by the contact angle formed by a drop of water on the resin-based composite surface. Composite resins are considered hydrophilic because the advancing contact angle of water on composite surface is $\sim 65^\circ$. The water sorption value is 0.6 mg/cm^2 and water uptake of composites is a function of polar groups in the polymer structure. The thermal conductivity of composites closely matches that of dentin and enamel. The coefficient of thermal expansion usually averages $26\text{--}40 \times 10^{-6} \text{cm}/\text{cm per}^\circ\text{C}$ for the range of $0\text{--}60^\circ\text{C}$, and recently marketed composite resins even more closely match the values normally obtained for tooth structure. A composite resin with a thermal coefficient of expansion similar to tooth structure ($10 \times 10^{-6} \text{cm}/\text{cm per}^\circ\text{C}$) would theoretically suffer from less margin microleakage. When there is a difference in the values, composite resin restorations may expand or contract more than tooth structure during temperature changes resulting with gap formation between the two substances. Margin microleakage may then occur. Polymerization contraction (% by volume) for the typical composite resin averages 3.2–3.8%. The amount of contraction due to polymerization is effected by the types of oligimers used in the matrix phase and the filler volume. As the composite resin shrinks, it pulls away from the walls of the cavity preparation in the tooth. This polymerization contraction encourages the ingress of salivary contaminants and bacteria (microleakage). Not all studies have shown a statistical correlation between polymerization shrinkage and microleakage (19).

Methods have been employed to reduce the stresses associated with polymerization shrinkage. Pulse or two-step light activation for polymerization initiation has been recommended (20,21). Other methods include incremental filling, directed shrinkage, void incorporation, filler and matrix alteration.

Mechanical Properties

Virtually all composite resin manufacturer's spend considerable resources on testing their own and competitor's products for mechanical properties. Dentists are continually

Table 1. List of Materials and Percentage of Fillers by Weight Determined by Ashing in Air^a

Material	Classification ^b	Manufacturer	Batch and Shade	% Fillers by Weight
Aeliteflo	VLC Hyb Flow CS	Bisco, Inc Itasca, IL, USA	039317 (A3)	54.9
Amelogen	VLC Hyb Univ CS	Ultradent products, UT, USA	2CPM (A2)	72.9
Arabesk	VLC Hyb Univ CS	Voco, Cuxhaven, Germany	70500 (A3)	71.6
Arabesk-Flow	VLC Hyb Flow CS	Voco, Cuxhaven, Germany	82777 (A3)	61.8
Arabesk-Top	VLC Hyb Univ CS	Voco, Cuxhaven, Germany	81594 (A3)	71.5
Ariston-pHc	VLC 'Smart Material'	Vivadent, Schaan, Liechtenstein	A00001 (-)	74.8
Brilliant-Dentin	VLC Hyb Univ CS	Colténe, Whaledent, Switzerland	GE931 (A3)	75.6
Brilliant-Enamel	VLC Hyb Univ CS	Colténe, Whaledent, Switzerland	GE902 (A3)	75.4
Charisma-F	VLC Hyb Univ CS	Heraeus Kulzer, Wehrheim, Germany	23 (A20)	76.4
Charisma-PPF	CC Hyb Univ CS	Heraeus Kulzer, Wehrheim, Germany	2 (A10)	68.3
Clearfil Photo Post	VLC Hyb Univ CS	Kuraray, Osaka, Japan	0035A (UL)	84.7
Clearfil Photo Ant.	VLC Hyb Univ CS	Kuraray, Osaka, Japan	0024C (A3)	59.9
Colténe-SE	VLC Hyb Univ CS	Colténe, Whaledent, Switzerland	FBJO1 (A3)	71.3
Concise	CC Conventional CS	3M, St. Paul, MN, USA	19970303(U)	80.2
Dyract-Flow	VLC Flow CM	Dentsply De Trey, Konstanz, Germany	9809000103 (A2)	55.4
Elan	VLC CM	Sybron/Kerr, Orange, USA	805872 (A3,5)	71.2
EXI-119 (Z-250) ^c	VLC Hyb Univ CS	3M, St. Paul, MN, USA	030998 (A3.5)	77.4
EXI-120 (P-60) ^c	VLC Hyb Pack CS	3M, St. Paul, MN, USA	030998 (A3,5)	78.9
F-2000	VLC CM	3M, St. Paul, MN, USA	19970905 (A3)	80.5
Glacier	VLC Hyb Univ CS	Southern Dental Industries, Australia	60506 (B3)	78.2
Metafil-CX	VLC Microfine CS	Sun Medical, Shiga, Japan	71201 (A3,5)	41.7
Pertac-II	VLC Hyb Univ CS	Espe, Seefeld, Germany	00634764 (A3)	70.0
Polofil-Molar	VLC Hyb Univ CS	Voco, Cuxhaven, Germany	63596(U)	78.5
Quadrant Anterior	VLC Microfine CS	Cavex Haarlem, Holland	22C (A2)	58.6
Quadrant Posterior	VLC Hyb Pack CS	Cavex Haarlem, Holland	30C (A2)	65.2
Revolution	VLC Hyb Flow CS	Sybron/Kerr, Orange, USA	710669 (A3)	53.9
Silux-Plus	VLC Microfine CS	3M, St. Paul, MN, USA	6DH (U)	54.8
Solitaire	VLC Hyb Pack CS	Heraeus Kulzer, Wehrheim, Germany	26 (A30)	64.3
Spectrum	VLC Hyb Univ CS	Dentsply De Trey, Konstanz, Germany	9608244 (A3)	75.3
Surefil	VLC Hyb Pack CS	Dentsply De Trey, Konstanz, Germany	980818 (A2)	79.4
Tetic-Ceram	VLC Hyb Univ CS	Vivadent, Schaan, Liechtenstein	900513 (A3)	75.7
Tetric-Flow	VLC Hyb Flow CS	Vivadent, Schaan, Liechtenstein	901232 (A3)	64.0
Wave	VLC Hyb Flow CS	Southern Dental Industries, Australia	80608 (A3)	60.7
Z-100	VLC Hyb Univ CS	3M, St. Paul, MN, USA	19960229 (UD)	79.6

^aReprinted with permission from Ref. 14.

^bVisible light cured = VLC, chemically cured = CC, composite = CS, compomer = CM, flowable = Flow, packable = Pack, universal = Univ, hybrid = Hyb.

^cThe P-60 (Posterior composite) and the Z-250 (Antero-Posterior composite), marketed by 3M were included in this study as experimental composites EXI-120 and EXI-119.

bombarded with advertisements touting a composite resin's "compressive strength". It cannot be denied that mechanical properties are important for consideration. On the other hand, it is quite evident that clinical success may not be solely correlated with laboratory results or easily predictable. The oral environment is rather hostile to inorganic or foreign materials. Corrosion or degradation occurs with most metals and resins used by dentists. Composite resins demonstrate a change in roughness of the surface with laboratory aging suggesting degradation (22). Mechanical properties of the composite resins have improved considerably the last decade. Unfortunately, the exact relationship between mechanical properties and clinical success has yet to be determined.

The compressive strength of composite resins is greatly superior to their tensile strength. The clinical significance of this is unclear. It may be assumed that tensile strength is of major concern when placing composite resins in posterior teeth as a sliding tooth cusp will cause shear stresses (23). Many manufacturers claim that the compressive strength of their composite resins equal silver amalgam (silver filling material). However, the clinical relevance of

this high compressive strength is not known due to a lack of data demonstrating the minimum strength required to resist the forces of occlusion and mastication. The modulus of elasticity is an important factor to consider when a composite resin is used on a stress bearing area. A composite resin with a very low modulus will deform under occlusal stress and increase microleakage or transfer the stress to the supporting tooth. This does not mean that low modulus composites cannot be used when well supported by adjacent tooth structure (14).

Composite resins may be classified by the component materials that comprise each of the three phases. The matrix phase may be described as being either hydrophilic or, as in the case of some experimental resins, hydrophobic. The interfacial phase is not as amenable as the matrix phase to classification. It is possible to describe the coupling mechanism as being chemical (either polymeric or silane treated) or mechanical. However, the most appropriate basis for classification would appear to be the size and chemical composition of the dispersed phase (24). Composite resins may be classified into five main categories: traditional hybrid, microfill (pyrogenic silica),

microhybrid, packable, and flowable (25). Hybrid resin composites have an average particle size of 1–3 μ and are 70–77% filled by volume. They present with good physical properties compared to the microfilled resin composites. Some composite resins used softer, radiopaque glass fillers averaging 1–5 μ m (26). Microhybrid resin composites have an average particle size of 0.4–0.8 μ and are 56–66% filled by volume. Microhybrid resin-based composites generally have good physical properties, improved wear resistance, and relatively high polishability. Microfill composite resins utilize pyrogenic silicon dioxide with an average particle size of 0.04–0.1 μ , but due to the increase in viscosity associated with an unacceptably low filler fraction (35–50% by volume), a modified method of manufacture was developed. Microfiller-based complexes consists of prepolymerized particles (pyrogenic silicon dioxide mixed with resin matrix and cured) and resin matrix with an equal concentration of dispersed microfiller (pyrogenic silicon dioxide) as was in the prepolymerized particles. This formulation allows for an increased filler fraction without the difficulty in manipulation due to high viscosity. The microfiller composite resins exhibit high polishability (27). The smoother the surface of the composite resin, the less that surface wear occurs and plaque accumulates (28,29). Packable composites have an average particle size of 0.7–20 μ and are 48–65% filled by volume. Their higher viscosity is considered desirable in establishing interproximal contacts and is the result of a higher percentage of irregular, fibrous, or porous filler in the resin matrix. Flowable composites have an average filler particle size of 0.04–1 μ and are 44–54% filled by volume. Their reduced filler volume decreases viscosity to ease placement in thin increments or small cavity preparations. They are also more flexible, with a decreased modulus of elasticity, to absorb stress in certain applications (Table 2).

CLINICAL APPLICATION

Bonding is a general term that is used to describe the joining, uniting, or attaching of adhesives to an adherend (6). Bonding describes the attachment of two materials, but not the mechanism by which the bonding occurred. A bonded assembly may be attached together by mechanical means or by physical and chemical forces. The bond that

occurs between composite resin and tooth enamel is mechanical and is achieved by the process of acid etching. The acid-etch technique was developed by Buonocore in 1955 for use in dentistry (31). An 85% phosphoric acid solution was applied to the teeth of volunteer subjects and greatly enhanced the bond strength of an unfilled acrylic resin to the enamel surface. Retief demonstrated the effectiveness of a 50% phosphoric acid pretreatment of enamel for bonding and also the reversibility of the process by the saliva's remineralization of the enamel (32–35). Bonding is now universally accepted by the dental profession.

Enamel bonding ushered in the age of cosmetic dentistry and popularized the use of composite resins. It was also desirable to develop materials that would adhere to the second layer of tooth structure, the dentin. This is especially applicable in situations where the available enamel for bonding is minimal. Commercially available dentin bonding systems have recently been developed. Bonding may occur by two different means, micromechanically or chemically. Adhesion may be chemically established to either the inorganic or organic portions of dentin. The dentin bonding systems have improved considerably in a short period of time showing great promise for reducing margin microleakage and increasing restoration retention. Composite resins have rapidly established themselves as an important segment of a dentist's restorative armamentarium. Composite resins were initially used solely as a replacement for the unfilled acrylic resins and silicate cements. Use was limited to the incisors where esthetics were of paramount importance. Even then, many dentists considered all resins to be strictly temporary at best.

The patient was often advised to consider a permanent restoration that was metallic, such as a gold foil (placed by compaction of overlapping gold increments directly into the prepared cavity of a tooth). The failure rate of the early resins was considered unacceptable by many dentists and composite resins had to prove themselves "worthy".

The introduction of "bonding" reduced the severity and occurrence of microleakage with the composite resin restorations. Bonding also provided a conservative means to retain a composite resin restoration without substantial sacrifice of tooth structure (by cutting undercut areas into the tooth). Composite resins soon became useful in restoring traumatically fractured teeth (34–38). Malformed or hypoplastic teeth were reconstructed using composite resins and bonding (39,40). The pigments incorporated

Table 2. Classification and Physical Properties of Resin-Based Composites^a

Composite Type	Average Particle Size (Micrometers)	Filler Percentage (Vol %) ^b	Physical Properties ^b		
			Wear Resistance	Fracture Toughness	Polishability
Microfill	0.04–0.1	35–50	E	F	E
Hybrid	1–3	70–77	F↔C ^c	E	G
Microhybrid	0.4–0.8	56–66	E	E	G
Packable	0.7–20	48–65	P↔G ^c	P↔E ^d	P
Flowable	0.04–1	44–54	P	P	F↔G ^d

^aReprinted with permission from Ref. 30.

^bSources: Kugel,⁴⁷ Wakefield and Kofford⁵⁰ and Leinfelder and colleagues⁵³.

^cE: Excellent; G: good; F: fair; P: poor.

^dVarying among the same type of resin-based composite.

into the composite resin matrix provide a wide range of natural shades. Additional tinted and opaquing resins are available that enable the dentist to achieve a life-like result closely mimicking enamel. The bonding of composite resins as a thin veneer to enamel was advocated for an aesthetic technique of restoring discolored teeth. Young discolored or malformed teeth could be given a natural appearance by covering the enamel surface with a thin "veneer" of the appropriately shaded composite resin (41,42). The public eventually became aware of dentistry's newest advancement and "cosmetic bonding" was "born".

Preventive dentistry benefited by the development and introduction of pit and fissure sealants. A pit and fissure sealant material is a BIS-GMA based resin that is introduced into the caries susceptible pits and fissures of teeth forming a barrier against the action of decay producing bacteria. Pit and fissure sealant resin is retained or bonded to the enamel surface of the teeth by the acid-etch technique. In the 1980s, a conservative cavity preparation that utilizes composite resins and bonding had been proposed by Simonsen (43). This technique involves the removal of only decayed tooth structure with the composite resin restoration bonded to the enamel surface sealing the pits and fissures from future decay. Simonsen termed this technique the "Preventive Resin Restoration" (43). Composite resins are also utilized as cementing or as luting agents.

Acrylic preformed laminate veneers were bonded with composite resins to discolored teeth as aesthetic restorations (44). Porcelain veneers have been bonded to discolored teeth with composite resins since 1983 (45). Composite resins have also been used to bond orthodontic brackets, and fixed partial dentures (Maryland Bridges) to teeth (46). Splints to stabilize periodontally or traumatically loosened teeth are constructed of composite resin. Restorations of decayed posterior teeth (molars) are more and more done with composite resins and bonding (47). A number of dentists and patients are reportedly concerned over the potential toxic effect of mercury, one of the constituents of silver amalgam filling material (48). Studies have shown that dentists with the highest exposure to silver amalgam also have the highest urinary mercury levels (49). The concern about mercury toxicity may yet be the necessary impetus to finally develop an ideal composite resin (50).

New low shrink composite resins consist of a bis-GMA matrix with a strontium glass filler of an average size of 1.1 μ . The composite is 86.5% filled by weight and has a compressive strength of 267.5 MPa. The composite has a volumetric shrinkage of 1.4%, which reduces the shrinkage stresses induced by polymerization. Also recently introduced have been the no-rinse conditioners (self-etching primers). No rinse conditioners are very convenient for the dentist and are very useful in patients with severe "gag" reflex (51).

The self-etching primer consist of a penetrating, polymerizable monomer and an acidic component. The pH of the self-etching primer is between 0.6 and 2, which is acidic enough to etch cut or prepared enamel (52). Several self-etching primers currently available will not etch unprepared enamel leaving resin extensions improperly bonded

(53). Self-etching primers that leave an acidic intermixed zone will demonstrate osmotic blistering (54). Osmotic blistering is the separation of the intermixed zone due to water penetrating the hydrophilic primers. Water is pulled from the dentin because of the high concentration of ionic species retained in the primer when no-rinse conditioners are used. The osmotic gradient from the ionic species and acid monomers creates pockets or blisters of water between the adhesive and composite resin. In addition, both the filler and the resin matrix may suffer degradation from the free acid radicals if the intermixed zone is left unbuffered. The shear bond strength of resin-based composite or sealants is reportedly less with the self-etching primers than the "total etch" technique, a pretreatment of both enamel and dentin with 32–40% phosphoric acid solution or gel (55). Immobilization of an antibacterial component into the resin matrix has also been achieved creating antibacterial composite resins. A new monomer, MDPB, has been produced by combining a quaternary ammonium with a methacryloyl group and incorporating it into the resin matrix for copolymerization with other monomers. The antibacterial effect is on contact only and does not dissipate by dissolution into the saliva (56).

BIBLIOGRAPHY

1. Anusavice K., editors. Phillip's Science of Dental Materials. 11th ed. Philadelphia: W.B. Saunders; 2003.
2. Paffenbarger G, Nelson R, Sweeney W. Direct and indirect filling resins: A review of some physical and chemical properties. *J Am Dent Assoc* 1953;47:516–521.
3. Stanford J. The current status of restorative resins. *Dent Clin N Am* 1971;15:57–66.
4. Bowen R. Dental filling Material comprising vinyl silane treated fused silica and a binder consisting of a reaction product of bisphenol and glycidyl acrylate. US Patent 3, 066, 112. Nov. 1962.
5. Ruyter I, Sjouik I. Monomer Composition of dental composites and sealants. *J Dent Res (Spec Iss A)* Jan. 1978; 57:249, Abstr. 700,
6. Craig R. Chemistry, composition, and properties of composite resins. In: Horn H, editor. *Dental Clinics of North America*. Philadelphia: W. B. Saunders; 1981. p 223.
7. Rock W. The use of ultra-violet radiation in dentistry. *Br Dent J* 1974;136:455–458.
8. Fan P, et al. Visible light-cured composites and activating units. *J Amer Dent Assoc* 1985;110:100–103.
9. Bowen R. Compatibility of various materials with oral tissues, I: The components in composite restorations. *J Dent Res* 1979;58:1493–1503.
10. Zinck J, Norling B, Buchanan R. Composite resins systems: A comparison. *Dent Stud* 1982;61:51–55.
11. Bowen R, Cleek G. A new series of X-ray opaque reinforcing fillers for composite materials. *J Dent Res* 1972;51:177–82.
12. McKinney J, Wu W. Chemical softening and wear of dental composites. *J Dent Res* 1985;64(11):1326–331.
13. Jorgensen K, Asmussen E. Occlusal abrasion of a composite restorative resin with ultra-fine filler—an initial study. *Quintessence Int* 1978;6:73.
14. Sabbagh J, Vreven J, Leloup G. Dynamic and static moduli of elasticity of resin based materials. *Dent Mater* 2002;18:64–71.
15. Arcis R, et al. Mechanical properties of visible light-cured resins reinforced with hydroxyapatite for dental restoration. *Dent Mater* 2002;18:49–57.

16. Skrtic D, Antonucci J, Eanes E. Improved properties of amorphous phosphate fillers in remineralizing resin composites. *Dent Mater* 1996;12:295–301.
17. Serman S, Marsden J. Silane coupling agents as integral blends in resin-filler systems. *Mod Plastics* 1963;49(11):125.
18. Lim B, Ferrocane J, Condon J, Adey J. Effect of filler fraction and filler surface treatment on wear of micro-filled composites. *Dent Mater* 2002;19:1–11.
19. Rosin M, et al. Polymerization shrinkage-strain and microleakage in dentin-bordered cavities of chemically and light cured restorative materials. *Dent Mater* 2002;18:521–528.
20. Yoshikawa T, Burrow M, Tagami J. A light curing method for improving marginal sealing and cavity wall adaptation of resin composite restorations. *Dent Mater* 2001;17:359–366.
21. Lim B, et al. Reduction of polymerization contraction stress for dental composites by two-step light-activation. *Dent Mater* 2002;18:436–444.
22. Powers J, Fan P. Surface degradation of composite resins. *J Dent Res (Spec. Iss. A)* 1979;58:328.
23. Soderholm K. Filler systems and resin interface. In: Vanherle G, Smith D, editors. *Posterior Composite Resin Dental Restorative Materials*. The Netherlands: Peter Szule Publishing; 1985. p 149.
24. Lutz F, Phillips R. A classification and evaluation of composite resin systems. *J Pros Dent* 1983;50:480–488.
25. Deliperi S, Bardwell D. An alternative method to reduce polymerization shrinkage in direct posterior composite restorations. *JADA* 2002;133:1387–1398. October.
26. Dogon I, Cross M, Douglas W. Clinical and laboratory studies on a fine grind composite. *J Dent Res* 1982;61:214 (Abstr. No. 320).
27. Christensen R, Christensen G. In vivo comparison of a micro-filled and a composite resin: A three year report. *J Pros Dent* 1982;48(6):657–663.
28. Shampanier A. A comparative Study of the surface resistance of various composite filling materials to toothbrushing abrasion. Masters dissertation. Chicago, (IL):Northwestern University Dental School; 1978.
29. Sotrez S, Van Huysen G, Gilmore N. A histologic study of gingival tissue response to amalgams, silicate, and resin restorations. *J Perio* 1969;40:543–546.
30. An alternative to reduce polymerization shrinkage in direct posterior composite restorations. *J Am Dent Assoc* 2002; 133.
31. Buonocore M. A simple method of increasing adhesion of acrylic filling materials to enamel surface. *J Dent Res* 1955;34:849–853.
32. Retief D. Effect of conditioning the enamel surface with phosphoric acid. *J Dent Res* 1973;52:333–341.
33. Retief D. A comparative study of three etching solutions. *J Oral Rehabil* 1974;1:381–390.
34. Jordan R, Suzuki M, Charles D, Gwinnett A. Esthetic and conservative restoration of the fractured incisor by means of microfilled composite materials. *A O* 1981;74:51–59.
35. Black J, Retief D, Lemons J. Effect of cavity design on retention of Class IV composite resins restorations. *J Am Dent Assoc* 1981;103:42–46.
36. Roberts M, Moffa J. Restoration of fractured incisal angles with an ultraviolet activated sealant and a composite resin: a case report. *J Dent Child* 1972;39:364–365.
37. Buonocore M, Davila J. Restoration of fractured anterior teeth with ultraviolet—light polymerized bonding materials: a new technique. *J Am Dent Assoc* 1973;86(6):1349–1354.
38. Hill F, Soetopo A simplified acid-etch technique for the restoration of fractured incisors. *J Dent* 1977;5:207–212.
39. Jordan R, Suzuki M, Gwinnett A. Restoration of fractured and hypoplastic incisors by the acid etch resin technique: A three year report. *J Am Dent Assoc* 1977;95:795–803.
40. Black J. Esthetic restoration of tetracycline-stained teeth. *J Am Dent Assoc* 1982;104:846.
41. Cooley R. Laminate preformed resin veneer. Proceedings of the American Dental Association Meeting; Anaheim, CA, Oct. 1978.
42. Faunce F, Myers D. Laminate veneer restoration of permanent incisors. *J Am Dent Assoc* 1976;93:790–792.
43. Simonsen R. Prevention resin restoration: Three year results. *J Am Dent Assoc* 1980;100:535–539.
44. Cooley R. Status report on enamel bonding of composite, preformed laminate, and laboratory fabricated resin veneers. *J Am Dent Assoc* 1984;109:762–764.
45. Horn L. A new lamination: Porcelain bonded to enamel. *Clin North Am* 1983;27:671–684.
46. Williams V, Dehehy G, Thayer K, Boyer D. Acid-etch retained cast metal prostheses: A seven year retrospective study. *J Am Dent Assoc* 1984;108:629–631.
47. Ernst C, Buhtz C, Rissing C, Willershausen B. Clinical performance of resin composite after 2 years. *Compendium* 2002;23(8):711–724.
48. Abraham J. The effect of dental amalgam restorations of blood mercury levels. *J Dent Res* 1984;63(1):71–73.
49. Naleway C, Sakaguchi R, Mitcheell E, Muller T, Ayer W, Hefferren J. Urinary mercury levels in U.S. dentists, 1976–1983: Review of Health Assessment Program. *J Am Dent Assoc* 1985;111:37–42.
50. Lobner D, Asrari M. Neurotoxicity of dental amalgam is mediated by zinc. *J Dent Res* 2003;82(3):243–246.
51. Cannon M. Advances in pediatric esthetic dentistry. *Compendium* 2003;24(8):34–39.
52. Wang Y, Sharp L, Suh B. The morphology study of several self etching adhesive systems. *J Dent Res* 2002; Abstr. 1898: (Spec. Iss. A).
53. Tay F, et al. Single step adhesives are permeable membranes. *J Dent Sep–Nov* 2002;30(7–8):371–382.
54. Tay F, et al. Osmotic blistering in enamel bonded with one-step self-etch adhesives. *J Dent Res* 2004;83(4):290–295.
55. Fuks A, Eidelman E, Lewinstein I. Shear strength of sealants placed with non-rinse conditioning compared to a conventional acid etch rinse technique. *J Dent Child Sept–Dec* 2002; 239–242.
56. Imazato S. Antibacterial properties of resin composites and dentin bonding systems. *Dent Mater* 2003;19:449–457.

See also BIOMATERIALS FOR DENTISTRY; BIOMATERIALS, TESTING AND STRUCTURAL PROPERTIES OF; ULTRAVIOLET RADIATION IN MEDICINE.

RESPIRATOR. See VENTILATORS, ACUTE MEDICAL CARE.

RESPIRATORY MECHANICS AND GAS EXCHANGE

JAMES LIGAS
University of Connecticut
Farmington, Connecticut

INTRODUCTION

Respiratory mechanics applies the principles of solid and fluid mechanics to pressure, flow, and volume measurements obtained from the respiratory system. The utility of the resulting mathematical models depends on how well they guide clinical decisions and prove consistent with the results of new experiments. Although very sophisticated

models of respiratory system structure and function exist, our focus here is on simple concepts used for pulmonary function testing and mechanical ventilation.

RESPIRATORY SYSTEM STRUCTURE

The airways, or bronchial tree, conduct air to the small air sacs, or alveoli, which comprise the lung parenchyma. The upper, larger airways are lined with ciliated epithelium and mucus secreting cells to warm, humidify, and filter small particles from the inhaled gas. The heated and humidified air is ultimately exhaled, resulting in a water loss from the body of $\sim 1 \text{ L} \cdot \text{day}^{-1}$. When the upper airways are bypassed by tracheostomy or by intubation for mechanical ventilation, drying of the respiratory tract can result in inspissation of secretions and obstruction of the airways. Extracorporeal humidification of the compressed, zero-humidity gases used in mechanical ventilation is a necessity (1).

The cross-sectional area of the trachea is roughly the size of a United States quarter dollar (0.25¢). Alveolar surface area is about the size of a tennis court, so that over a very short distance the airways branch repeatedly to bring inhaled gases into contact with a large surface area for diffusion into the blood. The branching is generally dichotomous and asymmetric. The number of branches from the trachea to an alveolus varies from 6 to 30. Small nerve fibers coursing through the walls of airways cause glandular secretion and determine muscle tone. The blood vessels bringing venous blood to the alveoli for gas exchange follow a similar branching structure so that air and blood flow are closely matched for efficient exchange of gases (2).

The last generation of bronchioles ends in sprays of alveolar ducts and sacs. There are some 300 million alveoli, each $\sim 300 \mu\text{m}$ in diameter at full inflation, forming a network of interconnecting membranes. Within those membranes pass the smallest vessels, the alveolar capillaries. Microscopic studies show that the membrane consists of two parallel tissue sheets separated by a series of tissue posts, much like the deck of a parking garage. Red blood cells spend $<1 \text{ s}$ in this structure, and diffusion of respired gases occurs across the $0.3\text{--}3.0 \mu\text{m}$ barrier between air space and red cells in the capillaries. At low intravascular pressures, some parts of this network remain collapsed, yet can be recruited if the pressure in the pulmonary vessels increases. In addition, lymphatics also drain the interstitial space and follow the structure of the bronchial tree. They are capable of removing large volumes of extravasated fluid if necessary.

A thin membrane, the pleura, covers the outer surface of the lungs and is reflected to line the inside surface of the thoracic cavity. Between these two pleural surfaces is a space $6\text{--}30 \mu\text{m}$ in thickness. A very small amount of fluid is normally present, providing lubrication that allows the lungs to slide freely over the interior of the chest wall (3).

The thoracic cage itself consists of the spine, sternum, ribs, and associated muscles of the chest wall. The diaphragm separates the abdominal cavity from the thoracic cavity and is the major muscle effecting resting ventilation (4). As muscular contraction expands the thoracic cavity,

the lungs follow, filling with air. When the subject is at rest exhalation is passive. The muscles relax, the thoracic cavity decreases in volume, and air flows out of the lung.

In addition to the solid mechanics of these structures, analyses of fluid flows are important. The behavior of fluids spans many regimes: convection-dominated flows in the large airways, gaseous diffusion at the alveolar level, interphase diffusion through the capillary walls into the blood, and viscous flows in the alveolar ducts, capillary spaces, and intrapleural space.

STATIC MECHANICS OF THE RESPIRATORY SYSTEM

The simplest test of respiratory system behavior is to ask a subject to inhale or exhale to different lung volumes (Fig. 1). The tidal volume, V_T , is the volume of air that moves in and out of the lungs in a normal breath. At the end of a normal exhalation all the muscles of respiration are relaxed. With the glottis open, no forces are exerted on the respiratory system and atmospheric pressure exists both in the alveolar spaces and on the body surface. The lung volume under those conditions is called the functional residual capacity, or FRC. One can exhale beyond FRC by forcing more air out of the lung. The amount that can be forced out is called the expiratory reserve volume (ERV). Even at that point, there is still residual air in the lungs: the residual volume (RV). If one inhales maximally, the volume of air in the lung is the total lung capacity, or TLC. Similar to ERV, there is an inspiratory reserve available to us beyond our usual tidal volume: the inspiratory reserve volume, or IRV. The total amount of air that could be inhaled from FRC is called the inspiratory capacity, and the total amount that could be exhaled from TLC is called the vital capacity (VC). Figure 1 shows that quantities labeled "capacity" are the sum of two or more quantities labeled "volume". Together with statistical tables relating normal values to gender, age, and size, this data provides one measure of any muscular weakness or structural impairment that might limit the ability to move air into or out of the respiratory system.

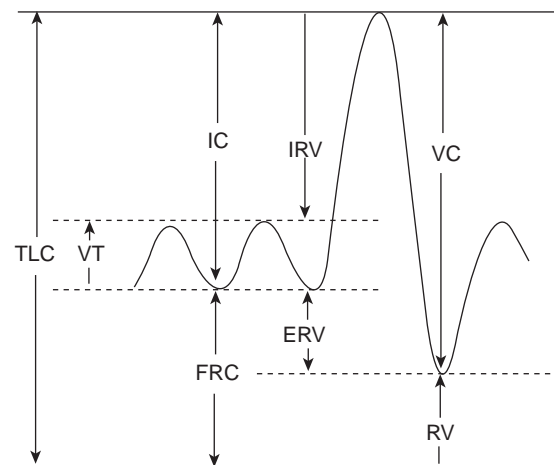


Figure 1. Lung volumes and capacities.

Beyond simply measuring these volumes, one can ask how much force is required to change the volume by a certain amount. These experiments have been performed upon excised lungs and intact animals and subjects.

Parenchyma

Experiments that applied static pressure to an excised, air-filled lung supported in air showed that the relationship between lung volume and the applied transpulmonary pressure difference, that is, airway pressure minus pressure at the pleural surface, is dependent on the volume history of recent expansion (Fig. 2). To achieve any given volume, the required pressure difference is greater for inflation than for deflation. A large part of this behavior is due to surface tension at the gas-liquid interface in the alveoli and smallest airways (5,6). When the lung is degassed and then filled with and surrounded by saline, inflation pressures are reduced and much of the hysteresis is eliminated (Fig. 2). The surface-active agent, or surfactant, responsible for most of the hysteresis consists of a phospholipid protein complex secreted by certain cells found in the alveolar epithelium. Surface tension-area data from experimental systems utilizing films of surfactant indicate that surface tension varies over the range of $2\text{--}50 \text{ dyn} \cdot \text{cm}^{-1}$, decreasing markedly as the surface area of the film decreases, and is independent of cycling frequency. Surfactant lowers the surface tension below that for a pure water-air interface, and therefore decreases the pressure necessary to inflate the lungs and keeps alveoli from collapsing during exhalation. Infants born prematurely can lack surfactant, making respiration difficult. The development of an artificial surface-active agent delivered by aerosol was an important advance in neonatal care (7).

Because one often deals with changes about a specified lung volume, the elastic behavior of the lung is routinely described in terms of the tangent to the pressure-volume curve at that point, referred to as the *local compliance* (C_L) of the lung,

$$C_L = dV/dP \quad (1)$$

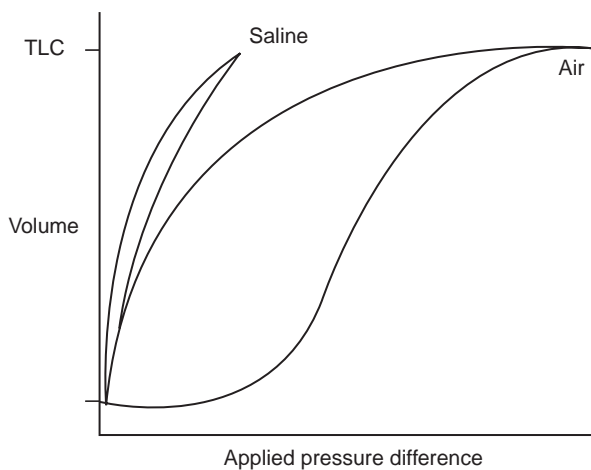


Figure 2. Pressure-volume curves for excised lungs filled with air and saline.

where P is the transpulmonary pressure (alveolar minus pleural surface pressure). Because of the nonlinear nature of the pressure-volume relationship, the compliance varies with lung volume. Thus the simplest model of lung elasticity is a single parameter derived from the transpulmonary pressure-volume curve of the liquid- or air-filled lung. The compliance for an air-filled lung will reflect both the mechanical properties of the tissue and the effects of surfactant.

Intrapleural Space

For spontaneously breathing subjects, the stresses generated by muscle contraction cause thoracic cavity expansion and are transmitted to the lungs through the intrapleural space. The thin fluid layer present between the membrane lining the lung (the visceral pleura) and the membrane lining the inside of the thoracic cavity (the parietal pleura) must in some way transmit those forces. Early models of this coupling postulated the concept of an intrapleural pressure (8). Because the minimal volume attainable at zero distending pressure for an excised lung (V_0) is below the residual volume of the intact subject, and because functional residual capacity is less than the volume of the thoracic cavity when the lungs are removed, the simplest model was that there must be a negative pressure, that is, a pressure less than atmospheric pressure in the pleural space (Fig. 3). In mechanical terms, any shear stresses in the minute amount of pleural fluid were neglected and the mechanics were modeled as a normal stress, the intrapleural pressure (P_{pl}).

Measurements of intrapleural pressure in intact animals or human subjects were traditionally made by placing

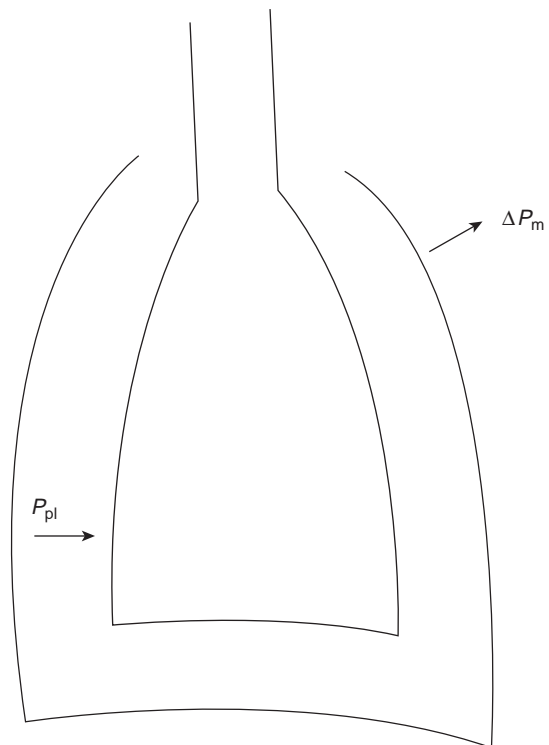


Figure 3. Lung-chest wall model. See text for details.

a catheter tipped with a small latex balloon into the thoracic esophagus. The balloon was partially inflated, but not stretched. The theory is that both the balloon and the esophageal wall are flaccid structures, and so pressure changes measured by this manometry system would reflect changes in intrathoracic pressure (9). These measurements were undertaken in intact animals and subjects to try to separate the mechanics of the chest wall from the properties of the lung itself.

Thoracic Cage

In a manner similar to the parenchyma, the chest wall was conceived of as an elastic structure and the slope of the pressure–volume curve for the relaxed chest wall was called the chest wall compliance. The active forces exerted by the muscles (ΔP_m) are conceived of as a normal force applied to change the intrapleural pressure. During inhalation, expansive chest wall forces decrease the intrapleural pressure, causing expansion of the lung. If the subject forcefully exhaled, or used the muscles to exhale to a volume below functional residual capacity, the intrapleural pressure would increase.

These were the simplest models for the static mechanics of the respiratory system. Pressure differences at various points in the system are related to volume changes, and the tissue properties were modeled as a simple elasticity although the compliance could be dependent on lung volume.

DYNAMIC EVENTS IN THE RESPIRATORY SYSTEM

Respiration requires the flow of gases into and out of the lungs. Many studies investigated the nature of the fluid flow, but early attention focused on a very simple model: that of a “resistance” to air flow. If a subject inhales to total lung capacity and then exhales to residual volume, the volume exhaled is the vital capacity. However, if performed with maximal expiratory effort, this experiment is called the forced vital capacity (FVC) maneuver (10). Mathematical analysis of the resulting data showed that the volume exhaled was almost exponentially related to time:

$$V = \text{FVC}(1 - e^{-kt}) \quad (2)$$

Where FVC was the volume eventually exhaled, t is time, and k was the constant in the exponential. Investigators acquainted with electrical analogues were quick to point out that this resembled the discharge of a capacitor C through a resistor R , with the time constant

$$k = 1/RC \quad (3)$$

The obvious analog was to equate C with the compliance of the respiratory system and R with the resistance to air flow. The pressure drop between the airway opening and the pleural surface was the driving force, so that the simple model for exhalation became:

$$P_{ao} - P_{pl} = R dV/dt + 1/CV \quad (4)$$

Where d/dt represents the time derivative, V the volume change of the lung, and C the compliance. The convention is

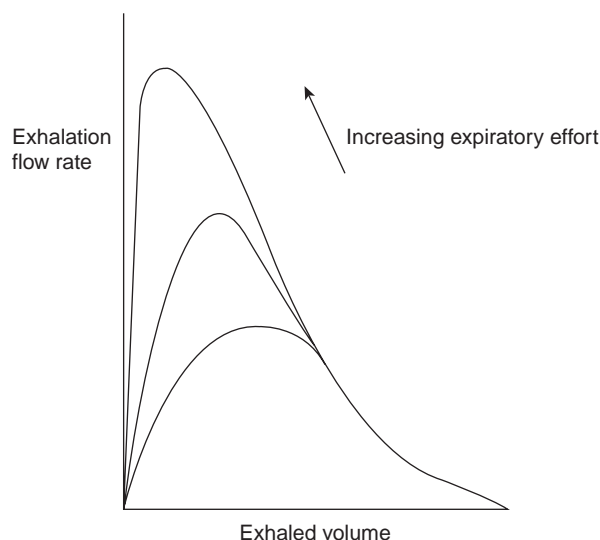


Figure 4. Maximum expiratory flow-volume maneuver at different degrees of effort.

that $dV/dt < 0$ represents exhalation. This equation fit conveniently with the classification of respiratory diseases into those that were restrictive and those that were obstructive. Restrictive diseases, such as muscle weakness, deformities of the chest wall, neurologic illnesses, and fibrosis of the lungs, were those that altered lung volumes and/or respiratory system compliance. Obstructive diseases such as asthma and bronchitis were those that interfered with the ability to quickly move air into, or especially out of, the lungs, leading to a high resistance.

Further support for these concepts came from the Maximal Expiratory Flow–Volume (MEFV) maneuver in which a subject performs the FVC maneuver while exhaling through a pneumotachograph. A plot of flow versus its integral (Fig. 4) has a straight-line portion, consistent with an exponential relationship, because if the pressure difference is constant the time derivative of equation 2 gives

$$dV/dt = kV \quad (5)$$

that is, a linear relationship between volume and flow rate. Once again, normal values are a function of race, gender, age, and size (11).

Equal Pressure Point Concept. Equation 4 would suggest that as the applied pressure difference is increased, the flow rate will increase indefinitely. However, the linear portion of the MEFV curve is relatively independent of effort. Beyond a certain point, trying to exhale more vigorously has no effect on increasing the expiratory flow. The idea of an “equal pressure point” was proposed to explain this flow limitation (12) (Fig. 5). If the subject closes the glottis, the pressure at the glottis will be equal to that in the alveoli because no flow occurs. Equation 4 shows that the pressure measured at the airway opening, or in this case just below the closed glottis, will be equal to the intrapleural pressure plus V/C . The quantity V/C is the elastic recoil force due to stretch in the alveolar walls. If the subject then performs a forced exhalation, intrapleural pressure

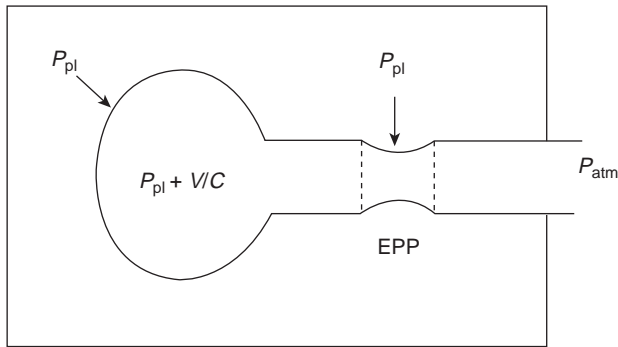


Figure 5. Equal pressure point concept. See text for details.

will be greater than atmospheric pressure and alveolar pressure is greater than this by V/C . Yet atmospheric pressure exists at the airway opening, so that at some point between the alveoli and the airway opening the pressure in the airway is equal to intrapleural pressure. This is the equal pressure point: The transmural pressure across the airway wall is zero, and if the EPP occurs in smaller airways unsupported by cartilage, they can collapse, impeding the flow. The model postulates that this occurs in such a fashion that the harder one tries to exhale, the more the airways are compressed, so that flow is limited. At a constant intrapleural pressure, as lung volume decreases during exhalation the equal pressure point moves toward the alveoli. Although the exact mechanism for flow limitation is undoubtedly much more complex (13,14), this model served as a simple explanation for the observed phenomenon. It also explained why flows would be reduced in patients with diseases such as emphysema, where airway obstruction from mucous and inflammation is not a major factor. In emphysema, fewer alveolar walls mean less elastic recoil, and so the equal pressure point occurs in smaller airways earlier in the exhalation.

MORE COMPLEX MODELS

These simple models underlie much of pulmonary function testing and mechanical ventilation. However, there is an extensive literature presenting more sophisticated analyses. Some of these are multicompartment versions of the simple models in that they postulate differing compliances and resistances in different regions of the lungs. Others propose nonlinear resistance and compliance elements, for example, a variable resistance representing the smaller, collapsible airways and a constant resistance for the more rigid upper airways (15). Some models are more consistent with physics in that they apply the principles of continuum mechanics, namely, conservation of mass, momentum, and energy, to the respiratory system. To do so, investigators have often had to restrict both the scope of the experiments and the portions of the respiratory system they model. Others have chosen to formulate input-output analyses without attempting to model the physical system (16). To date, none of these models has achieved wide acceptance or clinical utility.

THE PULMONARY VASCULATURE

Analyses of blood flow and mechanical changes in the pulmonary vasculature began with a concept analogous to airway resistance: the vascular resistance. The right ventricle pumps blood to the lungs, and the pulmonary veins return the blood to the left atrium. Measurements of mean pulmonary artery pressure (MPAP) and estimates of pressure in the left atrium (LAP) together with the determination of cardiac output (CO) allow the calculation of resistance as pressure drop is divided by flow:

$$\text{PVR} = (\text{MPAP} - \text{LAP})/\text{CO} \quad (6)$$

This is the clinically used model. Analyses of the “sheet flow” concept (17) have been used to explain the mechanics of blood flow through the lungs, but once again clinical applicability has been limited.

GAS TRANSPORT AND EXCHANGE

Although understanding the mechanical behavior of the respiratory system is important, ultimately one must both understand normal gas transfer, and then account for the effects of diseases on gas exchange. These concepts are among the most difficult to master in all of physiology. They bear detailed discussion because in practice physicians use the results of gas-exchange measurements to infer changes in mechanics rather than measure such changes directly. Once again, simpler concepts are used clinically although more complex and physically correct models do exist.

Respiratory control centers in the brain sense the partial pressures of carbon dioxide and oxygen in the blood and drive the respiratory muscles to move air into and out of the lungs. The amount of air that must move to the gas-exchanging alveolar surfaces each minute is known as alveolar minute ventilation, V_A . Some of the inhaled gas resides in the larger airways, where no gas exchange can occur. The volume of these airways is known as the anatomic dead space, V_D . If the gas were to flow in and out of the airways with no mixing, a concept known as “plug flow”, the total amount of air that is inhaled is composed of that which reaches the alveoli and that which is wasted in that it never reaches a gas-exchanging surface,

$$V_{\text{min}} = V_A + f V_D \quad (7)$$

Where f is the respiratory rate. Unless oxygen tensions are very low, carbon dioxide is the major driving force for respiratory effort. The partial pressure of carbon dioxide in the blood for normal people is ~ 40 Torr (5.320 kPa) and is closely regulated near that level.

The structure of the airways and blood vessels is such that local air flow and blood flow are closely matched: for the normal lung the ratio of ventilation to perfusion for any portion of the lung is nearly uniform even though different areas of the lung receive markedly different amounts of air and blood. The pulmonary vessels are capable of constriction in response to low oxygen tension, a process known as “hypoxic pulmonary vasoconstriction”.

This process tends to keep local blood flow well matched to local air flow (18). So, for the purposes of estimating gas transfer, one can lump all 300 million alveoli together and treat the normal lungs as if there were only one large alveolus receiving all the blood flow and all the air flow. One can then estimate (1) what level of alveolar ventilation would be required to maintain a normal partial pressure of carbon dioxide, and (2) what level of blood oxygenation would be expected for a given inhaled oxygen concentration.

Partial Pressures of Gases Related to Volume and Composition

It is molecules of gas that flow in the respiratory system, not volumes of gas. Pressure, volume, and the number of molecules are related by the equation of state:

$$PV = nRT \quad (8)$$

Where P is the pressure, V is the volume, n is the number of moles of gas, T is the temperature, and R is the gas constant. When physiologists speak of volume rather than moles, they must specify the pressure, temperature, and humidity at which that volume is measured. Two sets of conditions are used in respiratory physiology. The first is BTPS, or body temperature and pressure, saturated (with water vapor). Volumes given as BTPS are those that the gases would occupy if they were at 37 °C, standard atmospheric pressure (760 Torr or 101.080 kPa), and fully saturated with water vapor. The partial pressure of water vapor at 37 °C, the normal human body temperature, is 47 Torr or 6.266 kPa. Volumes given as STPD are standard temperature and pressure, dry, that is, at 0 °C, standard atmospheric pressure, with all the water vapor removed. The equation of state can be used to convert between conditions.

As an example, consider what would occur if an 80 kg human were to use glucose as the only fuel. A moderately ill person requires $\sim 25 \text{ cal} \cdot \text{kg}^{-1} \cdot \text{day}^{-1}$ (104.6 J). The metabolism of glucose generates roughly $4 \text{ cal} \cdot \text{g}^{-1}$ (16.75 J) and consumes an amount of oxygen equal to the carbon dioxide produced:



To generate 2000 cal (8373.6 J) would require use of 500 g, or 2.8 mol, of glucose/day. This would in turn use 16.8 mol of oxygen and generate 16.8 mol of carbon dioxide.

Using the equation of state, or remembering Boyle's law, we know that 1 mol of gas at STPD occupies 22.4 L. During a day, our patient would use 376 L of oxygen and generate an equal amount of carbon dioxide. On a per-minute basis, oxygen consumption and carbon dioxide production would be

$$\dot{V}_{\text{O}_2} = \dot{V}_{\text{CO}_2} = 260 \text{ mL} \cdot \text{min}^{-1}$$

These are volumes at STPD. What would the volumes be at BTPS? At STPD,

$$n_{\text{O}_2} = (760 \text{ Torr})(260 \text{ mL})/R(273 \text{ }^\circ\text{C}) \quad (10)$$

These same molecules heated to body temperature and humidified would occupy a larger volume. Under these

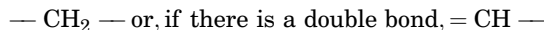
new conditions, the pressure due to the dry gases is lower and the temperature is higher.

$$n_{\text{O}_2} = (760 - 47)V/R(310 \text{ }^\circ\text{C}) \quad (11)$$

The number of molecules is the same, so equating 9 and 10 yields

$$V = (760 * 260/273) * (310)/(760 - 47) = 314 \text{ mL} \quad (12)$$

For these calculations, glucose was the sole fuel used to meet the daily energy requirement. However, normally fat forms part of the fuel supply and because the units of a fatty acid chain lack the oxygen atom:



fat as an energy source consumes more oxygen per carbon dioxide produced, and therefore lowers the respiratory quotient (RQ). We generally estimate

$$\text{RQ} = \dot{V}_{\text{CO}_2}/\dot{V}_{\text{O}_2} = 0.8$$

when performing calculations, so that if

$$\dot{V}_{\text{O}_2} = 310 \text{ mL} \cdot \text{min}^{-1} \text{BTPS}$$

Then,

$$\dot{V}_{\text{CO}_2} = 250 \text{ mL} \cdot \text{min}^{-1} \text{BTPS}$$

Alveolar Minute Ventilation

With each breath, we inhale surrounding air with 21% oxygen. We do not use all of it (only about a quarter) and exhale the remainder. The amount of air we need each minute is usually determined by the need to rid ourselves of carbon dioxide. What volume of gas would have to reach the alveoli each minute to remove $250 \text{ mL} \cdot \text{min}^{-1}$ of carbon dioxide while keeping blood, and hence alveolar, carbon dioxide tension at 40 Torr?

The fraction of alveolar gas that is carbon dioxide must be

$$F_{\text{ACO}_2} = \dot{V}_{\text{CO}_2}/\dot{V}_A \quad (13)$$

But partial pressures are related to the proportion of the gas in the total mixture:

$$P_{\text{ACO}_2} = F_{\text{ACO}_2}(P_{\text{atm}} - P_{\text{H}_2\text{O}}) \quad (14)$$

Combining 12 and 13,

$$P_{\text{ACO}_2} = \dot{V}_{\text{CO}_2}/\dot{V}_A(P_{\text{atm}} - 47) \quad (15)$$

Some textbooks show P_{ACO_2} multiplied by a constant, 0.863, in this equation. In those books, CO_2 production is given at STPD and \dot{V}_A is at BTPS. The constant results from using the equation of state to convert \dot{V}_{CO_2} to BTPS.

Substituting standard atmospheric pressure of 760 Torr or 101.325 kPa, a desired alveolar CO_2 tension of 40 Torr (5.332 kPa), and CO_2 production of $250 \text{ mL} \cdot \text{min}^{-1}$ leads to the conclusion that alveolar ventilation must be

$$\dot{V}_A = 4.45 \text{ L} \cdot \text{min}^{-1} \text{BTPS}$$

If the inhaled oxygen is 21% of the respired gases and oxygen utilization is $310 \text{ mL} \cdot \text{min}^{-1}$, less than a one-third of the inhaled oxygen is consumed and the rest is exhaled. Alveolar minute ventilation is therefore usually determined by the need to exhale the carbon dioxide produced.

Expected Oxygen Tension

We next ask what degree of blood oxygenation we would expect if all the alveoli had the same ventilation/perfusion ratio, that is, if our assumption that the normal lung can be modeled as a single alveolus is correct. Consider the steps involved in inhalation:

1. Gas is brought into the nose and upper airways, heated to body temperature, and humidified.

The gases we inhale are almost completely composed of nitrogen and oxygen at the local barometric pressure. When water vapor is introduced, the partial pressure of oxygen is the fraction of oxygen inhaled times the sum of the partial pressures of the dry gases, which is barometric pressure minus the water vapor pressure:

$$P_{I_{O_2}} = (P_{\text{atm}} - 47)FI_{O_2} \quad (16)$$

Where the subscript I denotes inhaled.

2. This gas reaches the alveoli, where a certain amount of oxygen is taken up and carbon dioxide is added. Although not strictly true (the correction is relatively minor), the simplest approach is to consider how much oxygen is used compared to carbon dioxide delivered. The respiratory quotient can be used

$$V_{O_2} = V_{CO_2}/RQ \quad (17)$$

and an estimate of how much the alveolar P_{O_2} falls compared to the inhaled P_{O_2} becomes

$$P_{O_2} = PA_{CO_2}/RQ$$

giving

$$PA_{O_2} = PI_{O_2} - PA_{CO_2}/RQ$$

or

$$PA_{O_2} = (P_{\text{atm}} - 47) * FI_{O_2} - PA_{CO_2}/RQ \quad (18)$$

For an RQ of 0.8, normal CO_2 tension, and breathing room air (21% O_2) this gives

$$PA_{O_2} = 100 \text{ Torr or } 13.300 \text{ kPa}$$

Note that the calculations of alveolar minute ventilation and expected alveolar oxygen tension were based upon the hypothesis that for a normal lung, ventilation and perfusion are closely matched. Some 300 million alveoli were modeled as a single unit in terms of gas exchange. Miraculously, for the normal lung this turns out to work quite well. The difference between measured arterial oxygen tension (PA_{CO_2}) and calculated alveolar oxygen tension (PA_{CO_2}) is known as the alveolar-arterial oxygen gradient, or A-a gradient. In young normal subjects it is usually $<10 \text{ Torr}$ (1.330 kPa).

The Role of the Cardiovascular System

Blood delivered by the heart brings carbon dioxide to the lungs for elimination and oxygen to the cells of the body. Figure 6 shows the relationship between the partial pressures of the gases and the amounts of those gases in the blood. For simplicity's sake, these are shown as single curves although the amount of one gas present does affect the carrying capacity for the other somewhat (the Bohr and Haldane effects). The curves are very different. That for carbon dioxide is almost linear, while that for oxygen is sigmoid shaped. This difference is crucially important. Together with the fact that the major regulator of alveolar ventilation, is carbon dioxide, mismatches in the distribution of ventilation and perfusion caused by acute lung injuries will have less effect on carbon dioxide tension in the blood than they will on oxygenation, as we shall see below.

The cardiac output is $\sim 5 \text{ L} \cdot \text{min}^{-1}$. Blood with a hemoglobin content of $14 \text{ g} \cdot \text{dL}^{-1}$ exposed in the lungs to 100 Torr (13.330 kPa) partial pressure of oxygen will contain $\sim 195 \text{ mL}$ of O_2 per liter of blood. Therefore, 965 mL of oxygen will be pumped to the body every minute. If oxygen consumption is $310 \text{ mL} \cdot \text{min}^{-1}$, 655 mL of oxygen will return through the venous circulation unused. This translates to a venous hemoglobin saturation of $\sim 70\%$ with a venous blood oxygen tension of 40 Torr (5.320 kPa).

The carbon dioxide tension of arterial blood is regulated near 40 Torr (5.320 kPa). The arterial blood content of carbon dioxide is therefore $\sim 350 \text{ mL} \cdot \text{L}^{-1}$. If carbon dioxide production is $250 \text{ mL} \cdot \text{min}^{-1}$ and cardiac output is $5 \text{ L} \cdot \text{min}^{-1}$, $\sim 50 \text{ mL}$ of carbon dioxide will be added to each liter of blood, so that venous partial pressure of carbon dioxide is $\sim 46 \text{ Torr}$ (6.118 kPa).

During light to moderate exercise, or with a moderate illness, the normal response to increased oxygen utilization and carbon dioxide production is an increase in alveolar minute ventilation and an increase in cardiac output rather than an increase in oxygen extraction at constant

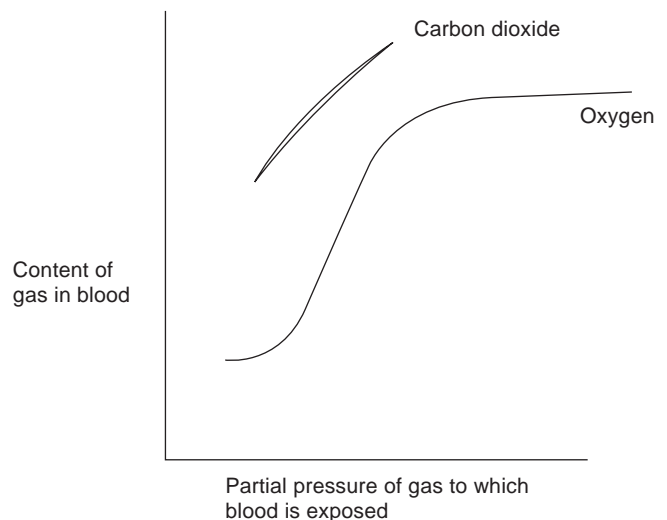


Figure 6. Carrying capacity of blood for carbon dioxide and oxygen.

cardiac output. If the cardiovascular system is impaired, oxygen extraction will increase.

RESPIRATORY MECHANICS IN DISEASE STATES

Several observations of mechanical behavior of the lung show that when diseases occur, the simple resistance-compliance model of the lung may cease to apply. The first concept relating to disease is that of dynamic compliance.

If a simple resistance-capacitor circuit is driven by a sinusoidal voltage, an analysis of voltage and current waveforms will give the same value for resistance and capacitance independent of the frequency of the sinusoid. However, attempts to calculate compliance from a diseased subject while varying the breathing frequency appeared to give a value for compliance that was dependent on the respiratory rate. This became known as “dynamic compliance”. The reason for this is that disease processes are seldom homogeneous. The inhomogeneity leads to varying mechanical properties, and a more appropriate model is then a sinusoidal driving force applied to several resistors and capacitors in parallel. In this case,

$$P \sin \omega t = R_1 dV_1/dt + 1/C_1 V_1 = R_2 dV_2/dt + 1/C_2 V_2 = \dots \quad (19)$$

Adding the equations above would give

$$aP \sin \omega t = (R_1 dV_1/dt + R_2 dV_2/dt + \dots) + (V_1/C_1 + V_2/C_2 + \dots)$$

where a is the number of parallel components. However, one measures only the total flow and total volume exhaled. Analyzing data obtained from an inhomogeneous lung using equation 4 would make the resulting R and C functions of all the resistances and compliances in addition to how the total volume and flow were distributed within the system, which itself would be a function of frequency. The term dynamic compliance is a misnomer: The existence of dynamic compliance is an indicator that a simple one-compartment model does not apply (19).

And as one would then expect, the MEFV curves from many subjects with obstructive lung diseases show that not only are the magnitudes of flows reduced, but the portion of the curve that was almost linear in normal subjects now has a definite curvature. This could result from having several compartments of different mechanical properties emptying at different rates.

These concepts fit well with observations of patients' breathing patterns. If increased airway resistance seems to be the issue, the patients breathe more slowly with a greater tidal volume, which might allow enough time for the slow compartments to empty. Equation 1 shows that complete emptying will require a time more than three time constants (RC). If retention of air is a problem because of loss of elasticity, such as in patients with emphysema and high FRC or with structural abnormalities lowering TLC, patients may breathe more rapidly with a smaller tidal volume. In addition, at times of increased airway resistance, some patients purse their lips while exhaling.

According to the Equal Pressure Point concept, creating a sharp pressure drop at the lips would raise the pressure throughout the airway system, moving the equal pressure point into larger airways that might not be so collapsible, and possibly allowing an increased expiratory flow rate. These patterns seem to be set by the respiratory controller in the brain in a way that may minimize the work and discomfort of breathing (20,21).

The simple dynamic models applicable to the normal lung have been extended in many ways to attempt to model the effects of disease processes. For example, investigators realized that at higher frequencies, such as those employed with jet ventilators or forced oscillatory ventilation in infants, the inertia of the fluid would play a role. They introduced the electrical analog of inertia, an inductance I , into their models. The models became more complex as investigators attempted to assign various components of the respiratory system their own electrical analog properties, and to allow these properties to vary in different areas of the lungs. However, measurements of lung volumes, compliance, and flows remain the mainstay of clinical pulmonary function testing and are based upon the simple concepts presented above.

EFFECTS OF DISEASE ON GAS TRANSPORT

In acute illnesses or acute exacerbations of chronic lung diseases, formal mechanical testing is usually not performed except in cases such as asthma, where abrupt and dramatic increases in airway resistance markedly reduce the expiratory flows. Clinicians draw conclusions about changes in mechanics from changes in arterial oxygen or carbon dioxide tensions. Understanding these changes is important for both assessing acute decompensations in ambulatory patients and for providing effective mechanical ventilation. These changes are perhaps the most misunderstood aspect of respiratory pathophysiology.

For the normal lung, ventilation and perfusion are matched to assure efficient gas transport. The elimination of carbon dioxide is the usual determinant of alveolar minute ventilation. Diseases affecting the respiratory system can decrease the total amount of air or blood delivered or disturb this matching. Although there are many different disease, the mechanisms are few. First, the membranes responsible for gas transport can be destroyed leaving fewer alveoli to do the job of gas transport, but without significantly affecting the matching. Second, the airways and/or the small blood vessels can become inflamed or blocked, either partially or completely, in a way that interferes with the normal matching of air and blood flows. This results in areas of the lungs with widely varying ventilation/perfusion ratios. Third, problems with the nervous system, with the musculature, or with the structure of the chest wall itself can limit alveolar ventilation. Interpreting the changes in blood gas tension using the simple model for the normal lung may lead to erroneous interpretations.

For example, consider what would happen if 50% of the blood flow went to alveoli, which were fluid filled or

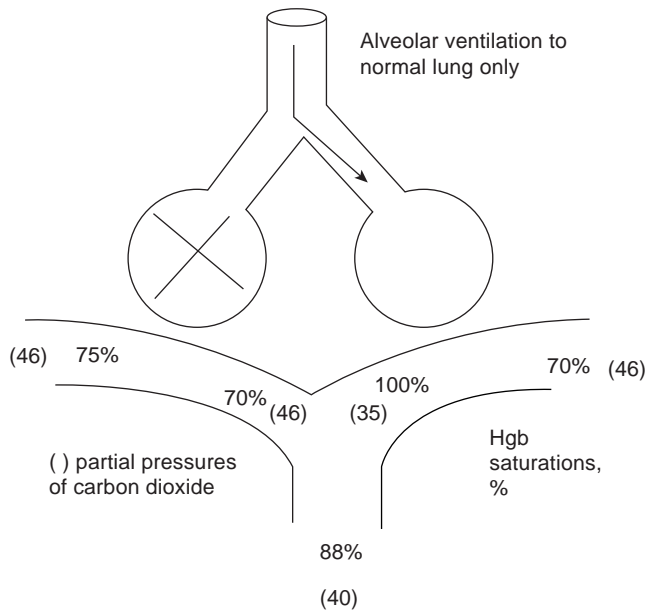


Figure 7. Effect of a 50% shunt on oxygen delivery and partial pressures of carbon dioxide. See text for details.

collapsed (Fig. 7). In the normal person, cardiac output is $\sim 5 \text{ L} \cdot \text{min}^{-1}$. Venous oxygen tension is ~ 40 Torr (5.320 kPa), and at that pressure the hemoglobin is $\sim 70\%$ saturated. Venous carbon dioxide tension is ~ 46 Torr (6.118 kPa). Normally, the lungs would deliver oxygen, remove carbon dioxide, and the arterial blood would have a P_{aO_2} of 100 Torr (13.330 kPa) and a P_{CO_2} of 40 Torr (5.333 kPa). Because respiration is driven by the carbon dioxide tension, the brain would try to keep arterial P_{CO_2} at 40 Torr (5.320 kPa). But one-half of the blood passing through the lung would maintain its venous content of carbon dioxide, which has a venous partial pressure of 46 Torr (6.118 kPa). This means that the respiratory controller would increase alveolar minute ventilation to the functional alveoli, and alveolar partial pressure of carbon dioxide in those alveoli would fall to something like 32 Torr (4.256 kPa). If one measured the increase in alveolar minute ventilation, it would amount to some 10–15% increase. If one attempted to interpret this according to the one-compartment model, which applied to normal lungs, one might conclude that some of the alveolar ventilation was wasted, that is, that it went to areas of the lung that did not participate in gas exchange (that there was an increase in physiologic dead space). However, that is not what the figure above shows: It shows that blood flow goes to unventilated areas of the lung, not that ventilation goes to unperfused areas of the lung. The true situation is a shunt, whereas the model would predict an increase in dead space.

What would happen to oxygenation? The venous blood would not pick up any oxygen in the unventilated alveoli. Because of the shape of the oxyhemoglobin dissociation curve, very little extra oxygen could be delivered to the blood flowing through the functional alveoli even if the inspired oxygen concentration were increased to 100%.

Note the implications of this. Any condition that causes a mismatch between the amount of blood flow to given alveoli and the amount of air flow will affect oxygenation much more profoundly than it will affect carbon dioxide elimination. There are two major reasons for this: (1) chemosensors are very sensitive to increases in CO_2 tension and will increase alveolar ventilation to compensate; and (2) this compensation is possible for CO_2 and not for O_2 because of the different shapes of the gas content: partial pressure curves for the two gases.

MECHANICAL VENTILATION

The mechanisms by which diseases affect gas transport can be used to understand mechanical ventilation. Because the primary determinant of carbon dioxide tension is alveolar ventilation, adjustments in respiratory rate and tidal volume will allow control of carbon dioxide levels. Because oxygenation is dependent on the fraction of inspired oxygen AND upon how ventilation and perfusion are distributed, the FI_{O_2} is just one of the controls affecting oxygenation. Positive end-expiratory pressure (PEEP) can be applied to try to recruit alveoli that might be collapsed and responsible for the shunting of blood. That is, end-expiratory pressure at the mouth is held above atmospheric pressure to hold alveoli open. The price paid is that normal alveoli can become overdistended. Therefore the exact level of PEEP that should be used is still controversial (22,23).

Although there are many types of ventilators that perform many sophisticated functions, there are two basic ways to ventilate a human being. For adults, volume-cycled ventilators are usually used. A tidal volume is set, and pressures measured for safety reasons. For infants, tidal volumes are too small to accurately measure, so that pressure-cycled ventilators are used. In this case, a peak inspiratory pressure and PEEP are chosen, and the volume actually delivered depends on the mechanics of the ventilator and of the infant's respiratory system (24). In this case, tidal volume becomes a variable so that safe pressure limits are never exceeded.

Although there is much written about "barotrauma" from mechanical ventilation, that is actually a misnomer. Experiments on cells from the lungs show that when a cell is stretched beyond a certain point, inflammatory mediators are generated and the cell can die (25). The correct term is "volutrauma", but because we cannot measure cell stretch adequately, pressure limits are generally respected (26).

CONCLUSION

This article does not attempt to summarize the literature on continuum mechanics, statistical models, or tissue mechanics of the respiratory system. However, the simple concepts described above are essential for understanding pulmonary function testing, respiration, and patient support with mechanical ventilators. Once these concepts are mastered, the complex and often conflicting literature can be critically read and understood.

BIBLIOGRAPHY

1. Kollef MH, Shapiro SD, Boyd V, Silver P, Von Harz B, Trovillion E, Prentice D. A randomized clinical trial comparing an extended-use hygroscopic condenser humidifier with heater-water humidification in mechanically ventilated patients. *Chest* 1998;113:759.
2. Huang W, Yen RT, McLaurine M, Bledsoe G. Morphometry of the human pulmonary vasculature. *J Appl Physiol* 1996;81: 2123.
3. Wang NS. Anatomy and physiology of the pleural space. *Clin Chest Med* 1985;6:3.
4. Lichtenstein O, Ben-Haim SA, Saidel GM, Dinnar U. Role of the diaphragm in chest wall mechanics. *J Appl Physiol* 1992;72:568.
5. Scarpelli EM. Physiology of the alveolar surface network. *Compar Biochem Physiol Part A, Mol Int Physiol* 2003;135: 39.
6. Escolar JD, Escolar A. Lung hysteresis: A morphological view. *Histol Histopathol* 2004;19:159.
7. Ainsworth SB, Milligen DWA. Surfactant therapy for respiratory distress syndrome in premature neonates: A comparative review. *Am J Resp Med* 2002;1:417.
8. Lai-Fook SJ. Pleural mechanics and fluid exchange. *Physiol Rev* 2004;84:385.
9. Milic-Emili J, Mead J, Turner JM, Glauser EM. Improved technique for estimating pleural pressure from esophageal balloons. *J Appl Physiol* 1964;19:207.
10. Pride NB. Tests of forced expiration and inspiration. *Clinics Chest Med* 2001;22:599.
11. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999;158:179.
12. Hyatt RE, Wilson TA, Bar-Yishay E. Prediction of maximal expiratory flow in excised human lungs. *J Appl Physiol* 1980;48:991.
13. Elliott EA, Dawson SV. Test of wave-speed theory of flow limitation in elastic tubes. *J Appl Physiol* 1977;43:516.
14. Kamm D. Airway wall mechanics. *Ann Rev Biomed Eng* 1999;1:47.
15. Polak AG, Lutchen KR. Computational model for forced expiration from asymmetric normal lungs. *Ann Biomed Eng* 2003;31:891.
16. Zhang Q, Suki B, Lutchen KR. Harmonic distortion from nonlinear systems with broadband inputs: Applications to lung mechanics. *Ann Biomed Eng* 1995;23:672.
17. Fung YC, Sobin SS. Theory of sheet flow in lung alveoli. *J Appl Physiol* 1969;26:472.
18. Brimiouille S, Lejeune P, Naeije R. Effects of hypoxic pulmonary vasoconstriction on pulmonary gas exchange. *J Appl Physiol* 1996;81:1535.
19. Otis AB, McKeerow CB, Bartlett RA, Mead J, McLroy MB, Silverstone NJ, Radford EP Jr. Mechanical factors in the distribution of pulmonary ventilation. *J Appl Physiol* 1956;8:427.
20. Oku Y, Saidel GM, Altose MD, Cherniack NS. Perceptual contributions to optimization of breathing. *Ann Biomed Eng* 1993;21:509.
21. Guz A. Brain, breathing, and breathlessness. *Respir Physiol* 1997;109:197.
22. Schmitt J-M, Vieillard-Baron A, Augarde R, Prin S, Page B, Jardin F. Positive end-expiratory pressure titration in acute respiratory distress syndrome patients: Impact on right ventricular outflow impedance evaluated by pulmonary artery Doppler flow velocity measurements. *Crit Care Med* 2001;29: 1154.
23. Rouby J-J, Puybasset L, Nieszkowska A, Lee Q. Acute respiratory distress syndrome: Lessons from computed tomography of the whole lung. *Crit Care Med* 2003;31(Suppl): S285.
24. Ligas JR, Moslehi F, Epstein MAF. Occult positive end-expiratory pressure with different types of mechanical ventilators. *J Crit Care* 1990;5:95.
25. Pugin J. Molecular mechanisms of lung cell activation induced by cyclic stretch. *Crit Care Med* 2003;31(Suppl): S200.
26. Acute Respiratory Distress Syndrome Network (ARDS). Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *NEJM* 2000;342:1301.

See also CONTINUOUS POSITIVE AIRWAY PRESSURE; HIGH-FREQUENCY VENTILATION; PNEUMOTACHOMETERS; PULMONARY PHYSIOLOGY.

RESPIRATORY MONITORING. See VENTILATORY MONITORING.

RESPIRATORY SOUNDS. See LUNG SOUNDS.

RESUSCITATION, CARDIOPULMONARY. See CARDIOPULMONARY RESUSCITATION.

RHEOLOGY, BLOOD. See BLOOD RHEOLOGY.

SAFETY PROGRAM, HOSPITAL

JOSEPH F. DYRO
Setauket, New York

INTRODUCTION

This article describes the broad scope of safety issues in the hospital and recommends ways in which hospitals should address these issues. Particular emphasis is placed on medical devices and instrumentation and how they interrelate with the hospital's environment, facility, patients, and device users and other personnel. Medical device safety depends in large part upon a comprehensive medical device technology management program, which includes elements ranging from technology assessment, evaluation, and procurement to device replacement planning and includes such components as user training, medical device incident investigation, and device quality assurance. The clinical engineer by education, training, and experience is shown to be ideally suited to implement and execute a technology management program and to lead in the effort to assure hospital-wide patient safety. Available resources are described including web-based training, distance learning, standards, publications, and professional organizations.

Innovative surgical techniques, improved invasive and noninvasive diagnostic procedures, advanced diagnostic and therapeutic medical devices, pharmacological processes continue to benefit the treatment of the sick; however, advanced technology has been a mixed blessing to recipients of these advances. Its positive impact on healthcare has been countered by the creation of problems engendered by an increasingly complex medical device and instrumentation environment. In the use of devices and techniques, misuse or procedural error can and does occur. Some level of risk is associated with everything that is done to care for the patient. Healthcare professionals must keep abreast of the advancing technologies, be aware of their limitations and risks, and manage them. The hospital safety program is an integral part of the system of risk reduction and can and should be broad in scope, not limited by undue attention to only one or two risk categories such as slips and falls or needle sticks. It should encompass means to address and control the risks of injury to patient and staff, risks that arise from all sources, not only those specifically related to medical devices, but those related to the performance of the individual healthcare giver.

THE PROBLEM WITH HEALTHCARE

The complexity and proliferation of the three components of technology (i.e., drugs, devices, and procedures) continues unabated. In the hospital complex, interrelationships exist among environment; facilities and utilities;

medical and nonmedical policies; and procedures, economics, ethics, and human performance. Inappropriate drugs, devices, and procedures; defectively designed, manufactured, and maintained medical devices; inappropriate staffing; inability to follow procedures, neglect of duties, inattention, and ignorance; and deficient education and training are among the factors that increase the risk of injury in the hospital.

The Institute of Medicine (IOM) of the National Academy of Sciences produced a report containing estimates as high as 98,000 deaths occurring annually in U.S. hospitals caused by medical errors (1). About 1 million medical injuries occur annually in the United States, with 200,000 involving some negligence (2). Such iatrogenic injury (injury caused by the physician, now more broadly defined as injury caused by giving care in a hospital) has been well documented (3). Medical devices used in delivering care have a finite probability of doing harm to the patient or to the person using the device (4) and are included as one source of iatrogenic injury. Estimates of the percentage of device-related incidents over all incidents vary from 1 to 11% (2).

Patient Safety Movement

The IOM report spawned a patient safety movement. Organizations, actively working to identify the weaknesses in the healthcare arena, are unified in their support of patient safety initiatives (5–7). The Institute of Medicine (9), in its sequel report, *Crossing the Quality Chasm: A New Health System for the 21st Century*, calls for a health care delivery system that is safe, effective, patient-center, timely, efficient, and equitable. Patient safety can be improved by implementing and executing effective hospital safety programs.

Safety Programs

Programs exist that address the overall safety concerns within a hospital. Under the umbrella of a *total hospital safety program*, there are various components, such as fire safety, infection control, medical device safety, and radiation safety. Guidelines for a comprehensive total hospital safety program incorporating all the various components are given below. Following this is presented a program that addresses medical device safety, in particular. Before elaborating upon these two safety programs, to obtain a clearer sense of the problem with health care the injurious forces and mechanisms present in a hospital are described and their interrelation with patients, staff, facilities, environment, medical technology are described.

Injurious Forces and Mechanisms

The hospital is a place where injury mechanisms abound. The nature of a hospital's business concentrates the injury mechanisms. For example, life-threatening situations demand rapid response by medical personnel often

utilizing a host of complex technologies. All components of the system must work for successful outcome. One element failure, for example, an error in device use, administration of the wrong drug, or inappropriate surgical technique, can produce disastrous results even though all other elements are working. Injury to patients, personnel, and visitors can arise from many different sources within a hospital. The following is a list of some of the many safety concerns and examples of hospital occurrences:

- **Fire:** Electrosurgical energy in an oxygen-enriched atmosphere has ignited a fuel source such as the hair of a patient resulting in a rapidly spreading, deadly conflagration.
- **Air, medical gases, and vacuum:** Crossed-medical gas pipelines have killed patients mistakenly given nitrous oxide instead of oxygen.
- **Water:** Water with inappropriate chemical content used in dialysis has injured patients.
- **Chemicals, solvents, sterilizing agents, skin preparation solutions, anesthetic gases:** Anesthetic gases can cause birth defects; ethylene oxide used in sterilization has caused severe skin lesions.
- **Drugs:** Administration of wrong or inappropriate amounts of medications have killed patients.
- **Filth, microorganisms, vermin:** Inadequately sterilized bronchoscopes have cross-contaminated hundreds of patients.
- **Waste, bodily fluids, sharps (needles, scalpels):** Improperly discarded needles have punctured hospital workers infecting them with human immunodeficiency virus (HIV).
- **Sound, noise:** Ambient noise levels have obscured audible alarms.
- **Ionizing and nonionizing radiation [X-rays, laser, ultraviolet (UV), visible light, infrared (IR), microwave]:** Interventional cardiologists have burned patients by using excessive X-ray exposure times during fluoroscopically guided catheter placement.
- **Electricity:** Lethal electric shock was delivered to a patient by improper connection of ECG leads to a power cord.
- **Natural and unnatural disasters:** Hurricanes have disrupted electrical power and back-up generators have failed to function.
- **Mechanical stress:** Static and dynamic forces on the body can result in injury. For example, excessive and prolonged mechanical pressure to parts of the body during surgery has resulted in pressure necrosis of tissue. Among the highest causes of injury are slips and falls and back strains.
- **People:** Human error, for example, administering the wrong medication or cutting-off the wrong leg, is the largest single cause of injury in the hospital. Abduction and elopement are also people-related safety concerns.
- **Devices:** A defectively designed check valve has failed preventing ventilation of a patient resulting in brain death.

From the above, it is clear that a hospital patient is exposed to a variety of risks many of which involve, in some way, medical devices and instrumentation. Human error resulting in patient injury often involves the use of a medical device in an inappropriate fashion. Such use may signify that the device was designed with inadequate attention paid to human factors considerations (9). A hospital safety program must pay due attention to the relationships among devices, people, places and things, and take a systems perspective on safety.

Systems Perspective on Safety

The hospital is a complex environment where medical devices and instrumentation utilized for diagnostic, therapeutic, and rehabilitative purposes interact with each other and with patients, staff, facilities, and the environment. A systems approach to understanding the mechanisms contributing to incidents and accidents aids in the development of programs to control and minimize safety risks (10). The efficacy and safety of the delivery of patient care depends, in general, on five main components: medical device; operator; facility; environment; and patient. The following is a description of these components along with examples of the contributions they can make toward creating hazardous situations.

Medical Devices

The device itself can be the sole cause of an iatrogenic injury. A medical device can be rendered injurious by the actions or inactions of the device inventor, designer, manufacturer, shipper, inspector, maintainer, or user. For example, the manufacturer may fail to properly assemble or construct an otherwise efficaciously designed medical device. The shipper who transports the device from the manufacturing plant to the hospital may damage a properly manufactured device. The inspector at the hospital may fail to properly inspect the device upon receipt, allowing a damaged or defective device to be placed into service. The maintainer may fail to keep the device operating within manufacturer's specifications by not adequately performing preventive maintenance and calibration procedures. The manufacturer may fail to provide adequate warnings and instructions. The user may abuse or misuse the device, rendering it hazardous to the patient. Finally, the device may not be retired when it has reached the end of its useful life and is subject to more failure modes and is not up to date with current medical practice.

Injury can result from the absence or failure to utilize medical devices that are accepted as the standard of care in the community. Physicians and nurses have an obligation to utilize technology that has been accepted as the standard of care, and liability is incurred if available technology is not used.

Definition

For an all-encompassing safety program, the definition of medical device adopted by the U.S. Food and Drug Administration (FDA) is recommended (11):

"The term device means an instrument, apparatus, implement, machine, contrivance, implant, in vitro

reagent, or other similar or related article, including any component part, or accessory, which is

1. Recognized in the official National Formulary, or the United States Pharmacopoeia, or as supplements to them.
2. Intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in humans or other animals.
3. Intended to affect the structure or any function of the body of humans or other animals, and that does not achieve any of its principal intended purposes through chemical action within or on the body of humans or other animals and that is not dependent on being metabolized for the achievement of any of its principal intended purposes.”

All devices used in the hospital or provided by the hospital must be addressed whether they are purchased, donated, leased, or owned by a physician or are under evaluation. Technically, a safety program must include devices ranging from tongue depressors to magnetic resonance imaging (MRI) units. Both reusable and disposable devices must be considered. As medical device technology becomes more complex and new modalities are employed for diagnostic or therapeutic intervention, the hospital safety program must adapt to these changes. It was not that long ago that such issues as laser safety, magnetic field strength hazards, and nonionizing radiation hazards were unknown concepts. A hospital safety program must also include a way of dealing with devices used outside the hospital, such as devices provided by or through the hospital to enable ambulatory and home care or devices used in the transport of patients to or between hospitals. A hospital department such as clinical engineering that extends its services to other entities such as a doctor's office or clinic must adhere to the same principles that guide its in-house safety program such as regular preventive maintenance schedules.

Electricity

Medical devices that are electrically operated can pose risk to patients and staff of electric shock. Incoming inspection and periodic inspections thereafter can assess the safety of such devices. As with all electrically operated devices, care should be taken to ensure that surfaces that may be contacted by patients and operators are not energized because, for example, of poor design, manufacturing defect, or deterioration and inadequate maintenance and repair. Electrical safety of medical devices is but one aspect of safety as is shown below. Safety programs should not be obsessed with electrical hazards, shown to be a relatively minor contributor to death and injury in hospitals since first introduced several decades ago as a major problem (12).

Chemicals

Many potentially injurious chemical compounds are or have been associated with the operation of medical devices and instrumentation. For example, ethylene oxide gas used in hospital sterilizers poses a risk to operators and to

patients if improperly utilized (13). Iatrogenic complications can be introduced by infusion pumps, widely used for intravenous and intraarterial fluid and drug administration. Incompatibility of certain drugs with the plastics from which infusion pump administration sets are made is a source of risk. Nitroglycerin in solutions administered by infusion pumps has been shown to interact with certain formulations of poly(vinyl chloride) (PVC) tubing, ultimately decreasing the potency of the drug and making the administration of a known amount impossible to gauge. Other examples include allergic reactions to latex gloves, skin preparation agents, and contrast media and adverse reactions to implantable devices such as artificial joints and nerve stimulators.

Sound and Electromagnetic Radiation

A neonate's auditory system is particularly susceptible to injury from high sound levels as can occur in infant incubators. Diagnostic ultrasound units must be properly designed and maintained to ensure that output levels remain within acceptable limits recommended in safety standards.

The hazards of ionizing radiation became well known not long after the development and use of the X-ray machine. A good radiation safety program is essential to ensure that radiographic equipment and protective measures in radiographic suites meet acceptable performance and safety standards and that operators are appropriately trained and follow established safety procedures.

Nonionizing radiation is a significant health hazard in all hospitals and includes UV, microwave, and laser radiation. Ultraviolet therapy is employed in the treatment of some skin disorders. However, UV radiation can adversely affect the eye (keratitis) and the skin (erythema and carcinogenesis). Microwave radiation is commonly used in physical therapy for diathermy treatment of patients. Microwave effects are largely thermal, with the eye being the most susceptible organ (cataractogenesis). Surgical lasers are particularly hazardous since the beam can travel large distances with little attenuation and can reflect off surfaces in the room. The intensity of the beam is of necessity sufficient to burn body tissue as is required in surgical procedures. Momentary eye contact with the beam can cause severe eye damage. A laser safety program is recommended wherever lasers are used. Eye protection and restriction of the area to trained personnel are essential steps in the program.

Alarms

The Joint Commission on Accreditation of Healthcare Organizations identified appropriate attention to alarms as one of six patient safety goals of 2003 (14). Alarms are sometimes ignored, unheard, unnoticed, misinterpreted, or disabled. A life-threatening alarm that goes unattended can have disastrous consequences for the patient. Alarms are rendered ineffective for a host of reasons, some of which are listed here. Frequent false alarms cause the care givers to relax their guard to the point of ignoring a real alarm condition when it occurs. Doors to patient rooms, closed to give the patient some respite from the noise in the corridor

and in other rooms, can attenuate a bedside monitor alarm such that it is not heard by nurses on the floor. Medical staff members, either through inattention, lack of knowledge, or intent, can disable the alarm function on a device. Mechanical or electrical failures could occur that render the alarm circuitry ineffective. Artifact present can be interpreted by a monitor as a real physiological signal and can cause a monitor to fail to alarm when the patient's physiological signal exceeds limits.

Mechanics

Device design should account for motion, maneuverability, and impact resistance. A transport stretcher that is difficult to maneuver can strain the muscles of the transporter. Devices should be made of materials with strength adequate for their intended purpose. Device mountings and supports must have adequate load-bearing characteristics. Inadequately secured physiological monitors have fallen onto patients from overhead shelves. Latches, locks, protective covers, and restraints must be designed with appropriate attention to human factors and conditions of use to prevent inadvertent opening and loss of protection from falls or tampering. Patients of all ages have died from asphyxiation after becoming entrapped by bed rails. Connectors and couplings abound in the hospital environment. Blood line disconnections have resulted in deadly air embolisms; airway disconnections have resulted in death and brain damage from oxygen deprivation; and reversed connections to anesthesia vaporizers have resulted in deadly anesthetic overdoses.

Operator

Lack of knowledge concerning the operation of a device is the most common failing of the operator. A safety program must recognize that the rapid introduction and wide proliferation of complex medical device technology taxes the operator's ability to remain competent and must assure that adequate educational programs are in place to address this need. Educational programs and credentialing requirements minimize user error caused by ignorance of device operation.

Failure to correctly diagnose is a claim made increasingly more frequently in malpractice actions. Risks associated with diagnostic medical devices such as ECG telemetry systems or clinical laboratory analyzers often relate to device set-up, calibration, and operation and data gathering, manipulation, display, storage, and retrieval. The harm that people (e.g., doctors, nurses, technologists), can do to the patient can relate directly to the presence, quality, and limitations of diagnostic data obtained by the device (15).

Errors in the operator's use of devices are more likely when manufacturers do not provide instructions or provide ambiguous or misleading instructions, labels, indicators, or controls on the device. Manufacturers have a duty to communicate word to their customers of upgrades and product improvements that address design deficiencies in the products that they have sold.

The manufacturer may contribute substantially to the operator's failure by not understanding the limitations of

the operator and designing a device without adequately addressing human factors engineering. The FDA recognizing the important role sound human factors engineering plays in safe medical device design has spearheaded efforts to develop standards and guidelines in this area (16). The degree to which human factor problems contribute to safety risks and incidents is indeterminate largely because those reporting device problems often lack the understanding of how faulty medical devices and instrumentation design contributes to user errors (17).

Drug and medication errors constitute one of the major sources of hospital incident reports. Safety problems include adverse drug reactions and inappropriate dosage, drug, frequency of administration, and route of administration.

Economic pressures influencing staffing patterns and work requirements result in understaffed and overworked operators who do not adhere to policies and procedures, but rather take more expedient, albeit more risky, avenues. For example, an ECG telemetry central station, which by the manufacturer's instructions and hospital policy should be monitored constantly, may go unattended because of competing demands for personnel resources. A study of anesthesia claims over an 8-year period revealed that 14% of the claims alleged failure to monitor (18). Personnel who do not adhere to hospital or departmental policies and procedures pose safety problems for the patient. Policies and procedures also address such issues as patient care procedures or surgical procedures.

Facility

The hospital's physical plant and its facilities can have a substantial effect upon safety. Some of the major facilities of a typical hospital are listed below followed by some examples of how these facilities can affect patient and personnel safety (19):

- Heating, ventilation, and air conditioning (HVAC)
- Medical gases
- Water
- Electric power
- Sanitation systems
- Transport and space

Heating, Ventilation, and Air Conditioning

Air pollution can adversely affect the compressed air supply that is needed to provide respiratory support and to power pneumatic devices. Inadequate ventilation has resulted in hazardous concentrations of ethylene oxide gas used in the sterilization of heat-labile devices. Failure to control relative humidity has resulted in water condensing on wrappings of sterile surgical instruments and thus breaching the sterility barrier.

Medical Gases

Switched oxygen and nitrous oxide supply pipelines have caused patient deaths. Poorly designed and maintained medical gas and vacuum distribution systems can fail to

provide adequate pressures and flows for the proper operation of such devices as suction machines, pneumatically powered surgical instruments, ventilators, and anesthesia machines.

Water

A hospital must have an adequate supply of water (20). Because it is in direct contact (through a membrane) with a patient's blood, the water used in hemodialysis must be monitored and its components controlled. Numerous adverse reactions from untreated water have occurred (21). Components that can affect the dialysis treatment have been identified and include insoluble particles, soluble organics and inorganics, heavy metals, bacteria, and pyrogens.

Electrical Power

The performance of sensitive electronic medical devices can be adversely affected by irregularities in electrical power distribution systems. Line voltage variations, line transients, and interruption of power can all result in harm to the patient by adversely affecting the performance of such devices as physiological monitors, ventilators, electrosurgical units, and clinical laboratory analyzers.

Sanitation Systems

The safe disposal of solid, liquid, and gas waste constitutes a significant environmental concern to the hospital, its patients and personnel, and the community in which the hospital is situated (22). It is common for clinical personnel to be stuck by needles in the course of either injecting a patient or disposing of a used needle. Housekeeping personnel are victims of improperly discarded needles, scalpels, broken glass, and biohazardous and radioactive waste. Chronic low level exposure to inhalation anesthetics is associated with spontaneous miscarriage, liver disease, cancer, and other physiological disorders.

Hepatitis C, HIV, SARS, influenza, and other infection agents pose threats to patients and personnel who come in contact with infected patients or their bodily fluids and waste. To add to the problem, infectious agents have developed into virulent forms that are resistant to standard antibiotics.

Transport and Space

Elevators, escalators, doors, stairwells, and staff and patient care areas affect the level of hospital safety. A patient room that is too small results in crowding of medical devices, personnel, and patient. Crowding restricts rapid access to necessary devices and instruments and interferes with access to the patient. Poorly maintained and designed elevators delay a caregiver's access to a patient. Fire doors propped open or emergency exit doors propped shut pose serious hazards in the event of fire. Fire is particularly hazardous in an environment in which an individual's ability to evacuate is compromised by illness or disability. A closed door to a patient room may afford the patient some quiet from the noise from the outside, but it can also attenuate a bedside alarm to make it inaudible to the staff.

Environment

Those elements within a hospital environment that can influence the safe operation of a medical device include such things as medical and nonmedical devices that are not directly involved in a patient's care, people, electromagnetic interference, cleanliness, psychological factors, biomechanics, and natural and unnatural disasters. The environment in which medical devices are used has grown beyond the hospital to now include emergency vehicles, ambulatory care facilities, long-term care facilities, and the home. The hospital, to the extent that it furnishes devices for patient use, must assume the responsibility for ensuring that these devices are safe and effective.

Device-Device Interaction

The hospital has become a complex electrical environment where incompatibility can exist between medical devices (24). For example, electrosurgical units have obscured electrocardiographic traces on cardiac monitors, defibrillator discharges have damaged internal pacemakers, and electrical transients have modified patient ventilator rates.

Nonmedical Devices

An ever-increasing variety and number of nonmedical devices are often requested to be brought into hospitals by patients. Some of these devices include cell phones, space heaters, electric heating pads, televisions, electric shavers, electric guitars, and video games. A hospital safety program must specify what is not allowed to be brought into the hospital or into specific areas of the hospital, must detail the means of enforcement, and give guidance on making exceptions in unusual circumstances. Policies and procedures will vary with the institution, medical condition of the patient, and inspection capabilities of the hospital's clinical engineering resources.

Hospital staff often desire to bring in personal-use devices such as fans, electric space heaters, radios, and hot plates. Office equipment such as personal computers and radios also may be a source of risk and should be considered in a comprehensive safety program. Equipment used by housekeeping and plant engineers such as floor buffers, vacuum cleaners, and power tools are a source of risk.

People

Besides the operator described above (i.e., the patient's nurses, doctors, and technicians), other people in the patient's environment can pose safety risks. These include other medical staff not directly involved in the patient's care; other patients; nonmedical hospital personnel (e.g., housekeeping, maintenance, administrators, and security personnel); and visitors, salespeople, and interlopers.

Visitors have altered controls on life support equipment to the detriment of the patient. Housekeepers have disconnected life-support ventilators from wall outlets in order to plug in floor sweepers. Premeditated or nonpremeditated attack either on a patient or by a patient is a possibility. Abduction of infants or elopement of mentally deranged patients occurs.

Poor communication has an adverse effect upon a patient's treatment. For example, a care giver's order for a test or request for a report that goes unheeded can be detrimental. Not communicating what has already been done for the patient may at best result in unnecessary and costly repeat procedures or at worse the repeat administration of medications and resulting overdose.

Electromagnetic Interference

Electromagnetic interference (EMI) is the disruption of the performance of one device by electrical energy emitted by another device or by another source of radiation such as a radio or television transmitter. EMI is becoming more of a problem as the number of electronic devices in the hospital increases (24,25). The utilization of an ever increasing amount of the electromagnetic spectrum and the proliferation of devices that are intended and unintended emitters and receivers of electromagnetic energy require that hospitals pay due attention to compatibility and interference (26).

The high strength magnetic fields about MRI units have drawn in from the surrounding environment ferrous metallic objects resulting in death and injury when these objects collide with great force against patients and personnel.

Cleanliness

Infection is a major complication of invasive arterial, central venous, and pulmonary artery monitoring. Nosocomial (hospital acquired) infection can be spread by way of medical devices that have not been properly handled, cleaned, disinfected, or sterilized (27). Vermin, animals, fomites (bedding, clothes), food, and people are other vehicles for the spread of nosocomial infection. Dirt and dust can contaminate wound sites, can damage delicate electronic instrumentation and computer data storage devices, and can harbor pathogenic organisms.

Psychological Factors

Periontogenic (having its genesis in the surroundings) illness afflicts patients and staff. Such illness has its roots in the stresses engendered by the high technology surroundings and includes such syndromes as intensive care unit (ICU) psychosis, the adverse psychological impact of the ICU setting. Dehumanization of patients can lead to such neuroses and psychoses as depression, denial, and dependency.

Biomechanics

Thousands of patients each year are injured in slips and falls while attempting to get out of bed to make use of bathroom facilities. Back injuries to staff from lifting heavy equipment or positioning patients are not uncommon occupational injuries.

Natural and Unnatural Disasters

Major safety problems are associated with a wide range of natural disasters such as tornado, hurricane, and earth-

quake and unnatural disasters such as terrorism, and nuclear power plant meltdown. Disaster planning will minimize loss in these circumstances (28).

Patient

Failure of patients to exercise their duties and responsibilities can lead to their own injuries. Patients can contribute to their own injury by failing to use reasonable care in attending to themselves. Such examples include deliberate removal or tampering with support equipment, failure to follow instructions to remain in bed, tripping or falling over obvious obstacles in hospitals, and introduction of unauthorized items into the environment of the hospital.

Systems Analysis

The interaction of the above five components (patient, device, facility, operator, and environment) can be demonstrated by considering the example shown in Fig. 1. A neonate (the patient) is in an infant radiant warmer (the device) electrically powered from the wall receptacle (the facility). The nurse (the operator) prepares to connect the skin temperature sensor to the warmer's control panel. A nearby examination lamp and external influences of a bright sun and electromagnetic waves emitted by a local station are shown (the environment). All five elements of the system can interact and give rise to a situation resulting in patient injury. For example, electromagnetic radiation can interfere with the operation of the radiant warmer's heater control. The nearby examination lamp can add to the heat transfer to the neonate leading to hyperthermia. The skin temperature sensor could be incompatible with the temperature monitoring system, but its connector could permit it to be connected to the system only to give false reading and result in excessive radiant heat output. Interruption of electric power because of a damaged electrical receptacle could result in loss of heater output and subsequent hypothermia.

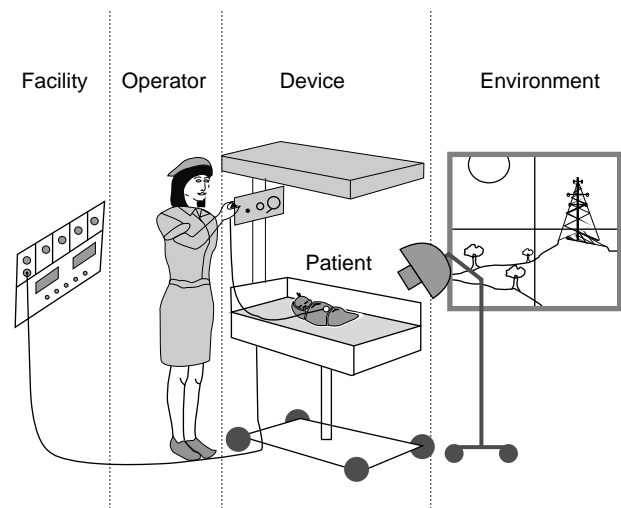


Figure 1. Systems approach to medical device safety.

TOTAL HOSPITAL SAFETY PROGRAM

As seen from the preceding section, the issue of total hospital safety encompasses human elements, machine elements, and environmental elements, all of which must be considered in a total hospital safety program (30).

JCAHO Recommendations

The *Comprehensive Accreditation Manual for Hospitals* (CAMH) (30) published by the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) includes a chapter entitled, “*Management of the Environment of Care*”. The standards contained in this article require hospitals to manage seven areas by the development and implantation of appropriate plans:

- Safety
- Security
- Hazardous materials and waste
- Emergency
- Life safety
- Medical equipment
- Utility systems

Safety Committee

The safety management plan requires a hospital to identify “*processes for examining safety issues by appropriate hospital representatives*”, that is, the safety committee, now often called the environment of care committee. The JCAHO refers to the safety committee as “*a multidisciplinary improvement team*” and defines its activities and responsibilities as follows:

- Safety issues are analyzed in a timely manner.
- Recommendations are developed and approved.
- Safety issues are communicated to the leaders of the hospital, individuals responsible for performance-improvement activities, and when appropriate, to relevant components of the hospital wide patient safety program.
- Recommendations for one or more performance improvement activities are communicated at least annually to the hospital’s leaders.

The safety committee serves as a forum for reports from the seven environment of care areas. For example, the safety committee will receive reports as specified in the hospital’s medical equipment management plan, including performance monitoring data and incident investigation reports. Serving as a reporting forum is an essential part of the safety committee’s role. However, the great value of the safety committee is found in its ability to subject these reports to multidisciplinary analysis. Careful analysis from a representative range of perspectives will produce communications, recommendations, and performance improvement activities that enhance safety throughout the hospital. Hospitals may assign managers to oversee

the implementation of the management plans in the above environment of care areas.

Cooperation and Communication

The hospital safety program requires the conscientious participation and work of a wide range of committees, departments, and individuals with full cooperation and support of the hospital’s administration. It involves a set of policies and procedures that cut through departmental barriers. Interdepartmental cooperation, involving virtually every clinical and support service, is imperative to adequately address the wide range of issues that bear upon the safety of patients and personnel. Hospital safety programs are inherently difficult to implement as a consequence of the diversity of their component parts.

A total hospital safety program should include patient and employee, environmental, and medical device safety. Hospital-operated and physician-directed safety programs should be linked in a single system. The hospital and medical staff should pool their resources and direct all patient safety information to a central point. At that point, data analysis can identify problem areas before compensable events occur. The components of a total hospital safety program that are actively involved in hospital safety issues are committees, departments, and administration.

Committees

The committee is a means by which formal lines of communication can be exercised. A hospital has various committees that relate in some fashion to an overall hospital safety committee (environment of care committee). Each committee necessarily deals for the most part with rather narrow issues. Those committees that have the strongest influence on overall hospital safety are listed in Table 1. No two hospitals are alike and the names of the committees may vary from one institution to the next.

Participation of key individuals, representing the entire spectrum of services provided within the hospital, in the work of the safety committee as well as in the work of the other committees listed is an integral component of the hospital safety program. Formal communication channels must be created between these committees and the safety committee. Committee attendance is an aid to interdepartmental communication. Transfer of information in an informal setting in a face-to-face meeting before or after a committee meeting can often be more effective, rapid, and accurate than written or telephone communications.

Table 1. Committees Concerned with Hospital Safety

Radiation safety	Specialty medical services	Infection control
Quality assurance	Emergency resuscitation	Utilization review
Standards and evaluation	Medical board	Capital equipment
Continuing medical education	Operating room	Credentials and privileges

Table 2. Departments, Offices, and Authorities Concerned with Hospital Safety

Risk management	Staff development	Materials management
Plant and facilities engineering	Environmental health and safety	Fire marshall
Medical departments	Quality assurance	Information services
Radiation safety	Infection control	Clinical engineering

Departments

The departments, offices, and authorities that play important roles in the hospital safety program are shown in Table 2.

Risk management is a key element in any hospital safety program (31). It has become not only a moral, but a financial necessity to provide a safe environment for the staff and patients. It is the risk manager who must make recommendations, in the face of cost containment efforts, on the investment of capital in monitoring systems and modern equipment as a means of reducing risk. Risk management can determine a course of action aimed at reducing risk through, for example, increased surveillance of the patient by appropriate monitoring devices such as electrocardiographs, capnometers, blood pressure monitors, and pulse oximeters. Should an analysis show that the main problems lie in human error, clinical engineering and staff development might implement an educational program as a response. For example, human error, rather than equipment failure, appears to be the primary factor in most preventable anesthesia mishaps. Increased risk management initiatives can reduce these adverse incidents. Cost-effective anesthetic practices actually require using the most up-to-date and well-maintained equipment. The risk management component is central to controlling the costs of safety. Care must be taken to deploy risk management resources judiciously. A disproportionately large amount of attention has been focused on some relatively minor safety issues, such as electrical safety, while by comparison few resources have been commanded to address issues that pose greater risks to patient safety, such as slips and falls (32).

Materials management is generally responsible for the acquisition and distribution to patient floors of a wide variety of device accessories and disposable products. Prepurchase evaluation of these items should not be overlooked. Clinical engineering is often in a position to lend its expertise in device testing to the evaluation of new products proposed for addition to the hospital's standing inventory.

Environmental health and safety is concerned with such issues as fire safety; air quality; spillage, handling, and storage of chemicals and unidentified substances; and noise abatement.

Radiation safety is concerned with the protection of patients and personnel from the effects of ionizing radiation (33). The primary aim of radiological protection is to provide an appropriate standard of protection for humans without unduly limiting the beneficial practices giving rise

to radiation exposure according to the International Commission on Radiological Protection (ICRP) (34). The ICRP system of protection is based on the following three objectives:

1. To prohibit radiation exposure to individuals unless a benefit to the exposed individuals and to the society can be demonstrated, and be sufficient to balance any detriment.
2. To provide adequate protection to maximize the net benefit, taking into consideration economic and social factors.
3. To limit the dose (other than from medical exposures) received by individuals as a result of all uses of radiation.

Infection control identifies the means of growth, reproduction, and transmission of pathogenic organisms. Infection control advises on methods for eliminating or controlling the reservoirs of microorganisms, on appropriate sterilization, disinfection, and sanitation techniques, and on disturbing or eliminating the means of transmission. Where this means of transmission involves the medical device, close cooperation with clinical engineering is advisable.

Staff development has a substantial contribution to make in training staff and in periodically reviewing competence levels in device use and patient treatment techniques. Staff development typically develops suitable curricula with the assistance of clinical engineering.

Plant and facilities engineering maintains the facility, ensures satisfactory heating, ventilation and air conditioning, electric power, water supply, medical gases and vacuum, and sanitation. Facilities engineering keeps abreast of the building codes pertaining to construction and renovation (35).

The *medical departments* constantly review the quality of the care administered and remain alert to indicators of inappropriate diagnoses or therapeutic intervention. The medical staff's quality assurance programs are coordinated with each other and integrated with hospital programs under the office of the safety director.

Information technology departments are responsible for the computerized acquisition, storage, and retrieval of patient information of both an administrative and clinical nature. Timely, accurate, and conveniently displayed information reduces medical errors and the risks associated with providing treatment on the basis of clinical laboratory and other diagnostic data.

Administration

The hospital administration must support the safety program by allocating adequate personnel and financial resources and by enforcing compliance with policies and procedures and assigning responsibility to program implementation to a hospital safety officer.

Safety Officer

The key individual in a hospital safety program is the hospital safety officer. This position should be held by a

full-time hospital employee, especially in large, tertiary care, teaching hospitals. For small hospitals, community hospitals, and surgical centers the responsibilities of the safety officer's position can be exercised by an administrative staff member. The safety officer develops the administrative policies and procedures to address all aspects of safety within the hospital. Each department will have its own particular policies and procedures governing safety as will multidisciplinary committees.

Performance Improvement

A fundamental concept underlying all JCAHO standards is *performance improvement*, which is a continuous process of monitoring and evaluating performance to identify and realize opportunities for improvement. The performance improvement cycle links the safety committee, the safety officer, and the managers of the seven environment of care management plans into a framework for coordinating the many separate activities that constitute a hospital-wide safety program.

SAFETY PROGRAM: MEDICAL DEVICES AND INSTRUMENTATION

The focus in the past of most hospital safety programs on employee and device safety has changed to include a broader range of patient safety issues. The hospital safety program must address all the risks to which patients and staff are exposed. The duties and responsibilities of the hospital safety officer should reflect this broader safety perspective. Clinical engineering contributes to the hospital safety program principally through its management of medical device technology. A broadening of the clinical engineering department's perspective could be required in some cases, particularly in a hospital in which undue attention is placed on electrical safety issues.

Clinical Engineering

A medical device safety program can be administered in several ways: through an in-house clinical engineering (CE) department, either centralized and full service or departmentalized and partial service, through outside service provided by manufacturers' programs, or through an independent service organization (ISO) (36). The hospital personnel required to support such a program will range from one part-time worker when outside services are employed to 30–40 full-time workers when, for example, an in-house program is utilized in a major medical center (38).

Clinical engineering departments are concerned primarily with medical device performance and safety. The responsibility for servicing nonmedical devices and medical and nonmedical facilities may rest within or outside of clinical engineering; however, of overriding importance are the concepts that the hospital safety program should include all medical devices, medical device systems, utilities and facilities that may directly affect the safety of hospital patients or staff, and that the hospital should have in place a system that addresses all of these safety

issues. The clinical engineering department's responsibilities with regard to the hospital safety program should include the implementation and execution of a medical device technology management program.

Technology Management

Medical device technology management is the sound foundation upon which any safety program must be built. Technology management is the management of medical devices throughout their useful life and includes selection and acquisition and eventual replacement (38). Elements of technology management follow:

- Strategic planning (technology assessment)
- Acquisition (specification, evaluation, purchasing)
- Utilization and asset management (impact analysis)
- Medical device service (inspection, maintenance, repair and modification)
- Replacement planning

When a technology management program is run effectively, a hospital can be assured that the need for all medical devices has been ascertained; that medical devices have been properly evaluated, selected, purchased, inspected, and calibrated for optimal performance and safety; that all users of the devices are appropriately trained; that the devices will be maintained on a regular basis; that service will be provided promptly so that patient care will not be disrupted; that a system of tracking and identifying every medical device will be in place; that the device is retired from service at the appropriate time; and that device replacement planning ensures continuity of technological support. Documentation of all elements of the system is essential. A deficiency in any one of the above elements can increase the exposure of a hospital patient or employee to injury. From technology management flow all other aspects of a safety program.

Strategic Planning

The first step in minimizing the risk of medical device technology is strategic planning and technology assessment, determining what services will be provided and what technologies are appropriate for delivering those services.

Hospitals need to monitor changes in technology to enable sound strategic planning. For example, examination of the effects of new therapeutic or diagnostic modalities, such as magnetism in the case of MRI, would allow the hospital to circumvent incompatible or health-threatening situations. The hospital should maintain ready access to technological assessment reports such as those published by the National Center for Technology Assessment. A *chief technology officer* may be warranted for large, university-affiliated teaching hospitals. The acquisition of appropriate medical device technologies can be aided by the use of such tools as the Essential Health Technology Package developed by the World Health Organization (39).

Acquisition

Acquisition of adequate medical device technologies starts with a statement of clinical needs. Detailed specifications

are then written, accurately reflecting the needs expressed by the clinical staff. Prospective vendors are then selected and invited to demonstrate their products. Clinical and laboratory evaluation of these products will determine suitability to the hospital's needs and adherence to safety and performance standards. The evaluation will often involve a multidisciplinary approach so that all aspects of the device (e.g., life cycle cost analysis, clinical efficacy, safety, and engineering design) can be evaluated. A hospital standards and evaluation committee is a good mechanism to control the acquisition of a broad range of medical devices from infusion pumps to gauze pads. Purchasing documents should contain device performance specifications and conditions of sale such as withholding payment until all in-coming acceptance inspections are complete and the device is installed and operational. Other conditions include provision of operator and service manuals, vendor training, and warranty periods.

Utilization and Asset Management

At the initial (in-coming) inspection of a device entering the hospital, a record of that device (asset) is entered into the management system. Generally, affixing an asset management control number to the device enables the identification of that particular device and allows a computerized medical device management system to store and retrieve device histories. Computerized Medical Device Management Systems greatly enhance the acquisition and analysis of data and thus improve the management process (40). The unique control number points to a file containing information about the device that is essential to management of that resource. Such information as generic name of the device, manufacturer, location, and phone number of the manufacturer or authorized service representative, applicable inspection procedure to be used during scheduled preventive maintenance, and cost of the device are generally included in the medical device data base. The management system linked to a *medical device recall system* (see below) enables rapid determination of whether or not the hospital possesses devices that are subject to a recall.

Asset management data of the acquired medical device technology can support utilization and impact analysis enabling the hospital to reap maximum economic benefit (41). Improvements in utilization can positively affect a patient's care by maximizing availability of diagnostic and therapeutic services.

Medical Device Service

Periodic inspection, maintenance and repair are essential elements of medical device safety. Whether these functions are carried out by an in-house department or by an outside service organization, it is crucial that the service be performed in a timely manner so that there is negligible impact on the delivery of care for want of the necessary device or instrument. Several guides to developing maintenance and repair programs are available (42). Techniques for obtaining maximum support from vendors and outside service providers have been described (43).

Replacement Planning

All medical devices reach the point in their lives where replacement becomes necessary because of decreased reliability, increased downtime, safety issues, compromised care, increased operating costs, changing regulations or, simply, obsolescence. Healthcare organizations have limited funding for capital purchases with many under strict budgeting guidelines. Replacement based on anecdotal and subjective statements and politics create havoc related to finances, morale, and operations. Clinical engineering departments are impacted if they do not have a replacement plan when major repairs occur in older equipment.

The ideal healthcare technology replacement planning system would be facility-wide covering all clinical equipment, utilize accurate, objective data for analysis, be flexible enough to incorporate nonequipment factors, and be futuristic by including strategic planning related to clinical and marketplace trends and hospital strategic initiatives related to technology. The plan should encompass many factors related to cost benefit, safety, support, standardization, and clinical benefit (44).

Medical Device Recall System

A medical device recall system should be in place in all hospitals to receive recall notices, hazard bulletins, and safety alerts and to disseminate them to the appropriate authorities (45). The system should have a means for documenting any corrective actions required and implemented. The pharmacy typically operates a separate drug recall system. The *medical device safety officer* (see below) should manage the recall system. Close interdepartmental communication is required. All clinical departments, ancillaries, and support or service departments within the hospital are typically involved in medical device recall system. Virtually all types of medical devices are subject to recall: reusables or disposables, accessories and supplies, and capital equipment. Every medical device from tongue depressor to computed tomography (CT) scanner are included in the system.

Several publications contain hazard and recall notices including *FDA Enforcement Report*, *Biomedical Safety Standards Newsletter*, and *Health Devices Alerts* (see Reading List).

Incident–Accident Investigation and Reporting

The hospital safety program should encompass the mechanism to respond to an accident or incident in which a patient or staff member is injured (46). Incidents and accidents associated with medical devices and instrumentation fall within the purview of the clinical engineering department. The Safe Medical Devices Act of 1990 requires that the manufacturer of a medical device be notified within 10 business days if one of its devices contributed to death or serious injury. The manufacturer, under FDAs Mandatory Device Reporting (MDR) regulation, is required to report these incidents to the FDAs Center for Devices and Radiological Health. Manufacturers are required to report to the FDA when they receive information from any source which reasonably suggests that a

device they manufacture contributed to death or serious injury (47).

The clinical engineering department has the most familiarity with medical device performance and safety features and would be best qualified to serve as incident investigation coordinator of all medical device-related incidents. Many clinical engineers are ideally qualified by education, training, and experience to investigate incidents. When in-house clinical engineering expertise is unavailable or when an outside, unbiased investigation is required, hospitals make use of independent third-party investigators.

Medical Device Accident and Incident Investigation Methodology

Impoundment of devices after incidents is recommended. Control settings, dials, knobs, and switches should not be changed. Photographic evidence should be obtained. Disposable devices must be saved along with packaging material that can often be the only indication of lot number. Cleaning, disinfection, or sterilization of the item(s) involved could hinder further investigation and should be done only under the guidance of the incident coordinator. Policies should be written advising clinical staff to preserve all equipment involved in an incident, including disposable devices and associated packaging. Relative positions of devices in the room should be noted. Staff should record all relevant identifying information, such as the manufacturer of the device, date of use, location, serial or lot number, and hospital equipment control number if applicable.

Incidents and accidents often go unreported for fear of legal repercussions and from reluctance to admit that a procedural error may have been committed. An error in the use of a device may be related more to the human factors engineering failings of the device than to the carelessness of the operator. Only through aggressive reporting of medical device problems can the need for technological improvements be identified and the likelihood of similar incidents occurring be eliminated.

Quality Assurance

A means by which the CE department can measure the quality of its service is recommended (48). A departmental quality assurance plan should aim to identify problems or potential problems related to the safe, effective, and appropriate use of patient care equipment. Quality assurance indicators would include such items as the number of in-service education sessions, the number of device prepurchase evaluations, the average length of time required to restore damaged or defective equipment to working condition, the scope of department services, productivity, and proficiency. A computerized medical device management system is an invaluable aid in developing the necessary statistics for measuring departmental performance.

Information Technology

Clinical engineering and information technology are converging as an increasing number of medical devices become integrated into hospital information systems (49). Sensi-

tive patient information is contained in many medical devices and thus the security provisions of the Health Insurance Portability and Accountability Act (HIPAA) apply (50). Inappropriate use of medical information is a patient safety hazard. For example, reporting a negative HIV result to a person who is actually HIV positive could result in that person not taking appropriate steps to reduce the spread of the disease.

Education and Training

Developing a comprehensive education program in the use of medical devices is one of the principal goals of the hospital safety program (51). Changing legal attitudes in response to technological advances have resulted in additional responsibilities for nurses. Training to enhance the recognition of the day-to-day condition of devices, the need for in-service, and alertness to the signs of incipient device failures are essential components of a sound medical device safety program.

Tools

Professional organizations provide a good forum for the exchange of safety information. Keeping current with developments in medical device technology, patient and hospital safety, and regulatory issues by attendance at meetings of these organizations is recommended. The American College of Clinical Engineering (ACCE) is the only international professional organization dedicated exclusively to representing the interest of clinical engineers (53). Other biomedical engineering, facilities engineering, and medical instrumentation organizations support clinical engineering efforts at the national, regional, and local levels. These include the Engineering in Medicine and Biology Society of the Institute of Electrical and Electronics Engineers, the American Society of Hospital Engineers of the American Hospital Association, and the Association for the Advancement of Medical Instrumentation.

Numerous resource materials are available to assist the clinical engineering department in administering a safety and technology management program. The Reading List section includes texts that contain instructional material relating to technology management and patient safety. Teleconferences, workshops, and Internet also serve as good sources of information (53–55). Manufacturers, recognizing the need to provide informative material to the user to promote a better understanding of safety issues, often provide beneficial educational material.

Medical device standards, technology management recommendations, and safety guidelines are promulgated by several professional and trade organizations. Clinical engineers are encouraged to participate in standards-setting activities to contribute their informed user opinion on medical device performance and safety issues.

ENHANCING PATIENT SAFETY

The clinical engineer has many opportunities to design means to reduce the occurrence of medical errors and to

enhance patient safety through the application of engineering skills to solve problems of medical device selection, application, utilization, and safety (56,57). The proactive engineer will implement solutions to problems before they occur. Root-cause-analyses and failure modes and effects analysis (FMEA) will aid the engineer in determining what is wrong with the system and what needs to be fixed (58,59).

Medical Device Safety Officer

The medical device safety officer (MDSO) will one day be commonplace in hospitals. The clinical engineer is well suited to serve in this role and to manage the medical device safety program.

Postmarket Surveillance

Improved postmarket surveillance of medical device performance and safety is needed to provide more device problem feedback to the manufacturer, regulatory authorities, and to the hospital itself (60). Such feedback enables enhanced device design, revisions in standards, and improvements in educational programs with the ultimate goal of reduction in hazards. Improved methods are required to determine the prevalence, nature, and causes of medical device-related errors.

BIBLIOGRAPHY

- Kohn LT, Corrigan JM, Donaldson MS, editors. *To Err is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 2000.
- Trandel-Korenchuk DM, Trandel-Korenchuk KM. Legal forum, malpractice and preventive risk management. *Nurs Admin Q* 1983;7:75–80.
- Steel K, Gertman PM, Crescenzi C, Anderson J. Iatrogenic illness on a general medical service at a university hospital. *N Engl J Med* 1981;304:638–641.
- Abramson N. et al., Adverse occurrences in intensive care units. *JAMA, J Am Med Assoc* 1980;244:1582–1584.
- Enhancing Patient Safety: The Role of Clinical Engineering. Plymouth Meeting, PA: American College of Clinical Engineering; 2001.
- JCAHO. 2001. Revisions to Joint Commission standards in support of patient safety and medical/health care error reduction. Joint Commission on Accreditation of Healthcare Organizations. Oakbrook Terrace, IL.
- Shojania KG, Duncan BW, McDonald KM, Wachter RM, editors. *Making Health Care Safer: A Critical Analysis of Patient Safety Practice*. Evidence Report/Technology Assessment No. 43 (Prepared by the University of California at San Francisco—Stanford Evidence-Based Practice Center under Contract No. 290-97-0013), AHRQ Publication No. 01-E058, Rockville, MD. Agency for Healthcare Research and Quality; 2001. p 668.
- Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press; 2001.
- Bogner MS, editor. *Human Error in Medicine*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1994.
- Shepherd M. Systems approach to medical device safety. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
- Federal Food, Drug, and Cosmetic Act, As Amended October 1976. Washington, DC: U.S. Government Printing Office; 1976.
- Ridgway MG. The great debate on electrical safety—in retrospect. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
- Dyro JF, Tai S. Evaluation of ethylene oxide sterilizers. *Proc Ann Conf Eng Med Biol* 1976;18:499.
- Special Report! 2003 JCAHO National Patient Safety Goals: Practical Strategies and Helpful Solutions for Meeting These Goals. Joint Commission Resources 2003;3(1):1–12.
- Hyman WA. Risks associated with diagnostic devices. *J Clin Eng* 1986;11:273–278.
- ANSI/AAMI Human factors design process for medical devices; ANSI/AAMI HE74:2001. Arlington, VA: Association for the Advancement of Instrumentation; 2001.
- Hyman WA, Wangler V. Human factors: Environment. In: Dyro JF editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
- Bowyer EA. Anesthesia claims study identifies recurring areas of loss. *Forum* 1985;6(2):3–5.
- Hyndman B. Physical plant. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
- Hernández D, Hernández A. Water. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
- Klein E, Evaluation of Hemodialyzers and Dialysis Membranes, DHEW/NIH 77-1294. Bethesda, MD: U.S. Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health; 1977.
- Brito LF, Magagna D. Sanitation. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
- Miodownik S. Interactions between medical devices. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
- Paperman WD, David Y, Hibbetts J. Electromagnetic interference in the hospital. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
- Tan K-S, Hinberg I. Electromagnetic interference with medical devices: In vitro laboratory studies and electromagnetic compatibility standards. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
- Witters D, Campbell CA. Wireless medical telemetry: Addressing the interference issue and the new wireless medical telemetry service (WMTS). In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
- Ben-Zvi S, Gottlieb W. Medical instrumentation and nosocomial infection. *J Clin Eng* 1979;4:135–145.
- Epstein A, Harding GH. Disaster planning. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
- Baretich MF. Hospital safety programs. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
- Comprehensive Accreditation Manual for Hospitals. Oakbrook Terrace, Illinois: Joint Commission on Accreditation of Healthcare Organizations; 2001.
- Harding GH, Epstein A. Risk management. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
- Ridgway MG. Hospital environmental safety and the safety codes. *J Clin Eng* 1977;2:211–215.

33. Strzelczyk J. Radiation safety. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
34. International Commission on Radiological Protection. *Radiological Protection and Safety in Medicine*. Oxford: Pergamon Press; ICRP Publication 73; Ann ICRP 26(2); 1996.
35. Goodman G. Hospital facilities safety standards. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
36. Smithson P, Dickey D. Outsourcing clinical engineering service. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
37. Soller I. Clinical engineering in an academic medical center. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
38. David Y, Judd TM, Zambuto RP. Introduction to medical technology management practices. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
39. Heimann P, Issakov A, Kwankam Y. The essential healthcare technology package. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
40. Cohen T, Cram N. Computerized maintenance management systems. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
41. Seaman G. Industrial/management engineering in healthcare. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
42. Hertz E. Medical equipment management program and ANSI/AAMI EQ56. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
43. Dyro JF. Vendor and service management. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
44. Clark JT. Healthcare technology replacement planning. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
45. Hyman WA, Schlain LA. User problems and medical device recalls. *Med Instrum* 1986;20:14–16.
46. Dyro JF. Accident investigation. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
47. Winston FB. United States Food and Drug Administration. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
48. Autio D. Clinical engineering program indicators. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
49. Cohen T, Ward C. The Integration and Convergence of Medical and Information Technologies.
50. Grimes SL. Health Insurance Portability and Accountability Act (HIPAA). In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
51. Bauld TJ, Dyro JF, Grimes SL. Clinical engineering and nursing. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
52. Ott JC, Dyro JF. American College of Clinical Engineering. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
53. Wear JO, Levenson A. Distance education. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
54. Dyro JF, Judd TM, Wear JO. Advanced clinical engineering workshops. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
55. Lozano-Nieto A. Emerging technologies: Internet and interactive videoconferencing. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
56. Patalil B. Patient safety and the clinical engineer. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.
57. Cram N, Stephens J-P, Lessard C. The role of clinical engineers in reducing medical error. *J Clin Eng* 2004;29(1):33–35.
58. *Root Cause Analysis in Health Care: Tools and Techniques*. Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organizations; 2000.
59. Stalhandske E, DeRosier J, Patalil B, Gosbee J. How to make the most of failure mode and effects analysis. *Biomed Instr Tech* 2000;34(2):96–102.
60. Cheng M. Post-Market Surveillance and vigilance on medical devices. In: Dyro JF, editor. *The Clinical Engineering Handbook*. Burlington, MA: Elsevier; 2004.

Further Reading

- Wildavsky A. *Searching for Safety*. New Brunswick, NJ: Transaction Publishers; 1989.
- ACCE News. Plymouth Meeting, PA: American College of Clinical Engineering (bimonthly).
- Accreditation Manual for Hospitals. Chicago: Joint Commission on Accreditation of Healthcare Organizations (annual).
- Biomedical Instrumentation & Technology. Arlington, VA: Association for the Advancement of Medical Technology (bimonthly).
- Biomedical Safety & Standards. Hagerstown, MD: Lippincott Williams & Wilkins (semimonthly).
- Gendron FG. *Unexplained Patient Burns: Investigating Iatrogenic Injuries*. Brea, CA: Quest Publishing Co.; 1988.
- FDA Enforcement Report. Rockville, MD: U.S. Food and Drug Administration (weekly).
- Health Devices Alerts. Plymouth Meeting, PA: ECRI (semimonthly).
- Health Facilities Management. Chicago: Health Forum, Inc.
- Bronzino JD. *Management of Medical Technology*. Boston: Butterworth-Heinemann; 1992.
- Kolka JW, Link DM, Scott GG. *European Community Medical Device Directives: Certification, Quality Assurance and Liability*. Fairfax, VA: CEEM Information Services; 1992.
- Joint Commission Perspectives on Patient Safety, Oakbrook Terrace, IL: Joint Commission Resources (monthly).
- Journal of Clinical Engineering. Lippincott Williams & Wilkins, Hagerstown, MA (quarterly).
- Shojania KG, Duncan BW, McDonald KM, Wachter RM, editors. *Making Health Care Safer: A Critical Analysis of Patient Safety Practice*. Evidence Report/Technology Assessment No. 43, AHRQ Publication No. 01-E058, Rockville, MD: Agency for Healthcare Research and Quality, 2001.
- Geddes LA. *Medical Device Accidents and Illustrative Cases*. 2nd ed. Tucson: Lawyers & Judges Publishing Company; 2002.
- MD&DI/Medical Device & Diagnostic Industry, Los Angeles, CA: Canon Communications (monthly).
- Medical device user facility and manufacturer reporting, certification and registration. 21 CFR Part 803, July 31, 1996.
- NCPS Triage Cards for Root Cause Analysis. National Center for Patient Safety, Department of Veterans Affairs, 2001.
- National Electrical Code: NFPA 70. Quincy, MA: National Fire Protection Association, 2002.
- Fish RM, Geddes L. *Medical and Bioengineering Aspects of Electrical Injuries*. Tucson: Lawyers & Judges Publishing Company; 2003.

Zambuto RP, When worlds collide, Health Facilities Management Magazine. April 2004.

Reiser SJ, Anbar M, editors. The Machine at the Bedside: Strategies for Using Technology in Patient Care. New York: Cambridge University Press; 1984.

Shepherd's System for Medical Device Incident Investigation and Reporting. Shepherd & Baretich. Baretich Engineering, 2003.

Standard for Health Care Facilities: NFPA 99, Quincy, MA: National Fire Protection Association, 2002.

Dyro JF, editor. The Clinical Engineering Handbook. Burlington, MA: Elsevier; 2004.

The Healthcare Failure Mode Effect Analysis Process. National Center for Patient Safety, Department of Veterans Affairs, 2002.

See also CODES AND REGULATIONS: MEDICAL DEVICES; EQUIPMENT MAINTENANCE, BIOMEDICAL; GAS AND VACUUM SYSTEMS, CENTRALLY PIPED MEDICAL; IONIZING RADIATION, BIOLOGICAL EFFECTS OF; RADIATION PROTECTION INSTRUMENTATION; X-RAY QUALITY CONTROL PROGRAM.

SCAFFOLD MATERIALS, STERILIZATION

OF. See STERILIZATION OF BIOLOGIC SCAFFOLD MATERIALS.

SCAFFOLDS. See BIOMATERIALS: TISSUE ENGINEERING AND SCAFFOLDS.

SCANNING ELECTRON MICROSCOPY. See MICROSCOPY, ELECTRON.

SCANNING FORCE MICROSCOPY. See MICROSCOPY, SCANNING FORCE.

SCANNING TUNNELING MICROSCOPY. See MICROSCOPY, SCANNING TUNNELING.

SCOLIOSIS, BIOMECHANICS OF

I.A.F. STOKES
C.E. AUBIN
Polytechnique Montreal
Montreal, Quebec, Canada

INTRODUCTION

Scoliosis deformity occurs in the spine quite frequently, especially in growing children. In spite of the attention given to this deformity over the past century, still very little is known about its etiology or progression mechanisms.

There are a number of health related consequences of a progressive scoliosis. In large scoliosis deformities there is impairment of respiratory and cardiovascular function. If progressive, the deformity produces thoracic compromise, pain, and degenerative changes in the spine. In large curves that have been surgically fused, there may be iatrogenic problems secondary to multisegmental arthrodesis. Although scoliosis is often perceived as having a minimal cosmetic impact, it has psychosocial consequences (1–3).

TERMINOLOGY AND GENERAL MORPHOLOGY OF SCOLIOSIS

Geometry of Scoliosis

Scoliosis literally means a lateral curvature of the spine, but as a deformity it includes overall asymmetry and lateral deviation of the trunk, as well as axial rotation of the spine and trunk. The total geometric description of scoliosis is of paramount importance and merits more detailed characterization than can be given by a single-plane X ray. Three-dimensional (3D) measurements of the complete spinal and thoracic geometry require complex techniques of radiographic and surface measurement. Some measurement techniques currently used in research may achieve clinical application in the future.

Frontal, Sagittal and Transverse Planes Components of Spinal Deformity

Frontal Plane. Scoliosis is strictly defined as a lateral curvature of the spine although the scoliosis deformity is a 3D abnormality of the axial skeleton. Scoliotic curves have been described as being primary or secondary, major or minor, or structural or compensatory. The structural curve is defined as having asymmetry not reversible in lateral bending. Usually, a compensatory curve is present to maintain the alignment of the body so that the head is over the pelvis. When two curves of the same spine are defined as structural, the resultant S-shaped deformity is called a double structural curve. The apex of a scoliosis curve is at the vertebra or disk that is the most laterally deviated from the vertical axis passing through the sacrum (base of the spine) (4).

The Sagittal Plane. The mid-sagittal plane is the plane that divides the left side of the body from the right. The terms kyphosis and lordosis (Fig. 1) are used to describe curvatures that exist normally in the sagittal plane, with the prefixes hyper- or hypo- describing abnormal curvatures. Curvatures in this plane interact with the abnormal lateral curvature of the spine in scoliosis. This interaction is termed coupling (see below).

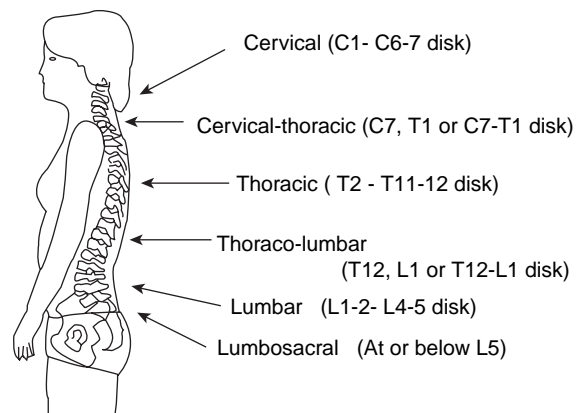


Figure 1. Lateral view of the spine defining regions of curvature.

Transverse Plane. Axial rotation of vertebrae accompanies the lateral deviation. The vertebral rotation can be measured from a frontal plane radiograph by grading the positions of the vertebral pedicles relative to the center of the vertebral body. Also, templates and formulas have been developed to convert these pedicle offsets into estimates of the rotation in degrees. The rotation can be measured more directly from transverse plane images obtained by computed tomography (CT) scanning or magnetic resonance imaging (MRI). The usual rotation of the vertebral bodies is to the convex side (5). Asymmetries of the vertebrae also develop, probably as a result of altered stress on these bones (6). The back surface rotates in the same sense, producing a “rib hump” (gibbosity). Back surface rotation correlates weakly with the scoliosis magnitude as measured by the Cobb angle (7), leading to the idea that curve magnitude could be measured topographically, without X rays. However, the back surface rotation is of lesser magnitude than the vertebral rotation (8), which in turn is of lesser magnitude than the Cobb angle, so this attenuation of the asymmetry as seen at the surface produces technical measurement difficulties.

Stereoradiography and stereophotogrammetry have been used to document the transverse plane deformity in patients with idiopathic scoliosis with simultaneous radiography to record the skeletal shape (9). The back surface rotation was reported to be greatest at the level of the apex of a lumbar scoliotic curve, but located as much as four vertebrae lower than the apex in thoracic and thoracolumbar curves (9). Thus the surface deformity is apparently augmented by the effects of the rib cage where the ribs point downward. In thoracic curves the radiographic rotation of vertebrae is reported to be greatest on average at the apex (defined as the most laterally deviated vertebra), but with a range of three vertebrae above to three below (9,10). In most of the lumbar curves the greatest surface rotation, the maximal vertebral rotation, and the apex coincides. The axial rotation of the surface had a variable relationship with the Cobb angle and in all cases was less than the Cobb angle (11). The amount of back surface rotation is less than the amount of axial rotation of the underlying vertebrae (9).

Coupling of Rotations. Coupling is the term used to describe the tendency of the spine while moving intentionally in one plane to produce secondary or coupled motion in another plane. Two kinds of coupling can be defined, orientation and kinematic. Orientation coupling describes the relationships between the linear or angular displacements of a single vertebra from its reference or nonscoliotic position. Kinematic coupling is defined as the association of two variables describing the relative motion of two vertebrae without regard to forces. Kinematic coupling has been described from *in vivo* measurements of normally curved spines (12–14). The relationship may be altered by adaptation of bone and soft tissue resulting from the scoliosis deformity.

Lateral bending of the healthy lumbar spine is accompanied by a rotation in the same direction as that seen in scoliosis (15); however, the coupling of axial rotation with lateral bending is in the opposite sense in the thoracic spine

(16). In normal spines, lumbar curves produced by lateral bending are associated with a vertebral body rotation toward the concavity of the curve (15). This is not the direction of the static rotation in a spine with scoliosis. The orientation coupling in scoliosis has the same direction as was noted in lateral bending of normal spines in extended positions.

In scoliosis, for each vertebra there are five orientation variables of interest: flexion–extension angle, lateral tilt angle, axial rotation angle, and lateral and anteroposterior deviations from the spinal axis. This leads to 10 possible coupling coefficients, which are ratios between pairs of these geometric variables. There is no reason *a priori* why orientation coupling in scoliosis and/or kinematic coupling coefficients in healthy spines should be related especially when ongoing growth complicates the relationship between different components of the deformity. However, after the end of growth, if a kinematic process governed by such coupling occurred during curve progression, then these relationships would be apparent.

Rib Cage and Back Surface Shape

The scoliosis deformity of the spine (curvatures and associated rotation) is associated with a complex pattern of rib cage deformity, evident as a thoracic rotation (rib hump). The rib hump becomes more subjectively evident in the back with scoliosis on forward bending of the trunk but its magnitude is in fact not increased (17). Thoracic rotation or rib hump can be measured on clinical examination in forward bending and recorded as the depression from the horizontal plane, or as the angular rotation from the horizontal (18) (Fig. 2). The balance of the torso or the extent to which the top of the spine deviates from being vertically above its base (sacrum). Several contacting and noncontacting (optical) methods are available to document the surface trunk deformity (reviewed below).

Classification of Scoliosis

Scoliosis is considered to exist if the Cobb angle is $>10^\circ$. The pattern of scoliosis can be classified by different criteria.

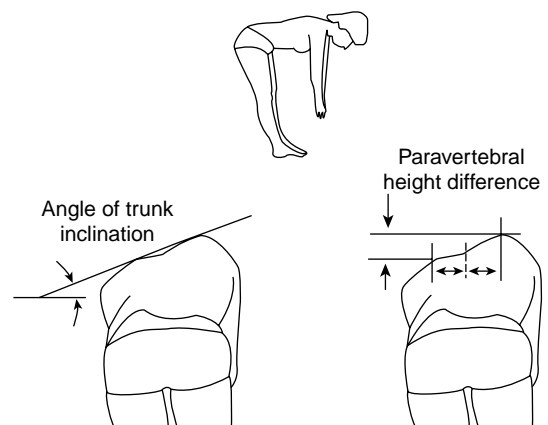


Figure 2. Rib hump seen in forward bend position.

Classification by Anatomical Region of the Curve. In classification of a curve by the location of the apex: the apex of a cervical scoliosis is between C1 and the C6–7 disk; cervicothoracic scoliosis between C7 and T1; thoracic scoliosis between T2 and the T11–T12 disk; thoracolumbar scoliosis between T12 and L1, lumbar scoliosis between the L1–L2 and L4–L5 disk spaces, and lumbosacral scoliosis below L5 (19).

Classification by Age. Scoliosis can be classified according to the skeletal age of the patient at diagnosis (not necessarily the same as the age of onset). Infantile scoliosis is diagnosed during the first 3 years of life, juvenile scoliosis between the ages of 4 and 10 years; adolescent scoliosis is a scoliosis diagnosed between ages 10 and 18 years, and adult scoliosis is diagnosed after age 18 years (19). A large percentage of infantile cases resolve spontaneously. These cases may be related to the position of the body in utero, but generally the likelihood of progression of the deformity is worse the earlier the age at which it develops, consistent with progression being associated with continued skeletal growth. Usually, skeletal maturation is complete after the age of 17 in females and after the age of 19 in males. For reasons that are unclear, in idiopathic scoliosis (81% of all scoliosis) the convexity is more common on the left side in lumbar curves and on the right side in thoracic curves.

Classification by Causation. In most instances, scoliosis can also be separated into one of the four major subgroups depending on the probable initial cause: idiopathic (spontaneous or unknown cause), neuromuscular (associated with disease of the neuromuscular system), congenital (associated with a prenatal developmental anomaly) (20), and iatrogenic (from radiation or thoracic surgery). Idiopathic scoliosis (for which no obvious cause is ascertained) represents 81% of the cases. There are biomechanical factors associated with scoliosis of any cause, since the deformity alters the biomechanics of the spine and trunk, and this in turn alters the processes of musculoskeletal growth, degeneration, and so on.

Classification by Curve Pattern. Classification systems are used to guide the management of scoliosis when a curve type can be consistently related to a different prognosis or management strategy. In surgical planning for the surgical management of adolescent idiopathic scoliosis, classification by radiographic measures is used to help decide on the extent of the arthrodesis. The King et al. classification (21) is probably still the most widely used in planning of spinal fusions, although it was originally developed to aid planning for Harrington instrumentation, that is seldom used now. It defines five thoracic scoliosis curve types, and an additional group called miscellaneous. The King et al. classification relies on subjective identification of the radiographic features that provide the measures used for classification. It may also require individual interpretation and memory of the classification criteria. As a result, there are numerous opportunities for variable implementation that produce inter- and intraobserver variability. A recent empirical study (22) of repeat curve-type classification

has demonstrated poor reliability. An algorithm (23) aims to overcome these problems by defining an objective classification procedure.

The Lenke (24) classification system was developed to provide a comprehensive and reliable means to categorize all surgical AIS curves, using postero anterior (PA), lateral, and side bending X-ray films. The three classification components are curve type (1–6), a lumbar spine modifier (A, B, C), and a sagittal thoracic modifier (–, N, +). The largest curve is considered to be the major curve, and minor curves are subdivided into structural and nonstructural types.

Axes and Coordinate Systems

The global coordinate system describes the positions and displacements of vertebrae relative to the whole body; regional coordinates systems employ a reference within a spinal curve or the entire spine, and local coordinate systems are fixed in a particular vertebra and aligned with its anatomic features are needed to describe motion between vertebrae (Fig. 3). The Scoliosis Research Society (SRS) (4,25) placed the origin at the center of the superior endplate of S1 for both global and spinal axis systems of patients with scoliosis. This is a commonly accepted convention. The ISO 2631 (VDI 2057) right-handed axis convention has x signifying anterior, y signifying left, and z signifying the cephalad direction. The global and spinal axis systems have the origin and sagittal plane defined by the pelvis, with the anterior superior iliac spines (ASIS) defining the transverse global (Y) direction. (This would normally be achieved by positioning the ASIS parallel to the X-ray film plane.) The other principal directions are aligned either with gravity (global system), or with spinal landmarks.

Plane of Maximum Curvature. The geometry of the spine can also be considered as a curved line in space with vertebrae positioned along this vertebral body line (4).

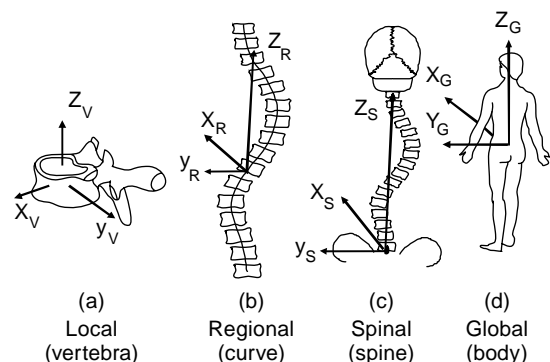


Figure 3. The hierarchy of four coordinate systems defining spinal geometry: (a) Local coordinates based on a vertebra. (b) Regional, curve-based coordinates defined by the end vertebrae of a curve or other spinal region. (c) Spinal coordinates, defined by the Z axis passing through the centers of the most caudal and cephalad vertebrae of the entire spine. (d) Global, whole-body-based coordinate system, with its origin at the base of the spine (S1), and the Z axis vertical (gravity line).

The vertebral body line in turn defines a plane of maximum curvature and best-fit plane (4,26), that lie in the mid-sagittal plane for an undeformed (symmetrical) spine, but are rotated from that plane when a scoliosis is present. In a straight, healthy spine the plane of maximal curvature is the sagittal plane, but in a scoliosis the maximum curvature is seen by viewing the spine from an intermediate angle that may be rotated from the sagittal plane by approximately the same number of degrees as the rotation of the apical vertebrae (27). In a purely geometrical sense, the trihedron axis system (28) provides the basis for 3D measurements, including the local curvature and the geometrical torsion of the vertebral body line. It also defines the osculating plane (local transverse plane), the tangent and normal directions, and the binormal. While providing a pure geometrical approach to defining spinal geometry (29), this approach has not been much used either in clinical or many research contexts.

Measurement of Curve Magnitude. The degree of the scoliosis deformity is most commonly measured clinically by the Cobb angle, which is the angle between lines drawn on the end plates of the most tilted vertebrae in the radiographic curve as shown in Fig. 4. The Cobb (30) measurement of a scoliosis is the sum of the angles of inclination in the frontal plane of the two end vertebrae. The end vertebrae are defined as those with the largest inclinations to the horizontal, and are typically close to the points of inflection of the curve. End vertebrae normally have no axial rotation. Reliability studies estimate the interobserver error of the Cobb measurement as 5° (23). The Cobb angle measurement summarizes curve magnitude, but ignores numerous important structural characteristics including the length of the curve (numbers of vertebrae in it), its flexibility, and the relative contributions of vertebral and discal wedging. While it is presently the major objective measurement used in planning management of patients with scoliosis, these shortcomings suggest that there is scope for improvements in clinical measurement.

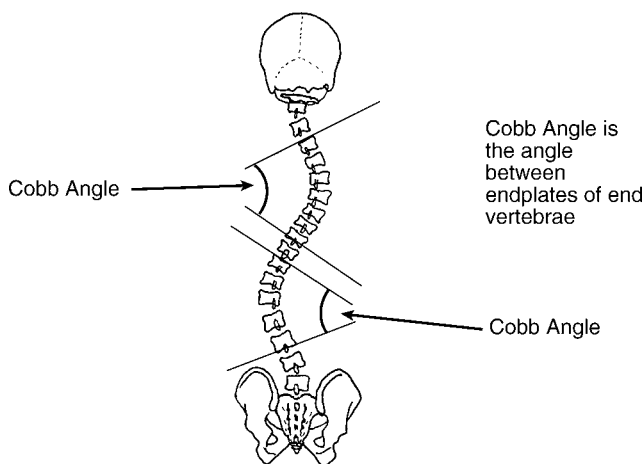


Figure 4. The Cobb angle as used to measure scoliosis curve magnitude. It is defined as the angle between the lines drawn on upper and lower end-vertebral endplates, where the end-vertebrae are those having the greatest inclination to the horizontal.

FUNCTIONAL ANATOMY

Spine: Vertebrae, Disks, Facet Joints

The structure and function of the spinal column depends on a complex interaction between the vertebral bodies, the articulations between their posterior elements (facet, or zygoapophyseal joints), interconnecting ligaments and intervertebral disks, the rib cage and the trunk musculature. In a scoliosis the vertebrae become wedged and twisted, the intervertebral disks become wedged, and there is a variable degree of adaptive changes in the soft tissues (ligaments, muscles, etc.) that probably increase with age. The ribs are more sharply angled on the convex side, and this asymmetry, together with the axial rotation component of the deformity produces the rib hump or gibbosity that is the most evident outward indication of the deformity. Surgical treatment of adult scoliosis is more difficult in part because of the adaptive changes and the typically greater curve magnitude and cardiothoracic compromise.

The Spinal Motion Segment

The spinal motion segment is a structural unit of the spine consisting of two vertebrae and the intervening soft tissues. It is a valuable tool in spinal biomechanics because it is the basic element of the spinal column. Once motion segment behavior is defined, the behavior of the whole spine can be represented by a series of fundamentally similar components. In the absence of scoliosis, the motion segment can be considered as being symmetrical about the sagittal plane, thus right and left lateral bends, axial rotation and shear are theoretically symmetrical. The presence of the posterior elements (zygoapophyseal or facet joints and the ligaments) introduces differences between the flexion and extension behaviors. The presence of these posterior structures also produces a posterior displacement (relative to the disk center) of the effective structural axis of the motion segment. The stiffness of the lumbar motion segments has been described by a stiffness matrix, and by an equivalent beam structure (31).

Ribcage and Costovertebral Articulations

Motion of the thoracic spine is connected to that of the rib cage by the costovertebral joints. The ribs articulate with the transverse processes, as well as with the vertebral bodies of the corresponding vertebra and that immediately cephalad. Motion of the ribs has been described as consisting of both pump-handle and bucket-handle motion. Although not well defined, these terms imply rotations about the horizontal (mediolateral axis) and an axis joining the ends of the ribs, respectively. The complex interactions in chest wall mechanics during breathing are difficult to explain in terms of muscular recruitments that involve the diaphragm, as well as the intercostal musculature and abdominal musculature. Rib deformation is part of the scoliosis deformity, thus complicating the progression of thoracic curves as well as their surgical treatment (32) where costoplasty is sometimes considered. The biomechanical consequences of costoplasty were analyzed biomechanically by Aubin and co-workers (33,34).

KINEMATICS OF THE SPINE AS IT RELATES TO SCOLIOSIS

Very little is known about how a scoliosis deformity specifically affects spinal kinematics, with little information either from cadaveric specimens or intraoperative measurements. Studies of gait (35) suggest that asymmetrical spinal axial rotation motion is the main kinematic consequence of a scoliosis. After spinal fusion, the overall motion of the spine is reduced as expected, but with reduced motion also in unfused segments (36).

Range of Motion, Coupled Motion, Intervertebral Stiffness Matrix

The intervertebral articulations have six degrees of freedom (three translations and three rotations) each of which has a measurable stiffness. Traditionally, the load-displacement characteristics of these joints have been described by a stiffness matrix (37). This stiffness matrix has off-diagonal terms, as well as diagonal terms. The off-diagonal terms resulting from this tendency of certain degrees of freedom to be associated with each other (especially axial rotation and lateral bending) has been referred to as coupling. The pattern of motion that occurs between two vertebrae depends on the combination of forces applied, and the axis of rotation is not fixed. It is only possible to define an instantaneous axis of rotation.

Most experimental motion segment stiffness data are limited as they do not include all six degrees of freedom, were obtained by inverting flexibility data, and were obtained without physiological levels of axial compressive force. Physiological axial compression has been observed to increase lumbar motion segment stiffness by a factor of 2 or more (31,38). Janevic et al. (38) reported that the stiffening of the motion segment with preload was approximately linear with preload magnitude. With 2200 N preload, rotational flexibility decreased on average 2.6 times, and shear flexibility 6.16 times, the effects being even greater at 4400 N preload. The stiffening of the motion segment with preload appears to result from nonlinearities in the intrinsic material properties of the annulus fibers, as well as the more easily visualized engagement of the facet joints (31).

METHODS TO DOCUMENT TRUNK DEFORMITY

Plane Radiography, and Derived Measures

The PA radiograph is normally used for documentation of scoliosis to show the lateral curvature and some indication of the axial rotation. The anteroposterior (AP) radiograph delivers higher radiation dose to breast tissue, sternum, thyroid, and ovaries (39), whereas the PA projection gives a higher dose to posterior structures, including the marrow of the spine. Radiographs used to monitor a scoliosis deformity can involve a cumulative skin entry dose of 2–3 rems per year during the growth years, or ~10 times the annual background dose; however, most of the body receives a much lower dose. Radiation exposure can be minimized by such techniques as the Ardran method (40), by using experienced technicians to minimize retakes, by

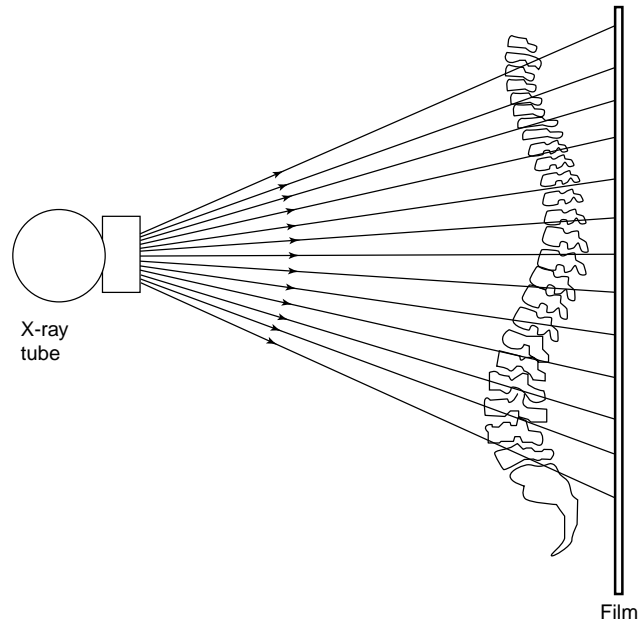


Figure 5. The distortion inherent in an X-ray image, due to the divergence of the X-ray beam relative to the object in the image, and the flat film plane.

proper collimation, and by proper monitoring of X-ray tube performance and filtration. Also, digital radiography with highly efficient X-ray detector arrays hold promise for further dose reductions relative to conventional passive systems with photographic films and intensifying screens.

Both PA and AP radiographs give a magnified distorted image because of divergence of the X-ray beam (Fig. 5). DeSmet et al. (41) compared the measurements of the Cobb angle using both AP and PA projections. There were small differences ($\sim 2^\circ$) in the lordotic lumbar region and kyphotic thoracic region. In the thoracolumbar region where the spine is more nearly vertical and the central X-ray beam is more perpendicular to the film, there was no difference between measurements in the PA and AP projections.

The frontal plane projections may also be used to estimate axial rotation of the vertebrae, based on the positions of the pedicles relative to the apparent center of the vertebral body (42). However, the exact relationship between rotation and this pedicle offset depends on the shape of vertebrae, which is a factor that can be taken into account by using statistical data for vertebral shape (43,44).

Multiview Radiography and 3D Reconstruction

In idiopathic scoliosis the standard frontal and lateral radiographs do not supply all the needed information to understand the 3D aspects of the deformities (e.g., intrinsic vertebral deformities, intervertebral disk wedging, spine torsion, rib cage and pelvis deformation, etc.). Computerized tomography and MRI can be used for 3D imaging, but these modalities have limitations. The image acquisition is normally done in the recumbent position (therefore not

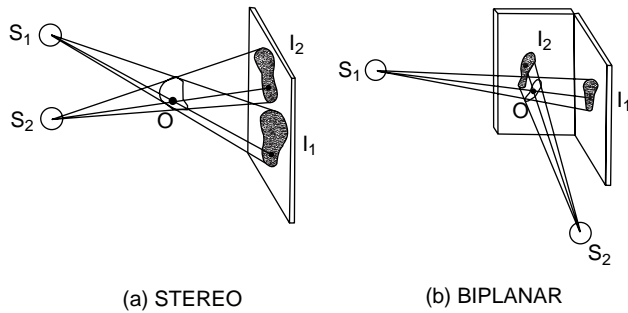


Figure 6. Principles of stereo and biplanar radiography. Stereophotogrammetry principles are used to identify the 3-D positions of object points. The location of the object points can be visualized as the intersections of lines joining the X-ray sources to the corresponding image points.

documenting the standing, functional position). The CT scans can require high X-ray dose of radiation, and both CT and MRI may be incompatible postoperatively with surgical metallic implants. The cost and the image acquisition time are considerable.

Stereo or biplanar techniques (with PA and lateral views, or PA and oblique views) permit 3D reconstruction of selected spinal landmarks using photogrammetric principles (Fig. 6). In the reconstruction, the position of a selected anatomical point visible in both views can be calculated in 3D space (45), using an algorithm such as direct linear transformation (DLT) (46). Dansereau and Stokes (46,47) developed an iterative method for line reconstruction that they applied to the 3D reconstruction of the rib cage based on the images of ribs as seen in PA and oblique radiographs. The average error of these reconstruction methods is ~3 mm, depending on the nature of the landmark (some are more precisely identified than others)(48,49).

These 3D reconstruction methods require calibration to take into account the relative positions of the X-ray sources and film planes. A calibration object with object points in known positions is normally employed. This process was substantially facilitated by the DLT method (46,47), and a recently described self-calibration algorithm (50) relies on point-correspondences between two views of the same scene, without using a calibration object. *A priori* knowledge of typically shaped vertebrae and the pelvis can be used to enhance the anatomical detail, employing a database of geometrical templates in an atlas derived from statistical anthropometrical data from the literature (51,52).

Tomography

Computerized tomography provides several measures of vertebral and thoracic rotation in the transverse plane (53). However, this method has technical difficulties because of the need to select the orientation of the image plane correctly: usually a global transverse plane is used that is not coincident with the local or regional planes in the deformed spine. While this technique can obtain rotation both at the vertebral level and in the trunk cross-

section, it is an expensive procedure, and the radiation dose can be high. It may be used in planning complex surgical procedures for spinal reconstruction after trauma or in cases of severe congenital spinal abnormalities.

Back Surface Topography

The topographical reconstruction of the human trunk has been developed in an attempt to improve documentation of the scoliotic deformity, to eliminate radiation exposure, and to document the external cosmetic aspects of the deformity. These methods may have potential clinical application in detection of scoliosis progression in serial back shape recordings using an automated surface shape measurement system (54). These noncontacting, optical techniques include Moiré fringe photography (55), laser scanning (56,57), and Raster photography (9,58). The principle behind the Moiré fringe technique is shown in diagrammatic form in Fig. 7. The interference patterns, or fringes, represent the surface height or topography of the surface upon which a least one set of lines is projected. Raster techniques involve projection of a structured light pattern (usually lines or dots) on to the body surface, and viewing from an oblique direction (Fig. 8). With suitable calibration, the distorted pattern of the projected lines or points can be converted into the three dimensional coordinates of surface points.

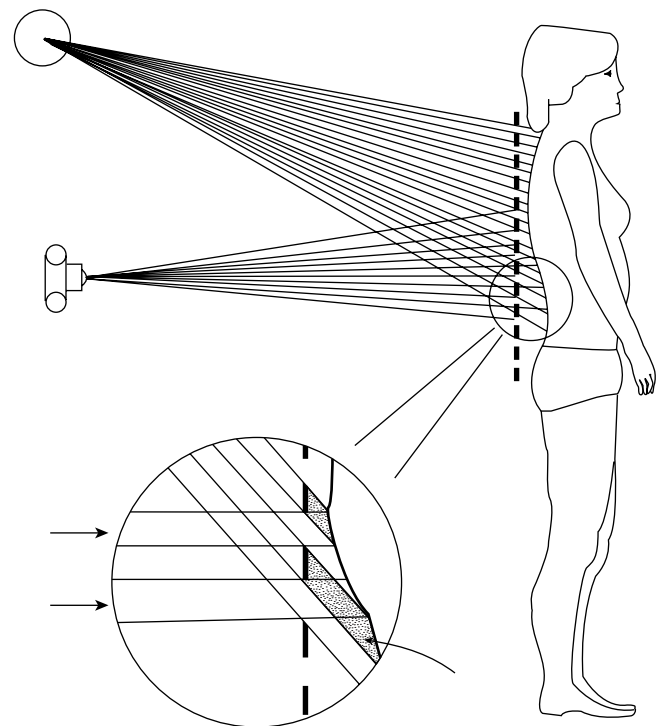


Figure 7. Principles of Moiré fringe photography for recording the shape of the back surface. The surface is viewed through a grid of parallel lines, through which also the surface is illuminated from a different angle. The Moiré fringe pattern results from the interference between the curved line shadows, viewed through the straight line grid.

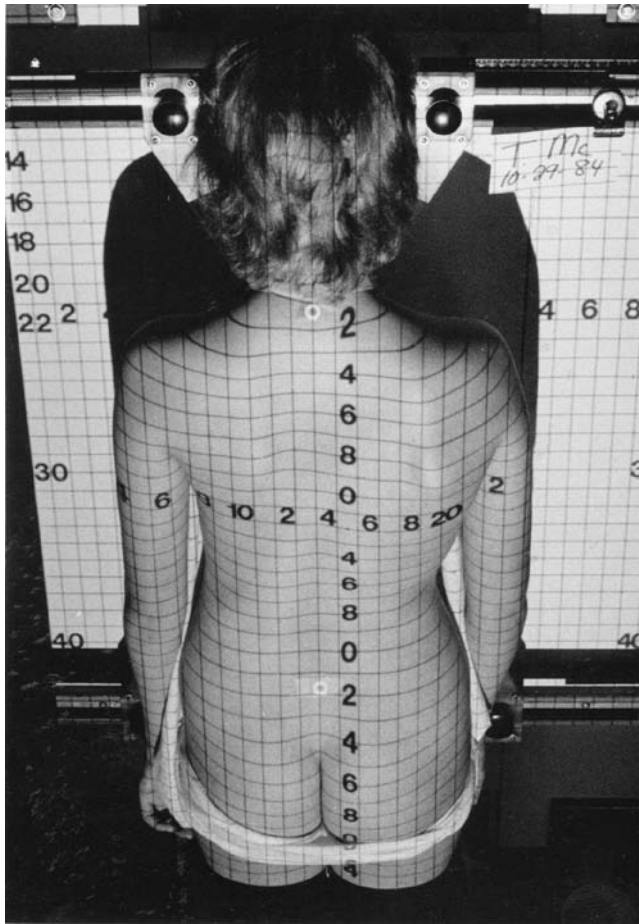


Figure 8. Example of raster stereophotograph of the back surface of a person with scoliosis. The back surface is illuminated by a square grating pattern of lines. These lines appear curved when viewed from a different angle. This system can be calibrated (by a priori knowledge of the geometry, or empirically by measuring the distortion of the grid when projected on to a surface of known shape. This measurement technique is suitable for automatic measurement, for example, by computer analysis of a digital image.

Recent developments in surface topography include the 360° reconstruction of the trunk to analyze the overall external asymmetry with laser scanning or range sensing (56,57,59) (Fig. 9), and its combination with a neural network approach to identifying the optimal relationship between body surface shape and spinal asymmetry.

THEORIES OF IDIOPATHIC SCOLIOSIS ETIOLOGY

The exact biomechanical contributions to etiology and progression of idiopathic scoliosis deformities are not well understood. Scoliosis consists of a lateral spinal curvature, together with transverse plane rotations of the spine and rib cage, and any complete theory concerning its etiology should be able to explain the originals of this complex geometry. In the absence of clear mechanism explaining

the etiology, it is considered that the causation is probably multifactorial (59,60).

Biomechanical Aspects

Several biomechanical factors have been invoked to explain the biomechanics of scoliosis. These include intervertebral motion coupling, spinal tethering and buckling, and mechanical influences on growth. While in some forms of scoliosis, the initiating causes are relatively clear (e.g., muscle imbalance or weakness in neuromuscular scoliosis), the etiology of idiopathic scoliosis remains obscure. Somerville (62) emphasized the importance of sagittal plane curvatures and demonstrated with an articulated model how a spine with tight posterior ligaments, which is then forced into flexion will become unstable and develop a rotated, scoliotic deformity.

Coupling. Note that both lateral bending of the spine and scoliotic deformities involve transverse plane rotation, but it appears unlikely that coupling of rotational motion in intervertebral segments controls the development of vertebral rotation in scoliosis. The normal kinematic relationships between lateral bending and axial rotation produce a different spinal shape than that seen in scoliosis (16,63,64). In particular these changes do not explain the pattern of deformity that develops in scoliosis with maximal vertebral rotation at the curve apex. Spinal tethering by posterior structures of the spine (62) has also been invoked to explain the spinal shape in scoliosis and its etiology. There are several parts to this theory. Tethering is thought to prevent flexion of the spine and lead to a hypokyphotic or lordotic shape, which then has a greater tendency to instability. Secondly, the tether is thought to maintain a straighter alignment of the posterior elements than of the vertebral bodies, thus the anterior part of the spine (the vertebral bodies) become more laterally deviated, producing the rotation of the vertebrae.

Buckling. The fact that sagittal plane curvature of the spine is flattest in the early teen years supports the idea that this shape places the spine at risk for development of scoliosis. The shape of the spine in scoliosis is reminiscent of a buckled beam, but buckling may not explain the development of lateral curvatures since buckling of the ligamentous spine first exaggerates the sagittal curvatures (64).

Nonbiomechanical Factors

Despite several recent promising developments concerning genetic (65–67) and systemic (68,69) anomalies in patients with AIS, the origin of the initial spinal asymmetry remains unknown. While the regulation of growth and development produces random distributions of right–left asymmetry that are commonly centered around a mean of perfect symmetry (called fluctuating asymmetry), the spine is apparently subject to a higher tendency to spontaneous development of right convex thoracic asymmetry, possibly linked to developmental or mechanical instability. There is some evidence of subtle disturbance



Figure 9. Range sensing topographic scanner (two of the six cameras that surrounds the subject are shown on left panel) and biplanar X-ray system (middle panel) for concurrent surface and spine imaging, showing typical superimposed 3D torso spine model (right panel).

of systemic growth factors (70), abnormal collagen synthesis (71), and abnormalities of muscles and neuromuscular control (72). Recently, Moreau et al. (73) reported that bone cells of patients having surgery for idiopathic scoliosis had an abnormal response to melatonin, implicating a dysfunction melatonin signaling in these patients. However, all these empirical findings have not yet been incorporated into a coordinated theory of the cause of idiopathic scoliosis.

BIOMECHANICS OF SCOLIOSIS PROGRESSION DURING GROWTH

Scoliosis progression is associated with the rate of skeletal growth (74,75). In the progression of a small deformity to a large one, it has been proposed that a small lateral curvature of the spine would load the vertebrae asymmetrically (76), leading to asymmetrical growth in the vertebral growth plates. This acceleration of the deformity associated with mechanically modulated growth has been termed a vicious cycle (77) starting after a certain threshold of spinal deformity has been reached. This theory of spinal growth sensitivity to loading asymmetry must, however, explain why the normal spinal curvatures (kyphosis and lordosis) in the sagittal plane do not progress into hyperkyphosis and hyperlordosis by the same mechanism, although progressive deformity in Scheuermann's kyphosis has been attributed to a similar mechanism (78). This concept of a biomechanical mechanism of progression of deformity is attractive intuitively, and has been incorporated into the rationale for brace and surgical treatments (79). However, it cannot be quantified without better knowledge of the normal spinal loading and the alteration of loading in scoliosis, and better understanding of the sensitivity of growth to the time course of mechanical load

and its magnitude. It is known that bone growth, including that of vertebrae, is modulated by sustained altered stress (increased stress slows growth). However, the mechanism of growth of intervertebral disks, and its possible modulation by altered mechanical conditions is unknown (80). Notably, there does not appear to be any sudden change in disk matrix synthesis at the time of skeletal maturity (81), although scoliosis progression slows considerably after cessation of growth (82).

Scoliosis progression by alteration of bone growth has been investigated both experimentally (83) and analytically. By using heuristic (84) or empirical (85) estimates of spinal loading and growth plate response to altered stress it has been possible to demonstrate the plausibility of the vicious cycle theory of mechanically modulated vertebral growth in progressive scoliosis. In these incremental analyses the asymmetrical forces acting on vertebrae are estimated, and a small increment of spinal growth is applied to the vertebrae, modulated by the applied load. Then the spinal geometry is updated, loading asymmetries recalculated, and the process is repeated until the simulated adolescent growth is achieved. Results are sensitive to numerous assumptions about the initial geometry, the prevailing loading of the spine, as well as the extent to which growth is modulated by chronically altered stress.

BIOMECHANICS OF CONSERVATIVE TREATMENT

Treatment for patients with scoliosis aims to reduce not only the lateral deviation of the axial skeleton, but also the cosmetic deformity that is associated with the axial torsion of the trunk. All the common treatment methods are mechanical in nature, yet the mechanical aspects of the treatment have only recently been studied. The magnitude, area of application and balance of the forces, and their

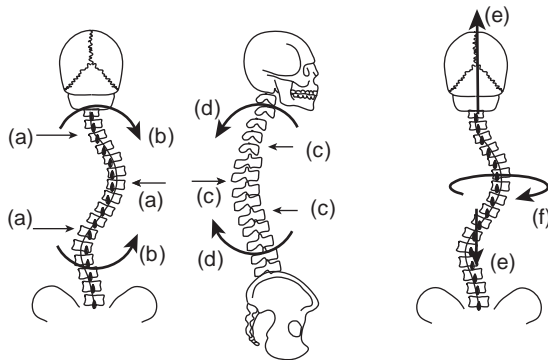


Figure 10. Possible correction force components applied by surgical instrumentation to the spine. Lateral force (a); coronal moment (b); sagittal force (c); sagittal moment (d); distraction force (e); axial moment (f).

moments about the spine are all important to the end result. The types of force applied to the spine for correction are shown in Fig. 10. These forces may either come from an external passive system (i.e., brace) or from an internal active mechanism (i.e. muscles control to shift the trunk away from the pressure areas of the brace).

Braces, Exercises, Electrical Stimulation

A brace is an orthosis that is used to prevent an increase in, or even to reduce, the curvature while waiting for the patient to reach skeletal maturity. The Boston brace, introduced in 1971 by Hall et al. (86) is widely used to treat moderate scoliosis. It has a plastic body envelope with pressure cushions adjusted in the back with straps. Its predecessor, the Milwaukee brace, is made of a neck ring extended by two metal bars and corrective pads. Braces are tailored specifically for each patient and are adjusted by the orthotist. Appropriate positions of the brace's pressure pads are determined from radiographs and clinical examination. Braces use a three- or four-point pressure principle (87), with pads placed generally over the posterolateral part of those ribs that connect to the apex of the thoracic spine deformity, and in the lumbar region, at the level of the scoliosis apex. The strap tension is determined empirically by the orthotist or the orthopedic surgeon (88).

Braces can improve the natural history of scoliosis according to case studies (87), a meta-analysis of published literature (89) and multicenter randomized and controlled studies (90). However, in 81% of the cases the correction is partial or the brace is not sufficient to stop the progression, which cast some doubt on the usefulness and efficacy of braces (91). Clinical studies may be biased by including patients having scoliosis curves that are not likely to progress, with the brace thereby erroneously deemed as effective. Also, it is apparent that many patients do not wear the brace as much as is prescribed (poor compliance), which may reduce the documented effectiveness of bracing. Physical exercises can also be recommended to strengthen back muscles and to improve the spine's stability as well as to reduce back pain. However, these exercises have not been proven to correct nor prevent the progression of

scoliosis deformity and should be used along with a more effective treatment like bracing.

Electrical stimulation of the back muscles has been used in the 1970s for the conservative treatment of scoliosis, but has not been found to be effective (90).

Biomechanical Evaluation of Brace Function

The mode of action of a brace is indirect, since it does not apply forces directly to the spine (92). Many factors influence brace efficiency, including the flexibility of the spine, the shape and stiffness of the brace shell, the location, size and thickness of the brace pads (or of the voids inside the brace), the strap tension adjustment, the biomechanical properties of trunk and thorax tissues, and the duration of the brace forces (93,94).

Analyses made using 3D reconstruction of the spine and rib cage of patients with scoliosis who were treated with the Boston brace have revealed that the brace produced immediate significant curve correction of the spinal deformity in the frontal plane at the expense of a significant reduction of thoracic kyphosis in the sagittal plane and without significant effect on the rib hump, vertebral rotation or frontal balance (95). Coupling mechanisms between the applied forces on the rib cage and the coronal deformity of the spine may explain partially the lack of effect in these planes (32).

The forces generated by braces have been evaluated by measurement of pressure distribution between the brace and the patient's torso (Fig. 11) (94,96) High force zones (20–113 N) are mostly located on both sides of the pelvis, on the lower anterior part of the abdomen, on the posterolateral part of the right thoracic zone, and on the lateral part of the left lumbar zone. However, in some cases unfavorable forces were measured on the left thoracic or on the right lumbar parts of the torso, which can explain some of the negative results of bracing. Correction of curves was not solely dependent on the level of force applied by the brace (96) and some patients with the greatest curves achieved little correction despite significant levels of applied force. Jiang (68) found that although high strap forces are necessary to ensure lateral and derotational forces on the spine, they also cause undesirable forces that induce lordosis. Large variability (8–81 N) was measured in the prescribed tension in the thoracic and pelvic straps of the brace (93), and significant relaxation of strap tension also was found a few minutes after adjustments had been made.

Computer Modeling of Bracing

Biomechanical finite element models simulating the orthotic treatment have been developed to analyze brace biomechanics. In the analyses, forces representing brace pads were applied to models representing specific patients' trunk geometry. The traction force applied by the mandible support of the Milwaukee brace was found to be of relatively minor importance in the correction of scoliotic curves, whereas the lumbar pad significantly affects correction in a lumbar curve and thoracic pad often completely dominates the nature of the correction produced (98). The best correction of scoliotic curves could be obtained using

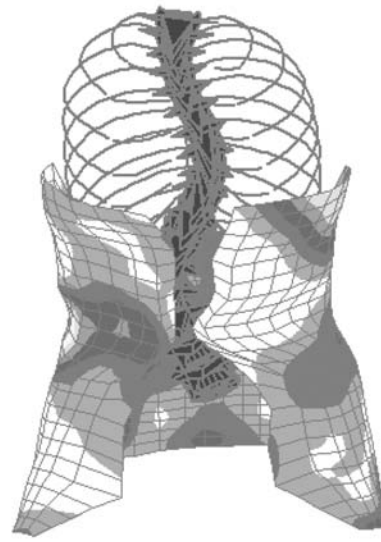


Figure 11. Analysis of brace effectiveness using a flexible matrix of 192 pressure sensors (at the brace-torso interface) and force cells (to measure strap tension). The right side panel shows a typical pressure distribution generated by a brace that is mapped onto a surfacic model of the brace envelope, and superimposed over the three-dimensional reconstruction of the spine and rib cage.

a thoracic pad without lumbar or subaxillary counterpad for thoracic curves, and a lumbar pad with a thoracic counterpad for lumbar curves (99). Application of passive primary forces on the convex side and counter forces on both thoracic and lumbar concave sides of a thoracic scoliotic curve had a substantial corrective effect on the Cobb angle and lateral alignment, as well as do active muscle forces (100). The Boston brace system also was modeled and the brace mechanisms investigated (96,101). An optimization approach of the Boston brace treatment of scoliosis showed that the optimal (most effective) brace forces were mostly located on the convex side of the spinal curves (102). However, the optimal configuration only achieved overall correction of 50% on average.

In previous models of bracing, the action of the brace was represented as force generators instead of passive deformable systems that interact with the flexible torso. Also the complex mechanical action of the brace on the entire torso was not completely addressed. Current modeling efforts are oriented toward the detailed explicit representation of the brace-torso interface (103) (Fig. 12) to improve their simulation of the complete brace system's interaction with the patient, and to optimize the brace design parameters.

BIOMECHANICS OF SURGICAL TREATMENT

In the case of severe spinal deformities (Cobb angle $> \sim 50^\circ$), surgical spinal fusion with metallic implants is often performed. The goals of surgical treatment are to prevent further worsening in the scoliosis, and if possible to straighten the spine. This is normally achieved through arthrodesis (bony fusion) of a region of the spine that spans one or more scoliosis curves. Implants used in surgical treatment of spinal deformities have two complementary mechanical roles. The first is to apply forces to

correct or reduce the spinal deformity intraoperatively and to maintain correction subsequently. The second is to create the correct mechanical environment for spinal fusion to occur by immobilizing the spine until bony fusion has occurred.

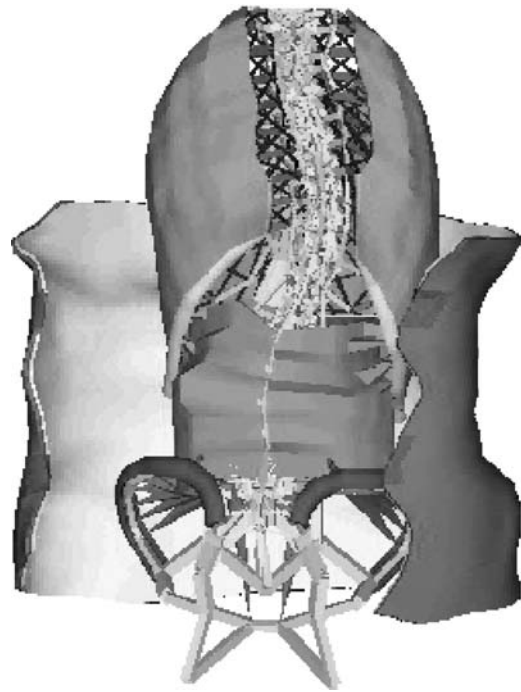


Figure 12. detailed finite element modeling of the brace-torso interface (pelvis not shown). The opened brace is first introduced onto the spine and rib cage model. When the brace is closing (when the opening forces are released), contacts are established at the brace-torso interface and reaction forces are generated. Strap tightening is further generating additional forces.

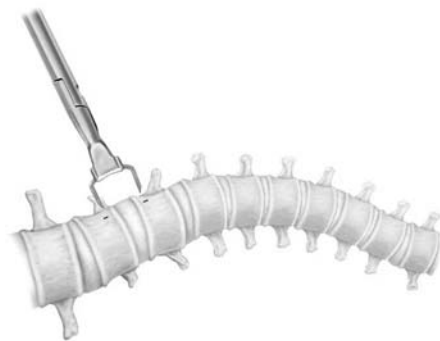
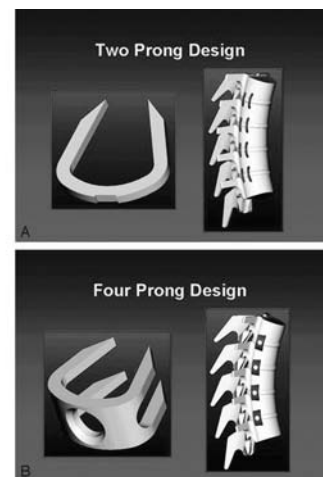


Figure 13. Stapling of intervertebral disk interspace. (Reproduced from Ref. 109, with permission.)



Types of Spinal Instrumentation and Surgical Planning

The Harrington instrumentation system was introduced in the 1960s, with many subsequent advances, including segmental instrumentation (attached to numerous vertebrae) and anterior spinal instrumentation systems (attached to the front of the spine). The main components of the Harrington instrumentation are hooks, rod with ratchets and cross members. An outrigger could be used to provide some predistracted and stretching of ligaments before the actual distraction rod is implanted on the concave side of the spine deformity. A second rod in slight compression can be installed on the convex side of the spine curve. The main disadvantages of this instrumentation are its limited number of attachments to the spine; consequently, it mainly corrected the deformity in the frontal plane and occasionally resulted in back flattening (104). In the mid-1980s, the 3D nature of scoliosis deformity was becoming more widely recognized and surgical systems were developed that addressed correction of the deformity in 3D. Luque (105) advanced the use of wires passed around the neural arch of vertebrae (sublaminar wiring) to provide multiple connections between a profiled (curved) rod and the vertebrae. The first 3D system was developed by Dubousset and Cotrel (106). It had a curved rod system that was rotated during surgery. A precontoured rod, to account for the deformity and a final kyphosis, is loosely attached to the hooks and then rotated $\sim 90^\circ$ and tightened. The belief is that the deformity in the frontal plane is rotated 90° to provide a natural kyphosis in the sagittal plane. Subsequent developments included pedicular screws that provide stronger anchorage to the vertebrae than simple hooks. With screws, direct derotation of the apical vertebra can be applied before securing the screws on the rods. Instead of additional hooks, sublaminar wires may be used to attach the spine to the rod.

Surgical release of anterior structures and fusion of adjacent vertebral bodies are sometimes necessary in conjunction with posterior fusion to prevent the crankshaft phenomenon. This occurs due to the continued anterior growth in skeletally immature patients. A completely anterior approach and fusion to correct scoliosis was first

introduced by Dwyer and Schafer (107) and was popularized with the Zielke instrumentation in the 1980s (108). Anterior fusion is considered controversial (e.g., because of risks to anterior structures including nerves and major blood vessels), so posterior instrumentation with fusion is still considered the gold standard.

Staples applied between adjacent vertebrae are a promising minimal invasive technology to correct spinal deformities (109) (Fig. 13). Compression staples are used on the convex side of a spine curve to reduce growth while distraction staples are placed on the concave side to accelerate growth.

Testing of Spinal Instrumentation (Construct Design)

Much of the understanding about results of spinal instrumentation and fusion is empirical, based on follow-up studies. Biomechanical principles should be able to provide additional information. The use of surgical instrumentation has a direct mechanical effect on the fused part of the spine, and an indirect effect on the unfused region. It remains as a challenge to biomechanics to further our understanding of the interactions between the spine, the instrumentation, and the muscular and other forces. The muscles, and their control presents the greatest difficulties, since CNS (central nervous system) control of trunk balance is so poorly understood.

There are three basic types of spinal implant construct testing: tests of individual implant components (hooks, screws, rods etc.), tests of the connections between an implant, and the spine and tests of complete instrumentation assemblies. All can be tested in simulations of *in vivo* conditions to evaluate their performance. Laboratory testing of a spinal construct (consisting of a test instrumentation applied to a standardized spinal specimen) can provide information about its flexibility, strength, and fatigue life. Components, such as pedicle screws, may be tested in pull-out (110) or bending. The challenge in all these tests is to establish conditions that are representative of the *in vivo* situation. Both strength and stiffness are desirable properties of instrumentation, and fatigue testing is important to establish whether instrumentation will mechanically fail

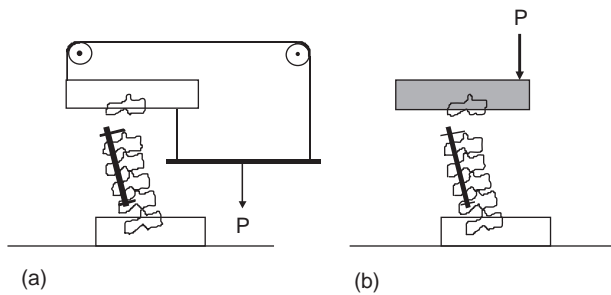


Figure 14. Mechanical test of spinal instrumentation construct, (a) with a pure moment imposing flexion-extension by means of a couple (equal parallel forces). (b) with offset compression loading.

or the conditions will lead to pseudoarthroses (failure to fuse this spine segment completely) with repetitive loading over a period of time.

In a typical construct test, instrumentation is applied to a spine specimen, which is then loaded. The resulting deformations are measured. Alternatively (but rarely), a construct can be tested by applying a displacement, and measuring the necessary forces: a flexibility test, which is the inverse of the stiffness test. In a typical stiffness test, a pure moment is applied via a couple consisting of two equal parallel forces. Alternatively, offset compression loading (Fig. 14) is intended to be more physiologically realistic by compressing the spine axially at the same time as creating a bending moment. Most reports then give the resulting rotations about the same axis as the applied torque, or collinear with the applied load. If additional rotations or displacements about other axes are recorded, these provide information about the coupled (as opposed to direct) stiffnesses. The motion is typically recorded by mechanical or opto-electronic methods, or by magnetic sensors of position and orientation.

Some information about *in vivo* loading of internal fixators has been provided by means of telemetry of signals from force transducers built into the instrumentation (111). These show forces on the order of 250 N and moments on the order of 6 Nm transmitted through the instrumentation during walking, but this constitutes an unknown proportion of the total load carried by the spine and the instrumentation.

Many instrumentation tests are performed with animal spines because human spines representative of the target population (of people requiring implants) are hard to obtain. Furthermore, human spinal specimens are more variable in their mechanical properties than animal spines of similar ages, taken from animals of the same breed. Bovine calf and pig spines have been shown to be a reasonable choice based on similar vertebral dimensions and flexibility properties.

Standardized instrumentation testing protocols (112–114) should be advantageous to facilitate comparisons between tests performed in different labs. However, standardization is difficult. In addition to defining the exact conditions for the tests, standardization has to take into account the difficulty of obtaining suitable human cadaveric material or the appropriateness of using animal spines. Standardization has been elusive, and to date only

the testing of instrumentation systems attached to plastic blocks (corpectomy model) has been provisionally standardized by the ASTM (111).

It remains unclear how to interpret the information in reports of stiffness and strength testing of spinal implants, in terms of how to select an instrumentation system and adapt it to the individual needs of a particular patient. It appears that stiffer instrumentation is better, providing the instrumentation size is not excessive. Given that there are six degrees of freedom possible in a stiffness test, it is not known which components of rotation and translational stiffness are the most important. An important aspect of the instrumentation is how easy it is to apply, with minimum blood loss and risk of injury, for example, to neural and vascular structures. Also, the ability to achieve the desired realignment of the spinal column is important. Biomechanical information can be used in conjunction with other information in making surgical decisions, and biomechanical testing ought to be used with the goal of reducing the need for empiricism in the development of knowledge about the outcomes of different surgical strategies.

Analytical Simulation of Surgical Maneuvers

Segmental instrumentation offers surgeons many variables and multistep maneuvers to adapt to individual patients' needs. Biomechanical modeling offers the possibility to explore these options in advance, and to assist in decision making. In theory, if the mechanical properties of both the spine and the instrumentation were understood completely, then biomechanical models could be built and be used to predict the outcome of surgery. However, the complexity of the surgical choices creates many unknown inputs for the biomechanical analyses, and difficulties in validation of model predictions. To simulate the procedures of the Cotrel–Dubousset instrumentation (Medtronic Sofamor–Danek, Memphis, TN), for example, the surgical maneuvers for the concave-side rod of a segmental instrumentation can be represented as four steps: (1) install the rod passively to the end hooks, then approximate the intermediate hooks to the rod; (2) displace the hooks along the rod to their final positions (hook distraction); (3) rotate the concave-side rod; (4) lock the hooks to the rod and relax the applied torque (spring-back). It has proven very difficult to quantify the required magnitude and direction of all these displacement inputs in the simulation. Another difficulty in using a mechanical deterministic model to predict the outcome of surgery for an individual patient is the unknown flexibility of each motion segment and of the rib cage.

Another possibility offered by analytical modeling is the ability to calculate stresses at selected sites in the spine and instrumentation. Finite element analyses were used to estimate stresses in internal fixation devices (115,116). These models have provided estimates of the change in bone stress for different vertebral injury situations, with inclusion of instrumentation components such as plastic washers in the construct, and after incorporation of bone graft. Biomechanical analyses have also been used to investigate the consequences of surgical variables such

as angulation of pedicle screws on the rigidity of the construct (110,117).

Most biomechanical models simulating surgical procedures have employed finite element models. A finite element model of the spine and rib cage (118) was adapted to investigate the biomechanics of Harrington instrumentation (119). The hooks were connected to the instrumented vertebrae (with rotation allowed along vertical axis), while the rod distraction maneuver was modeled as a thermal expansion of the rod to reach its final shape. The biomechanics of Cotrel–Dubousset instrumentation also was studied with this model, in an idealized geometry (120). This study was extended by simulating the surgical maneuvers on fifteen surgical cases (121,122). Each model's geometry was personalized to each patient's 3D skeletal geometry, built from preoperative stereo X rays, and results were compared with documented post-surgical geometry. The simulations of surgical maneuvers showed generally good agreement (on average) with measured effects of surgery in the frontal plane. However, in the sagittal and transverse planes, the response was dependent on model's boundary conditions, motion segments' mechanical properties and instrumentation variables.

A kinematic model including flexible elements to represent each motion segment and kinematic joints and sets to model the instrumentation (flexible multibody approach) was developed by Aubin et al. (123,124). This was intended to overcome limitations of finite element structural modeling that occur when there are large differences in the structure's stiffness (the surgical construct has properties of a mechanism as well as of a structure). Fig. 15 shows preliminary results of this modeling approach for 1 of 20 cases. In the simulations, geometric indices such as Cobb angles showed trends similar to those observed during surgery: gradual correction during surgical procedures and partial loss of correction after hooks/screws lock-up, when the strains induced by the correction maneuvers and stored in the noninstrumented part of the spine were released. To complement this kind of surgical simulation, a spine surgery simulator and a virtual reality training simulator (Fig. 16) currently are under development (125).

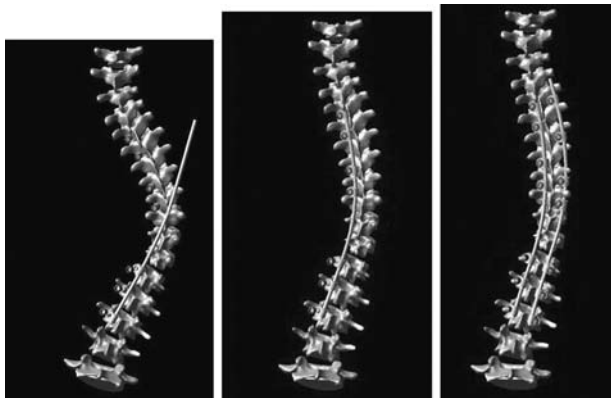


Figure 15. Biomechanical simulation of a spinal instrumentation procedure. The left and central panels show the attachment of the first rod, while the right panel shows the installation of the second rod.



Figure 16. Virtual reality spine surgery simulator. The left panel shows one part of the virtual surgery room with the operating table, the patient and the exposed spine, as well as the radiographs and a magnification of the selected vertebra. The right panel shows an operator that is performing a virtual reality surgery (using stereo stereoscopic glasses and a 3D mouse).

Once validated, these combined tools will allow surgeons to test various surgical options before the real surgery and determine the best scenario for each patient's spinal deformity.

Simulation of the mechanical action of various designs of costoplasties (operations to change the shape of the ribs) by Grealou et al. (33) and Carrier et al. (34) is another recent development. They analyzed different procedures (side, location, length, and number of ribs to resect or to graft), to investigate the biomechanical effect on the spinal and thoracic shape, and to test the idea that spinal deformity can be treated by operations on the rib cage alone.

CONCLUSION

Scoliosis is a costly problem in terms of both healthcare resources and human suffering. Many cases have unknown etiology, and there may be a genetic component in their causation that is modified by environmental factors.

Mechanical analyses of scoliosis as a buckling phenomenon have not been very illuminating, possibly because they are single plane models and do not incorporate rotation or other coupling. The interactions between biomechanical factors and spinal stability during growth are probably important in most scoliosis deformities, irrespective of their initial cause. Biomechanical factors including motion segment coupling, spinal length and slenderness, muscle asymmetry, and postural sway, are apparently relevant in the development and progression of these deformities.

Geometrical aspects of scoliosis can be better documented and understood with 3D measurement including low dose 3D X-ray techniques. These provide a more complete picture of the deformity than the relatively inaccurate single plane Cobb angle measurement. Noninvasive surface measurement techniques show promise, but there is still a need to establish their relationship to radiographic measurements, especially during progression of a curve.

Although many experimental and biomechanical models of scoliosis have been developed, the mechanisms behind the spinal deformity process and its treatment

are not yet clear. Developing a better understanding of the mechanical aspects of the deforming scoliotic spine will lead to a more selective and appropriate means of treatment, especially if the mechanisms that produce the deformity are confirmed to be reversible. Certainly, the empirical evidence from successful brace or stapling treatments suggests that this is the case. Early detection and treatment by mechanical methods show promise for prevention of progression of the deformity.

BIBLIOGRAPHY

- Danielsson AJ, Nachemson AL. Back pain and function 22 years after brace treatment for adolescent idiopathic scoliosis: A case-control study-part I. *Spine* 2003;28:2079–2085.
- Danielsson AJ, Nachemson AL. Childbearing, curve progression, and sexual function in women 22 years after treatment for adolescent idiopathic scoliosis: A case-control study. *Spine* 2001;26:1449–1456.
- Noonan KJ, Dolan LA, Jacobson WC, Weinstein SL. Long-term psychosocial characteristics of patients treated for idiopathic scoliosis. *J Pediatr Orthop* 1997;17:712–717.
- Scoliosis Research Society. Three-dimensional terminology of Spinal Deformity. Available at <http://www.srs.org/professionals/glossary/3-d.asp>. Accessed 2005.
- Armstrong GW, Livermore NB 3rd, Suzuki N, Armstrong JG. Nonstandard vertebral rotation in scoliosis screening patients. Its prevalence and relation to the clinical deformity. *Spine* 1982;7:50–54.
- Parent S, et al. Morphometric analysis of anatomic scoliotic specimens. *Spine* 2002;27:2305–2311.
- Thometz JG, Lamdan R, Liu XC, Lyon R. Relationship between Quantec measurement and Cobb angle in patients with idiopathic scoliosis. *J Pediatr Orthop* 2000;20:512–516.
- Stokes IA. Axial rotation component of thoracic scoliosis. *J Orthop Res* 1989;7:702–708.
- Stokes IA, Armstrong JG, Moreland MS. Spinal deformity and back surface asymmetry in idiopathic scoliosis. *J Orthop Res* 1988;6:129–137.
- Acaroglu E, et al. Does transverse apex coincide with coronal apex levels (regional or global) in adolescent idiopathic scoliosis? *Spine* 2001;26:1143–1146.
- Thulbourne T, Gillespie R. The rib hump in idiopathic scoliosis. Measurement, analysis and response to treatment. *J Bone Joint Surg Br* 1976;58:64–71.
- Pearcy MJ, Whittle MW. Movements of the lumbar spine measured by three-dimensional X-ray analysis. *J Biomed Eng* 1982;4:107–112.
- Pope MH, Wilder DG, Mattern RE, Frymoyer JW. Experimental measurements of vertebral motion under load. *Orthop Clin N Am* 1977;8:155–167.
- Stokes IA, Wilder DG, Frymoyer JW, Pope MH. 1981 Volvo award in clinical sciences. Assessment of patients with low-back pain by biplanar radiographic measurement of intervertebral motion. *Spine* 1981;6:233–240.
- Lovett RW. The mechanics of the normal spine and its relation to scoliosis. *Boston Med Surg J* 1905; 153.
- White AA 3rd. Kinematics of the normal spine as related to scoliosis. *J Biomech* 1971;4:405–411.
- Stokes IA, Moreland MS. Measurement of the shape of the surface of the back in patients with scoliosis. The standing and forward-bending positions. *J Bone Joint Surg Am* 1987;69:203–211.
- Bunnell WP. An objective criterion for scoliosis screening. *J Bone Joint Surg Am* 1984;66:1381–1387.
- Scoliosis Research Society Terminology Committee. Revised Glossary of Terms. <http://www.srs.org/professionals/glossary/glossary.asp>. Accessed 2005.
- Goldstein LA, Waugh TR. Classification and terminology of scoliosis. *Clin Orthop Relat Res* 1973;93:10–22.
- King HA, Moe JH, Bradford DS, Winter RB. The selection of fusion levels in thoracic idiopathic scoliosis. *J Bone Joint Surg Am* 1983;65:1302–1313.
- Richards BS, Sucato DJ, Konigsberg DE, Ouellet JA. Comparison of reliability between the Lenke and King classification systems for adolescent idiopathic scoliosis using radiographs that were not premeasured. *Spine* 2003;28:1148–1156. discussion 1156–1157.
- Stokes IA, Aronsson DD. Computer-assisted algorithms improve reliability of King classification and Cobb measurement of scoliosis. *Spine* (in press).
- Lenke LG, et al. Curve prevalence of a new classification of operative adolescent idiopathic scoliosis: Does classification correlate with treatment? *Spine* 2002;27:604–611.
- Stokes IA. Three-dimensional terminology of spinal deformity. A report presented to the Scoliosis Research Society by the Scoliosis Research Society Working Group on 3-D terminology of spinal deformity. *Spine* 1994;19:236–248.
- Peloux J du, Fauchet R, Faucon B, Stagnara P. Le plan détection pour l'examen radiologique des cyphoscolioses *Rev Chir Orthop* 1965;51:517–524.
- Raso J, Gillespie R, McNiece G. Determination of the plane of maximum deformity in idiopathic scoliosis. *Orthop Trans* 1981; 4.
- Kreuzig E. *Advanced Engineering Mathematics 7th ed. and Mathematics Manual to Accompany Advanced Engineering Mathematics 7th ed.* New York: John Wiley & Sons Inc; 1995.
- Poncet P, Dansereau J, Labelle H. Geometric torsion in idiopathic scoliosis: Three-dimensional analysis and proposal for a new classification. *Spine* 2001;26:2235–2243.
- Cobb JR. Outline for the Study of Scoliosis. The American Academy of Orthopaedic Surgeons: Instructional Course Lectures 5 48: p 261–275.
- Gardner-Morse MG, Stokes IA. Structural behavior of human lumbar spinal motion segments. *J Biomech* 2004;37:205–212.
- Aubin CE, Dansereau J, de Guise JA, Labelle H. Rib cage-spine coupling patterns involved in brace treatment of adolescent idiopathic scoliosis. *Spine* 1997;22:639–635.
- Grealou L, Aubin CE, Labelle H. Rib cage surgery for the treatment of scoliosis: A biomechanical study of correction mechanisms. *J Orthop Res* 2002;20:1121–1128.
- Carrier J, Aubin CE, Villemure I, Labelle H. Biomechanical modelling of growth modulation following rib shortening or lengthening in adolescent idiopathic scoliosis. *Med Biol Eng Comput* 2004;42:541–548.
- Kramers-de Quervain IA, Muller R, Stacoff A, Grob D, and Stussi E. Gait analysis in patients with idiopathic scoliosis. *Eur Spine J* 2004;13:449–456.
- Lenke LG, et al. Prospective dynamic functional evaluation of gait and spinal balance following spinal fusion in adolescent idiopathic scoliosis. *Spine* 2001;26:E330–E337.
- Panjabi MM, Brand RA. Jr, White AA 3rd. Three-dimensional flexibility and stiffness properties of the human thoracic spine. *J Biomech* 1976;9:185–192.
- Janevic J, Ashton-Miller JA, Schultz AB. Large compressive preloads decrease lumbar motion segment flexibility. *J Orthop Res* 1991;9:228–236.

39. Nash CL. Jr, Gregg EC, Brown RH, Pillai K. Risks of exposure to X-rays in patients undergoing long-term treatment for scoliosis. *J Bone Joint Surg Am* 1979;61:371–374.
40. Ardran GM, et al. Assessment of scoliosis in children: low dose radiographic technique. *Br J Radiol* 1981;53:146–147.
41. DeSmet AA, Goin JE, Asher MA, Scheuch HG. A clinical study of the differences between the scoliotic angles measured on posteroanterior and anteroposterior radiographs. *J Bone Joint Surg Am* 1982;64:489–493.
42. Perdriolle R, Vidal J. [A study of scoliotic curve. The importance of extension and vertebral rotation (author's transl)]. *Rev Chir Orthop Reparatrice Appar Mot* 1981;67:25–34.
43. Drerup B. Principles of measurement of vertebral rotation from frontal projections of the pedicles. *J Biomech* 1984;17:923–935.
44. Stokes IA, Bigalow LC, Moreland MS. Measurement of axial rotation of vertebrae in scoliosis. *Spine* 1986;11:213–218.
45. Brown RH, Burstein AH, Nash CL, Schock CC. Spinal analysis using a three-dimensional radiographic technique. *J Biomech* 1976;9:355–365.
46. Marzan GT. Rational design for close-range photogrammetry Ph.D. dissertation, University of Illinois, 1976 Xerox University, Microfilms, Ann Arbor MI 1976; p. 46.
47. Dansereau J, Stokes IA. Measurements of the three-dimensional shape of the rib cage. *J Biomech* 1988;21:893–901.
48. Aubin CE, et al. Morphometric evaluations of personalised 3D reconstructions and geometric models of the human . *Spine. Med Biol Eng Comput* 1997;35:611–618.
49. Delorme S, Petit Y, de Guise JA, Aubin CÉ, Dansereau J. Assessment of the 3-D reconstruction and high-resolution geometrical modeling of the human skeletal trunk from 2-D radiographic images. *IEEE Trans Biomed Eng* 2003;50:989–998.
50. Cheriet F, et al. Towards the self-calibration of a multi-view radiographic imaging system for the 3D reconstruction of the human spine and rib cage. *Int J Pattern Recog Arti Intell* 1999;13:761–779.
51. Aubin CE, et al. [Geometrical modeling of the spine and the thorax for the biomechanical analysis of scoliotic deformities using the finite element method]. *Ann Chir* 1995;49:749–761.
52. Dansereau J, Labelle H, Aubin CE. 3-D personalized parametric modelling of reconstructed scoliotic spines. IVth International Symposium on Computer Simulation in Biomechanics. 1.6–1.9 p. 93.
53. Aaro S, Dahlborn M. The longitudinal axis rotation of the apical vertebra, the vertebral, spinal, and rib cage deformity in idiopathic scoliosis studied by computer tomography. *Spine* 1981;6:567–572.
54. Theologis TN, Fairbank JC, Turner-Smith AR, Pantazopoulos T. Early detection of progression in adolescent idiopathic scoliosis by measurement of changes in back shape with the Integrated Shape Imaging System scanner. *Spine* 1997;22:1223–1227; discussion 1228.
55. Stokes IA, Moreland MS. Concordance of back surface asymmetry and spine shape in idiopathic scoliosis. *Spine* 1989;14:73–78.
56. Jaremko JL, et al. Genetic algorithm-neural network estimation of Cobb angle from torso asymmetry in scoliosis. *J Biomech Eng* 2002;124:496–503.
57. Jaremko JL, et al. Estimation of spinal deformity in scoliosis from torso surface cross sections. *Spine* 2001;26(14):1583–1591.
58. Frobin W, Hierholzer E. Analysis of human back shape using surface curvatures. *J Biomech* 1982;15:379–390.
59. Pazos V, et al. Accuracy assessment of human trunk surface 3D reconstructions from an optical digitising system. *Med Biol Eng Comput* 2005;43:11–15.
60. Nachemson A, Sahlstrand T. Etiologic factors in adolescent idiopathic scoliosis. *Spine* 1977;2:176–184.
61. Robin GC. The Aetiology of Scoliosis. A Review of a Century of Research. Boca Raton (FL): CRC Press: 1990.
62. Somerville EW. Rotational lordosis; the development of single curve. *J Bone Joint Surg Br* 1952;34-B:421–427.
63. Stokes IA, Gardner-Morse M. Analysis of the interaction between vertebral lateral deviation and axial rotation in scoliosis. *J Biomech* 1991;24:753–759.
64. Veldhuizen AG, Scholten PJ. Kinematics of the scoliotic spine as related to the normal spine. *Spine* 1987;12:852–858.
65. Axenovich TI, et al. Segregation analysis of idiopathic scoliosis: demonstration of a major gene effect. *Am J Med Genet* 1999;86:389–394.
66. Bashiardes S, et al. SNTG1, the gene encoding gamma1-syntrophin: A candidate gene for idiopathic scoliosis. *Hum Genet* 2004;115:81–89.
67. Justice CM, et al. Familial idiopathic scoliosis: Evidence of an X-linked susceptibility locus. *Spine* 2003;28:589–594.
68. Bagnall KM, et al. Melatonin levels in idiopathic scoliosis. Diurnal and nocturnal serum melatonin levels in girls with adolescent idiopathic scoliosis. *Spine* 1996;21:1974–1978.
69. Lowe T, et al. Platelet calmodulin levels in adolescent idiopathic scoliosis: Do the levels correlate with curve progression and severity? *Spine* 2002;27:768–775.
70. Skogland LB, Miller JA. Growth related hormones in idiopathic scoliosis. An endocrine basis for accelerated growth. *Acta Orthop Scand* 1981;51:779–781.
71. Aigner T. Variation with age in the pattern of type X collagen expression in normal and scoliotic human intervertebral discs. *Calcif Tissue Int* 1998;63:263–268.
72. Veldhuizen AG, Wever DJ, Webb PJ. The aetiology of idiopathic scoliosis: biomechanical and neuromuscular factors. *Eur Spine J* 2000;9:178–184.
73. Moreau A, et al. Melatonin signaling dysfunction in adolescent idiopathic scoliosis. *Spine* 2004;29:1772–1781.
74. Little DG, Song KM, Katz D, Herring JA. Relationship of peak height velocity to other maturity indicators in idiopathic scoliosis in girls. *J Bone Joint Surg Am* 2000;82:685–693.
75. Lonstein JE, Carlson JM. The prediction of curve progression in untreated idiopathic scoliosis during growth. *J Bone Joint Surg Am* 1984;66:1061–1071.
76. Stokes IA, Gardner-Morse M. Muscle activation strategies and symmetry of spinal loading in the lumbar spine with scoliosis. *Spine* 2004;29:2103–2107.
77. Stokes IA, Spence H, Aronsson DD, Kilmer N. Mechanical modulation of vertebral body growth. Implications for scoliosis progression. *Spine* 1996;21:1162–1167.
78. Scoles PV, et al. Vertebral alterations in Scheuermann's kyphosis. *Spine* 1991;16:509–515.
79. Roaf R. The treatment of progressive scoliosis by unilateral growth-arrest. *J Bone Joint Surg Br* 1963;45:637–651.
80. Urban JP, Roberts S. Development and degeneration of the intervertebral discs. *Mol Med Today* 1995;1:329–335.
81. Antoniou J, et al. Elevated synthetic activity in the convex side of scoliotic intervertebral discs and endplates compared with normal tissues. *Spine* 2001;26:E198–E206.
82. Weinstein SL, Ponseti IV. Curve progression in idiopathic scoliosis. *J Bone Joint Surg Am* 1983;65:447–455.

83. Mente PL, Aronsson DD, Stokes IA, Iatridis JC. Mechanical modulation of growth for the correction of vertebral wedge deformities. *J Orthop Res* 1999;17:518–524.
84. Villemure I, Aubin CE, Dansereau J, Labelle H. Simulation of progressive deformities in adolescent idiopathic scoliosis using a biomechanical model integrating vertebral growth modulation. *J Biomech Eng* 2002;124:784–790.
85. Stokes IA. Biomechanical spinal growth modulation and progressive adolescent scoliosis—a test of the ‘vicious cycle’. Electronic Focus Group report. *Eur Spine J* (in press).
86. Hall JE, Miller ME, Schumann W, Stanish W. A refined concept in the orthotic treatment management of scoliosis. *Orthot Prosthet* 1975;4:7–13.
87. Emans JB, et al. The Boston bracing system for idiopathic scoliosis. Follow-up results in 295 patients. *Spine* 1986;11:792–811.
88. Watts HG. Bracing in spinal deformities. *Orthop Clin N Am* 1979;10:769–785.
89. Rowe DE, et al. A meta-analysis of the efficacy of non-operative treatments for idiopathic scoliosis. *J Bone Joint Surg Am* 1997;79, 664–674.
90. Nachemson AL, Peterson LE. Effectiveness of treatment with a brace in girls who have adolescent idiopathic scoliosis. A prospective, controlled study based on data from the Brace Study of the Scoliosis Research Society. *J Bone Joint Surg Am* 1995;77:815–822.
91. Edgar MA. To brace or not to brace? *J Bone Joint Surg Br* 1985;67:173–174.
92. Ogilvie J. *Spinal Orthotics. An overview. The pediatric spine: principles and practice 1787–1793.* New York, Raven Press Ltd.; 1994.
93. Aubin CE, et al. Variability of strap tension in brace treatment for adolescent idiopathic scoliosis. *Spine* 1999;24:349–354.
94. Mac-Thiong JM, et al. Biomechanical evaluation of the Boston brace system for the treatment of adolescent idiopathic scoliosis: Relationship between strap tension and brace interface forces. *Spine* 2004;29:26–32.
95. Labelle H, Dansereau J, Bellefleur C, Poitras B. Three-dimensional effect of the Boston brace on the thoracic spine and rib cage. *Spine* 1996;21:59–64.
96. Perie D, et al. Boston brace correction in idiopathic scoliosis: A biomechanical study. *Spine* 2003;28:1672–1677.
97. Chase AP, Bader DL, Houghton GR. The biomechanical effectiveness of the Boston brace in the management of adolescent idiopathic scoliosis. *Spine* 1989;14:636–642.
98. Andriacchi TP, Schultz AB, Belytschko TB, Dewald R. Milwaukee brace correction of idiopathic scoliosis. A biomechanical analysis and a retrospective study. *J Bone Joint Surg Am* 1976;58:816–815.
99. Patwardhan AG, et al. A biomechanical analog of curve progression and orthotic stabilization in idiopathic scoliosis. *J Biomech* 1986;19:103–117.
100. Wynarsky GT, Schultz AB. Optimization of skeletal configuration: studies of scoliosis correction biomechanics. *J Biomech* 1991;24:721–732.
101. Perie D, et al. Personalized biomechanical simulations of orthotic treatment in idiopathic scoliosis. *Clin Biomech (Bristol, Avon)* 2004;19:190–195.
102. Gignac D, Aubin CE, Dansereau J, Labelle H. Optimization method for 3D bracing correction of scoliosis using a finite element model. *Eur Spine J* 2000;9:185–190.
103. Perie D, et al. Biomechanical modelling of orthotic treatment of the scoliotic spine including a detailed representation of the brace-torso interface. *Med Biol Eng Comput* 2004;42:339–344.
104. Humke T, Grob D, Scheier H, Siegrist H, Cotrel-Dubousset and Harrington Instrumentation in idiopathic scoliosis: a comparison of long-term results. *Eur Spine J* 1995;4:281–283.
105. Luque ER. The anatomic basis and development of segmental spinal instrumentation. *Spine* 1982; May–Jun; 7(3):256–259.
106. Dubousset J, Cotrel Y. Application technique of Cotrel-Dubousset instrumentation for scoliosis deformities. *Clin Orthop Relat Res* 1991; 103–110.
107. Dwyer AF, Schafer MF. Anterior approach to scoliosis. Results of treatment in fifty-one cases. *J Bone Joint Surg Br* 1974;56:218–224.
108. Metz P, Zielke K. [First results of the Luque operation (author’s transl)]. *Z Orthop Ihre Grenzgeb* 1982;120:333–337.
109. Betz RR, et al. An innovative technique of vertebral body stapling for the treatment of patients with adolescent idiopathic scoliosis: A feasibility, safety, and utility study. *Spine* 2003;28:S255–S265.
110. Krag MH. Biomechanics of thoracolumbar spinal fixation. A review. *Spine* 1991;16:S84–S99.
111. Rohlmann A, Bergmann G, Graichen F. Loads on an internal spinal fixation device during walking. *J Biomech* 1997;30: 41–47.
112. ASTM. Standard test method for PS5-94 static and dynamic spinal implants assembly in a corpectomy model. American Society of Testing and Materials; 1997.
113. Ashman RB, et al. *In vitro* spinal arthrodesis implant mechanical testing protocols. *J Spinal Disord* 1989;2:274–281.
114. Panjabi MM. Biomechanical evaluation of spinal fixation devices: I. A conceptual framework. *Spine* 1988;13:1129–1134.
115. Goel VK, Pope MH. Biomechanics of fusion and stabilization. *Spine* 1995;20:85S–99S.
116. Skalli W, Lavaste F, Robin S, Dubousset J. A biomechanical analysis of short segment spinal fixation using a 3D geometrical and mechanical model. *Spine* 1993;18:536–545.
117. Ruland CM, et al. Triangulation of pedicular instrumentation. A biomechanical analysis. *Spine* 1991;16:S270–S276.
118. Stokes IA, Laible JP. Three-dimensional osseo-ligamentous model of the thorax representing initiation of scoliosis by asymmetric growth. *J Biomech* 1990;23:589–595.
119. Stokes IA, Gardner-Morse M. Three-dimensional simulation of Harrington distraction instrumentation for surgical correction of scoliosis. *Spine* 1993;18:2457–2464.
120. Gardner-Morse M, Stokes IA. Three-dimensional simulations of the scoliosis derotation maneuver with Cotrel-Dubousset instrumentation. *J Biomech* 1994;27:177–181.
121. Stokes IA, et al. Biomechanical simulations for planning of scoliosis surgery. *Res Spinal Deformities II, IOS Press* 1998;59:343–346.
122. Grealou L, Aubin CE, Labelle H. Biomechanical modeling of the CD instrumentation in scoliosis: A study of correction mechanisms. *Arch. Physiol Bioche* 2000;108:194.
123. Aubin CE, et al. Biomechanical modeling of posterior instrumentation of the scoliotic spine. *Comput Methods Biomech Biomed Eng* 2003;6:27–32.
124. Poulin F, et al. [Biomechanical modeling of instrumentation for the scoliotic spine using flexible elements: A feasibility study]. *Ann Chir* 1998;52:761–767.
125. Plouznikoff A, Aubin CE, Ozell B, Labelle H. Virtual reality scoliosis surgery simulator. *Inter Res Soci Spinal Deformities* 2004; 139–142.

See also HUMAN SPINE, BIOMECHANICS OF; REHABILITATION, ORTHOTICS IN; SPINAL IMPLANTS.

SCREEN-FILM SYSTEMS

ZHENG FENG LU
Columbia University
New York, New York

INTRODUCTION

Projection radiography generates the majority of the imaging volume in a typical hospital-based radiology department. Figure 1 illustrates the basic geometry for projection radiography. As shown, an X-ray beam incidents upon and transmits through the patient. Differential attenuation occurs as X-ray photons interact with tissues in the patient. Consequently, an altered X-ray distribution pattern is generated behind the patient. This altered X-ray pattern can be recorded by an image receptor placed underneath the patient.

Film has been utilized as an image receptor since the very beginning of radiographic imaging. The first radiograph of a living hand was exposed by Roentgen in 1895 using a photographic plate (1). Film is sensitive to both X rays and light photons. However, the sensitivity of film to direct X-ray exposure is low. In order to overcome this problem, the fluorescent screen was developed shortly after the discovery of X rays to be combined with the film as an image receptor. Fluorescence, as used in radiography, refers to the phenomenon that the absorption of X rays by crystals of certain inorganic salts (called phosphor) is followed by the emission of light photons (2). The emitted light photons are then utilized more efficiently to expose the film that is sandwiched between the screens (see Fig. 2). These fluorescent screens are also called intensifying screens. Initially, calcium tungstate was the phosphor of choice for screens in radiography (3). Over the years, especially since early 1970s, more efficient phosphor materials have been developed, such as rare earth phosphors (2,5-7). The "rare earth" elements are found in a row of the periodic table of the elements with the atomic numbers 57-71. The most common rare earth phosphor currently available is gadolinium oxysulfide (4). Phosphors, other than rare earth types, are used as well, such as yttrium tantalate (5-7).

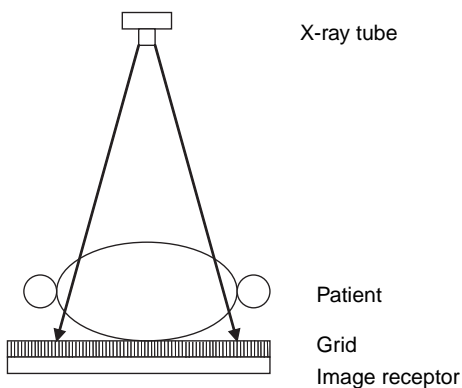


Figure 1. Illustration of the basic geometry for projection radiography.

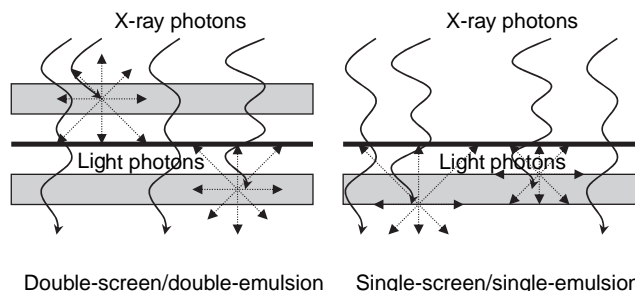


Figure 2. X-ray photons are absorbed by the intensifying screens and the emitted light photons from the screens expose the film to form the image.

Despite numerous improvements made in technology since the discovery of X rays, screen-film systems had remained, for a long period of time, as the main recording image receptors for radiography until recently. Since the 1980s, the essential role of the screen-film system has been challenged by the new invention of computed radiography (CR) and digital radiography (DR) (8). In comparison to screen-film systems, the CR and DR system has a much larger dynamic range. It has been reported that a linear response over more than four orders of magnitude of radiation exposure is achievable (2). This allows CR and DR systems to have a high tolerance for variations in radiation exposures. Therefore, the repeat rate due to underexposure or overexposure is significantly reduced. While the CR and DR systems provide tremendous potentials in image manipulation associated with digital imaging, their current spatial resolution is usually inferior to that of the screen-film systems (9,10). This explains the fact that the last stronghold of the screen-film system is in mammography, where the spatial resolution is on high demand. However, the digital technology continues to evolve, and the results of several large clinical trials of digital versus screen-film mammography suggest that the diagnostic accuracy of the digital mammography "is similar to" that of the screen-film mammography (11). In recent years, more and more radiology departments moved toward "filmless" with picture archiving and communication systems (PACS) (8). It is expected for DR and CR to become more widely available, and even replace screen-film systems completely in the near future (12).

In this article, we will outline the physical and photographic properties of intensifying screens and film. The steps for film processing will be delineated. Performance evaluation of the screen-film system will be discussed in relation to image quality and patient radiation dose. Finally, quality assurance and quality control testing will be described for the applications of screen-film systems in radiology.

INTENSIFYING SCREEN

Since only a small percentage (2-6%) of the X rays can be absorbed by the emulsion of the film, the sensitivity of the film alone is very low (2). With the screen-film combination,

X rays are absorbed more efficiently by the intensifying screen. After absorption, the screen furnishes a light image converted from the X ray image. The light image is then recorded on the film with much higher sensitivity. The main purpose of using the intensifying screen with a combination of film, instead of film alone, was to raise the sensitivity of this image receptor in order to reduce the radiation dose to patients. A factor of 50–100 dose reduction is achievable. This dose reduction is often depicted by the intensification factor that is defined as the ratio of the X-ray exposure required to produce a film density (e.g., 1.0 o.d. net density) without screens versus the exposure required to produce the same film density with screens (2):

$$\text{Intensification factor} = \frac{\text{exposure required without screen}}{\text{exposure required with screen}} \quad (1)$$

Also, lower exposure required by screen-film systems can be achieved by using a shorter exposure time that has the added benefits of minimizing patient motion artifacts and reducing X-ray tube heat capacity. However, these advantages are gained at the cost of a decrease in spatial resolution and an increase in noise on the image.

Physical Properties of Intensifying Screen

As shown in Fig. 3, an intensifying screen is composed of four layers (2): a base, a reflecting coat, a phosphor layer, and a plastic protective coat. The total thickness of a typical intensifying screen is ~ 0.3 – 0.6 mm.

The base is the screen support. The materials to make the screen base may be high grade cardboard or polyester plastic (2). The thickness of the base varies among manufacturers. The approximate range is 0.2–0.3 mm.

As described by Curry et al. (2), the reflecting coat is made of a substance, such as titanium dioxide (TiO_2), and is spread in a thin layer (~ 30 μm in thickness) between the phosphor layer and the base. Because light from the phosphor is emitted in all directions, including the direction away from the film, this reflecting coat acts to reflect those light photons back toward the film so that they may also contribute to the exposure on film. Although this increases the sensitivity of the screen, it also increases the blurring effect due to light spreading (shown in Fig. 3). Therefore, some manufacturers choose not to include a reflecting coat in the screen in order to improve image sharpness. This is normally the case for those screens that emphasize spatial resolution, such as those for bone imaging. Consequently, more X-ray exposure is needed to compensate for the loss in

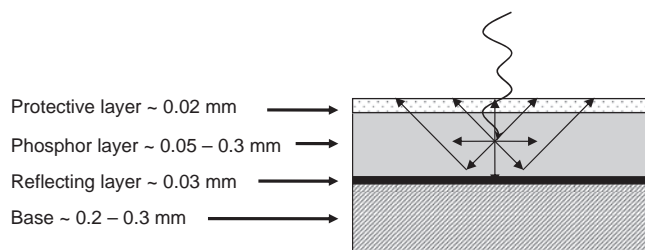


Figure 3. A cross-sectional view of an intensifying screen.

the reduced intensification factor of the screen if the reflecting coat is not added.

The phosphor layer, the key component of the screen, contains phosphor crystals that generate fluorescence. It is used to absorb X-ray photons and convert the absorbed X-ray energy into light photons that expose the film. The thickness of the phosphor layer varies in a range 0.05–0.3 mm, depending on the desired screen intensification factor. Usually, a thicker phosphor layer leads to a faster screen and a thinner phosphor layer leads to a slower screen.

The outmost layer consists of a protective coat that is applied over the phosphor crystals. This protective coat serves three major functions as outlined by Curry et al. (2): (1) it protects the delicate phosphor crystals; (2) it provides a surface that can be cleaned; (3) it helps to prevent static electricity that may generate image artifacts. This layer is usually made of very thin plastic.

Absorption Efficiency and Conversion Efficiency

The process of screen intensification is twofold. First, X-ray photons are absorbed by the phosphor, quantified by the absorption efficiency. Second, the absorbed X-ray energy is converted into light photons that expose the film, quantified by the conversion efficiency. The overall efficiency of a screen is the product of the absorption efficiency and the conversion efficiency, which leads to the intensification factor of the screen mentioned in equation 1. Raising either the absorption efficiency or the conversion efficiency or both can increase the screen intensification factor; thus raising the speed of the screen-film system.

The absorption efficiency, describing how efficiently the phosphor crystals absorb the incident X-ray photons, is a function of the X-ray photon energy and varies with the phosphor type (2–7,13–14). As we learn from the X-ray interactions with matter, X-ray photons are most likely to undergo the photoelectric effect when the X-ray photon energy is just slightly greater than the K-shell binding energy. This forms an abrupt increase in the X-ray photon absorption above the K-shell binding energy, called the K-edge. Obviously, it is desirable to have the K-edge of the phosphor coincide with the effective energy of the X-ray beam so that a high percentage of X-ray photons may interact with K-shell photoelectric effect. The effective energy of an X-ray beam is usually one-third to one-half of the peak in kilovolts. Most X-ray beams commonly employed in diagnostic radiology are operated in a kVp range of 50–120 kVp. Therefore, majority photons incident upon the image receptor are <60 keV. As a result, screen phosphors with K-edges <60 keV may have improved absorption efficiencies. As shown in Table 1 (7,13,14), the K-edge of the conventional calcium tungstate is 69.5 keV, above the energy of majority photons in a typical diagnostic X-ray beam spectrum, especially those X-ray beams with low kVp settings. In contrast, the K-edges of the others are <60 keV. Consequently, the calcium tungstate screen does not absorb X rays as efficiently as those screens with lower K-edges. Other factors that change the absorption efficiency are the thickness of the phosphor layer, the composition of the phosphor crystal, and

Table 1. Physical Properties of Some Common Phosphors^a

Phosphor	Atomic Number of Heaviest Element	K-Edge, keV	Conversion Efficiency, %	Light Emission Spectrum
Calcium tungstate, CaWO ₄	74	69.5	3.5	Blue (340–540 nm)
Barium strontium sulfate, BaSO ₄ :Eu	56	37.4	6	Blue (330–430 nm)
Barium fluorochloride, BaFCl:Eu	56	37.4	13	Blue (350–450 nm)
Gadolinium oxysulfide, Gd ₂ O ₂ S:Tb	64	50.2	15	Green (400–650 nm)
Lanthanum oxybromide, LaOBr	57	38.9	13	Blue (360–620 nm)
Lanthanum oxysulfide, La ₂ O ₂ S:Tb	57	38.9	12	Green (480–650 nm)
Yttrium oxysulfide, Y ₂ O ₂ S:Tb	39	17.0	18	Blue (370–630 nm)

^aSee Refs. 13 and 14.

grain size. A thicker phosphor layer will result in greater absorption efficiency. However, the disadvantage of doing so is the reduction in spatial resolution because light diffuses laterally as it propagates through the screen phosphor layer. The thicker the phosphor layer, the wider spread the light diffusion will be. For a single-screen single-emulsion system, the film is placed closer to the patient prior to the screen (shown in Fig. 2). This arrangement is based upon the fact that X-ray interaction with the phosphor reduces exponentially with the depth. Therefore, the majority of the X rays are absorbed in the top most region of the phosphor layer. Placing the intensifying screen underneath the film ensures a shorter light diffusion path. Consequently, a better spatial resolution is achieved.

The conversion efficiency is also increased from a calcium tungstate screen to a rare earth screen. As defined, the conversion efficiency is the fraction of absorbed X-ray energy converted to light energy for exposing the film. As shown in Table 1, the conversion efficiency varies according to the composition of the phosphor material.

An increase in conversion efficiency results in a greater intensification factor. However, since fewer X-ray photons are involved in image formation, an increase in conversion efficiency also results in an increase in quantum mottle in the image. Quantum mottle is the image noise caused by statistical fluctuation of the finite number of X-ray photons that form the image on film (2). As the conversion efficiency increases, the number of the X-ray photons detected by the intensifying screen reduces; thus the quantum mottle increases due to the reduced number of X-ray quanta that form the image. Unfortunately, an increased quantum mottle is associated with a decrease in low contrast detectability.

FILM

Physical Characteristics of Film

Figure 4 illustrates the cross-section of a film. The film base is to support the fragile photosensitive emulsion. Several conditions for film base materials have been outlined by Curry et al. (2) and include the following: (1) It must be transparent to visible light so that when a developed film is viewed (e.g., on a light box) the base will not interfere with the visual pattern (e.g., the image) recorded in the emulsion layer. (2) It must be strong enough to endure the film

developing procedure, and, yet at the same time, be flexible and easy to handle. (3) It must be physically stable, that is, the shape of the film base must not distort during the developing process and over the long period of film storage. Historically, glass was utilized for film base. Although it was transparent, glass was too fragile and difficult to handle. In 1914, cellular nitrate was adapted to replace glass for X-ray film. But cellular nitrate was flammable and caused possible fire hazards. Later in 1924, less flammable cellulose triacetate base was developed to replace cellular nitrate (2). In 1960, polyester plastic was first introduced to make the X-ray film base. Thereafter, film base has been made of a thin, transparent sheet of plastic; the thickness of the base being ~0.2 mm.

Firmly attached to the base through adhesive substances are photosensitive emulsion layers. Because the emulsion materials are delicate in nature, a supercoat layer is used to prevent damages to the emulsion. If emulsion layers are coated on both sides of the base, the film is called double-emulsion film. Sometimes, such as in mammography, only one side of the base is coated with emulsion. Thus, the film is called single-emulsion film. Intuitively thinking, the single-emulsion film would be less sensitive than the double-emulsion film because it has only one emulsion layer. However, it gains in spatial resolution. The thickness of the emulsion layer is usually very thin, ~10 μm.

The emulsion, the key component of the film, consists of two major ingredients: gelatin and grains made of silver halide. The gelatin is used as the binder for the silver halide grains in order to keep them well dispersed. Because the gelatin is stable in nature, it provides protection and stability for the film emulsion layer before and after the film development process that we will discuss in detail in the section Film Processing. Also, the gelatin allows easy penetration for the film-processing solutions (13).

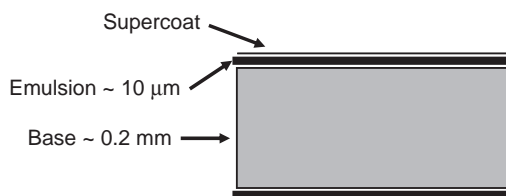


Figure 4. A cross-sectional view of an X-ray film.

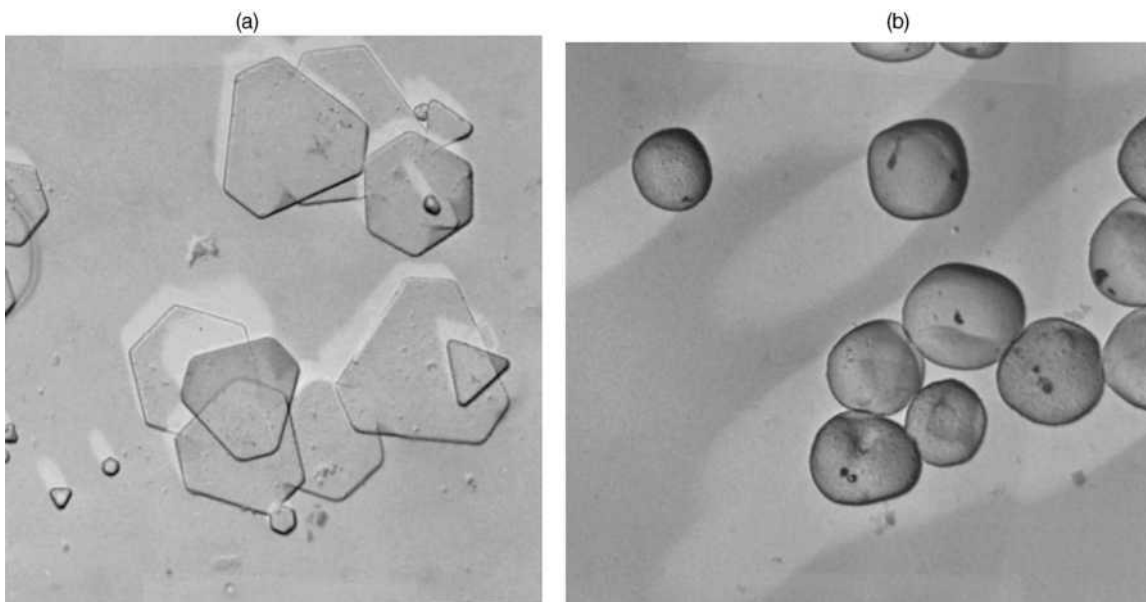


Figure 5. Photomicrographs of flat, tabular-shaped grains with the crystals orientated with the flat side parallel to the film base (a) and three-dimensional (3D) silver halide grains (b). (Courtesy of Dr. R. Dickerson of Eastman Kodak Company.)

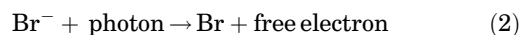
The light sensitive silver halide is in the form of small crystals whose sizes vary in the order of 0.1–1.0 μm in diameter (15). The grain sizes, size distribution, and their shapes play a major role in determining the film speed and contrast. Generally, the larger the grain size, the greater the film sensitivity will be. This may be changed if the grain shape varies at the same time. Figure 5 shows photomicrographs of two types of crystals in the film emulsion. One contains flat, tabular-shaped grains, and the other contains the conventional, 3D silver halide crystals. As explained by Dickerson (16), the use of tabular grains increases the ratio of the surface area to volume, which leads to a significant improvement in the film sensitivity. Also, the tabular grains provide better image sharpness by reducing the light “crossover”, which refers to the light emitted not by the screen in contact with the film emulsion, but by the screen opposite the film emulsion and passed through the film base. Newer technologies in 1990s developed a zero-crossover system that the film emulsion on one side of the film base was isolated from the film emulsion on the other side of the film base by adding a light-stopping dye to the film base (16,17). Obviously, the added dye must be removable during the film processing so that it will not be visible on the developed film.

Silver halide crystals can be grown in a variety of sizes and shapes by choice of emulsion precipitation conditions. Details of emulsion making were described by Wayrynen (15) and Dickerson (16). The film designers vary the emulsion grain morphology to meet various needs in film speed, resolution, contrast, and latitude.

The silver halide consists of, predominantly, silver bromide (AgBr) and a very small fraction of silver iodide (AgI) or silver chloride (AgCl), added as a sensitizer. All ions are bonded to form a cubic crystalline lattice.

Latent Image Formation

The silver halide crystals in the film emulsion also contain silver sulfide (Ag₂S) molecules randomly distributed on the crystal surface, which tend to trap free electrons from the bromide ion during X-ray exposure. The trapping site is called the sensitivity speck. This is the location where the latent image formation begins. The mechanism of latent image formation was initially described by the Gurney-Mott theory and has remained a topic of research by many over the years (18). The Gurney-Mott theory is accepted as incomplete, but basically correct. The theory describes the latent image formation as two steps: an electronic excitation and an ionic migration. First, a visible light photon interacts with a bromide ion, forming a bromide atom and releasing an excited mobile electron:



The bromide atom, Br, then migrates out of the crystal into the gelatin while the free electron becomes subsequently trapped at a sensitivity speck. The trapped electron (negatively charged) at the sensitivity speck attracts a mobile silver ion, Ag⁺, (positively charged), and the two combine to form a silver atom, Ag, at the sensitivity speck:



This single silver atom then acts as an electron trap for subsequent freed electrons. There is a continued accumulation of silver atoms at the sensitivity speck following repeated trapping of electrons and their neutralization with silver ions. At least 3–5 silver atoms must be accumulated at the sensitivity speck in order to form a valid latent image center that can become a clump of the silver atoms after the film is developed. The film darkening is produced by the accumulated silver atoms.

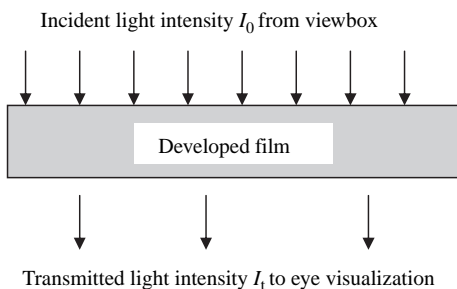


Figure 6. Illustration for the definition of the optical density of a developed X-ray film.

Usually, the film is developed immediately after the X-ray exposure is made. If the film is delayed in developing, the film speed and contrast will be reduced. This phenomenon is called latent image fading that can also be explained by the mechanism mentioned above.

Optical Density

In photographic print film, the more exposure the film receives, the brighter the print film will be. In screen-film radiography, it is the opposite because X-ray film is a negative recorder. An increased X-ray exposure produces darker film. The degree of darkness of the developed film is defined by a term called optical density or, simply, density (4). Shown in Fig. 6, if the incident light intensity from a viewbox, where the film is read, is given by I_0 , and the transmitted light intensity through the film for one particular spot is given by I_t , then the optical density (O. D.) at that location is given as the 10 based logarithm of the ratio of the incident light intensity versus the transmitted light intensity:

$$\text{O.D.} = \log_{10} \left(\frac{I_0}{I_t} \right) \tag{4}$$

For example, if the transmittance at one film location is 10%, then the optical density of that location is 1.0. Notice the relationship of the transmittance and the optical density is logarithmically compressed. Hence, if the optical density is 3.0, then the transmittance is only 0.1%. The optical density is often measured by a device called a densitometer. A densitometer has a small sensitive area (aperture), where the light intensity without film in the light beam— I_0 is measured; then with the film inserted and the region of interest covering the aperture, the densitometer, measures the light transmitted through the film— I_t . The optical density is read accordingly when the read-out button is pressed on the densitometer.

Table 2 lists a range of optical densities. Optical density 0.0 is reached only when the light transmission through the film is 100%, which is not practically possible. Even when a fresh unexposed film is processed, it will still have a low film density of ~ 0.15 O.D., which is known as the film base plus fog. The base density is caused by the absorption of light through the film base material. The current polyester film base is usually slightly blue-tinted in order to ease

Table 2. Optical Density Range of an X-Ray Film

Optical Density	Comments
0.0	100% transmission (practically impossible)
0.15–0.20	Typical base + fog
0.45–2.20	Useful density range
3.0	Very dark; needs a hot lamp for high luminance
> 3.5	Typical maximum darkness

the human eye visualization and help cancel distractions from faint chemical residue resulted from film processing. The fog density is mainly due to unintended exposure or the development of unexposed silver halide grains in the film emulsion. The fog may be affected by conditions related to the film storage and the film processing. It is desirable to keep the film base plus fog < 0.20 O.D. (19).

Table 2 lists the useful density ranges between 0.45 and 2.20 O.D. In this range, the human eye can appreciate the shades of contrast with a standard viewbox of which the luminance level is ~ 1500 nits. For mammography, in which films are generally darker and with more demands of low contrast perception, the viewbox luminance level should be at least 3500 nits (19). A brighter viewbox improves the ability to appreciate the contrast in the dark density regions. Sometimes, hot lamps are utilized to provide more intense incident light to view the film. It is typically utilized to check the dark regions on film such as those specific regions with optical densities > 3.0 O.D. The maximum darkness a film can reach is called D_{max} . It varies from film to film. A mammography film may reach a D_{max} of 3.5 O.D. or greater; while a regular film may barely reach 3.0 O.D. For comparison, print films are normally saturated at a density of 1.7 or 2.0 O.D. at best (2).

The Hurter and Driffield Curve (H&D Curve)

The relationship between the exposure to a film and the film density corresponding to that exposure can be plotted as the characteristic curve or H&D curve (named after Hurter and Driffield who were the first to use such a curve) (3). The exposure is plotted along the horizontal axis in

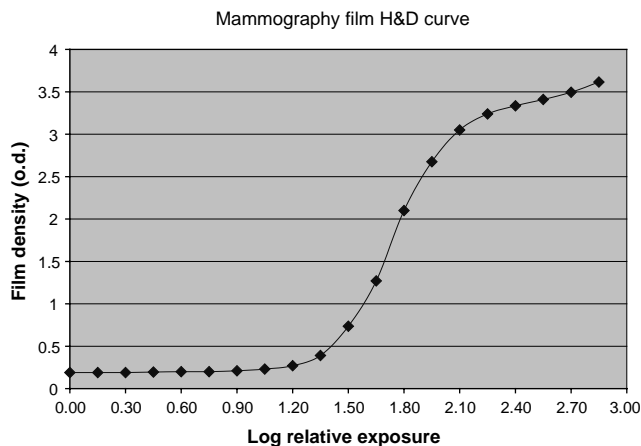


Figure 7. An example of H&D curve of an X-ray film.

Comparison of two films using light sensitometry

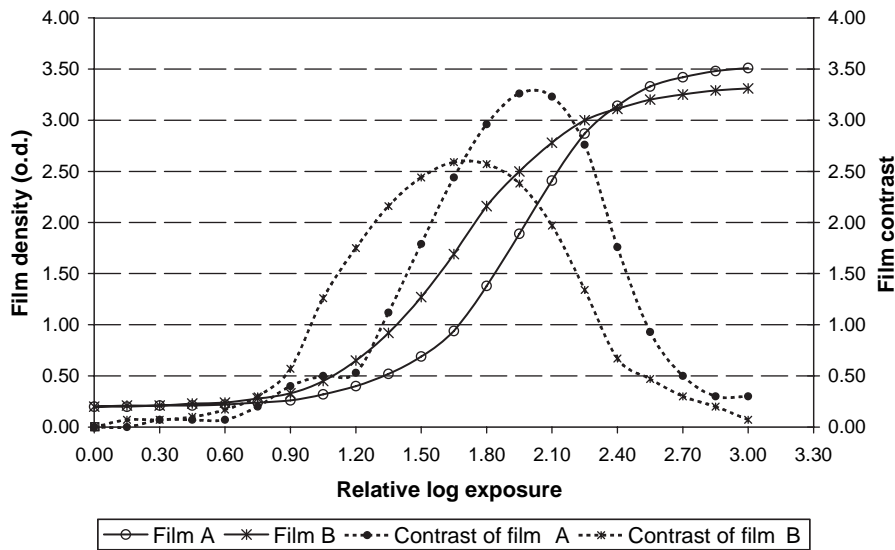


Figure 8. A comparison of two films using their H&D curves obtained by light sensitometry.

logarithmic scale, and the film density is plotted along the vertical axis. Figure 7 shows an example of the H&D curve of an X-ray film using light sensitometry, where relative exposures are utilized instead of actual exposures, because, in practice, the relationship between two exposures is often more important than the actual exposure. The H&D curve illustrates how much the optical density may change corresponding to the exposure change. As outlined by Bushberg et al. (4), the curve has a sigmoid shape, beginning with the base plus fog for no exposure, which cannot actually be plotted on a logarithmic scale, and the “toe” for the low exposure region. In the “toe”, the film density rises very slowly as the exposure increases. This region corresponds to “underexposed” film regions, such as the mediastinum on a chest radiograph. Beyond the “toe”, the film density escalates linearly to the increase of logarithmic exposure. This is called the “linear region” or the “straight-line portion” of the H&D curve. In this region, the film contrast is at its best and the film density fits in the useful density range as shown in Table 2. Ideally, a radiograph image should be exposed with proper technique to have densities within this region. Further beyond the “linear region”, the curve flattens out. This region is called the “shoulder”. It corresponds to the “overexposed” region where the film contrast is compromised.

Film Contrast and Latitude

The image contrast in film radiography is shown as the difference in film densities at various locations on the image. It is contributed by both the subject and film contrast. The subject contrast is determined by the difference in X-ray intensity after transmitting through various regions of the patient. It is primarily determined by the differential attenuation of tissues. For example, a solid tumor may attenuate more X rays than its surrounding tissue; thus leaving a lower X-ray intensity behind the tumor. The subject contrast relies on factors such as kVp,

differences in patient part thickness, atomic number, and density of tissues. Film contrast represents the ability of the film to manifest the subject contrast into the image contrast on film. Choosing the correct film will enhance the subject contrast on the final image. The film contrast is determined by the slope of the H&D curve: a steeper slope leads to a higher film contrast. For a given H&D curve, the slope is a function of film density, being low in the toe region, peaking in the straight-line portion, then being low again in the shoulder region (shown in Fig. 8). Based upon the H&D curve, the film contrast is often quantified by two commonly used parameters: gamma and average gradient. Gamma is defined in the straight-line portion by the maximum slope of the H&D curve:

$$\gamma = \frac{D_2 - D_1}{\log E_2 - \log E_1} \quad (5)$$

where D_2 and D_1 outline the straight-line portion of the H&D curve and E_2 and E_1 are exposures needed to produce the film densities of D_2 and D_1 . The average gradient is often calculated over a density range of 0.25–2.0 O.D. both above the base plus fog, because such a density range is considered the useful density region:

$$\text{Average Gradient} = \frac{D_S - D_B}{\log E_S - \log E_B} = \frac{1.75}{\log E_S - \log E_B} \quad (6)$$

where D_S is 2.0 O.D. above the base plus fog and D_B is 0.25 O.D. above the base plus fog. The parameters E_S and E_B are exposures needed to produce the film densities of D_S and D_B .

Very often, a special term called “latitude” is also cited for film photographic characteristics. The latitude is usually defined as the exposure range that produces a certain density range over which a usable image can be made on film (20). If this density range is between the upper and lower optical density limits on an image, then the corresponding exposure range is also called the

“dynamic range”. Obviously, the film latitude changes inversely with the film contrast. Films with wide latitude exhibit lower film contrast than films with narrow latitude. Sometimes, a wide latitude is preferred in spite of compromising the film contrast. For example, in chest imaging, wide latitude films are desirable because of the large dynamic range needed to cover areas behind lungs as well as mediastinum. In other words, a wide latitude film is capable of keeping the lung regions below the shoulder and the mediastinal region above the toe. With the new inventions of digital imaging (e.g., CR, DR), the wide linear dynamic range of these technologies has proved to be advantageous and ultimately solve the conflict between the film latitude and contrast.

The characteristic curves are often employed to compare films regarding various properties including speed, contrast, base plus fog, and maximum density. Figure 8 shows the characteristic curves of two films. The H&D curves are shown in solid lines and the film contrasts as functions of the film density are shown in dashed lines. The graph has dual vertical axes with the left axis for the film density and the right axis for the film contrast. The graph demonstrates Film B has a faster speed, but a lower contrast than Film A. Both have similar base plus fog. Film A can reach a higher D_{max} than Film B. Note that the characteristic curves shown in Fig. 8 are from light sensitometry. A device called a sensitometer is utilized to expose the film in a “step” fashion with a range of constant light intensities that each differs by a factor of the square root of 2 (i.e., 0.15 increment on the logarithmic exposure scale of the H&D curve). The sensitometer simulates the light emitted from the intensifying screens. Nevertheless, because the simulated light is not the same as the emitted light from screens, the response of film to the sensitometer may differ from that to the actual light from the intensifying screens caused by X-ray exposure. Special cautions are needed when comparisons are made based upon the light sensitometry (21).

The contrast of the image receptor of the screen–film system, as a whole, is primarily dependent on its film contrast; although there are exceptions, such as Kodak InSight asymmetric film–screen systems, of which the contrast can be varied by changing the intensifying screens (16,17).

Spectral Emission and Spectral Sensitivity

Obviously, the wavelength of light emitted by an intensifying screen should correspond closely with the spectral sensitivity of the film used with the screen. Otherwise, the total photographic effect is decreased, and the patient has to suffer more radiation exposure to get the proper film density. Usually, the sensitive spectrum of the film is matched to cover all the wavelengths emitted by the screen in order to maximize the speed of the screen–film combination. Shown in Fig. 9 is a spectral match of blue sensitive X-ray film with the calcium tungstate spectral emission that is continuous with a peak in the blue region (16). The invention of rare earth phosphors made the intensifying screen more efficient in both absorption and conversion of X-ray photons. However, the luminance from the majority

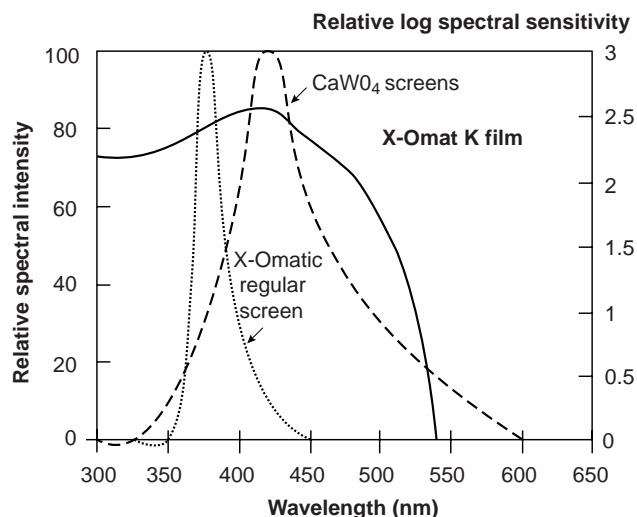


Figure 9. Relative spectral match of Kodak X-Omat K film with ultraviolet (UV) and blue-emitting screens. (Courtesy of Dr. R. Dickerson of Eastman Kodak Company.)

of the rare earth screens was changed to the color green, which made it necessary to change the spectral sensitivity of X-ray films accordingly. The extension of the film spectral sensitivity to longer wavelengths was achieved by adding spectral sensitizing dyes absorbed by silver halide (13). Figure 10 shows a spectral match of green sensitive X-ray film with the spectral emission of a rare earth intensifying screen. Note that, in Fig. 10, despite the fact that the screen emission spectrum was beyond the green region, the film spectral sensitivity was topped at the green region in order to have a safelight zone in the color red. The safelight in the darkroom is for the convenience of handling the film without film fogging. For many years, amber safelights were utilized to handle blue sensitive X-ray films. As the upper limit of the film spectral sensitivity moved up from the blue region to the green region, the

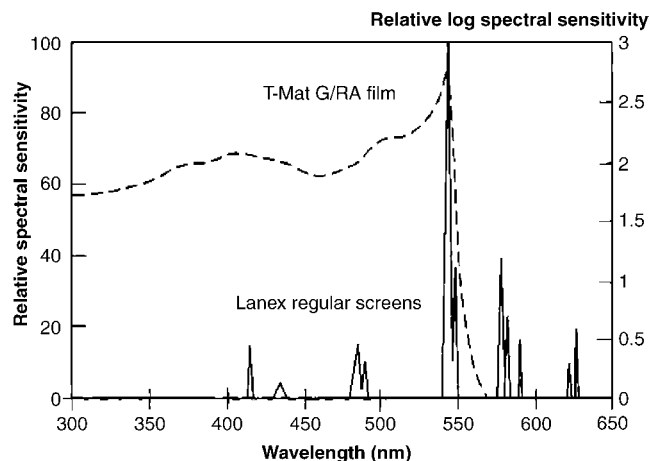


Figure 10. Relative spectral match of Kodak T-Mat G film with Kodak Lanex regular screens. (Courtesy of Dr. R. Dickerson of Eastman Kodak Company.)

darkroom safelight was necessary to change from the amber to the red safelight.

Speed and Resolving Power

The speed of a film, also known as the system sensitivity, is defined in the unit of $1/R$ as the inverse of the exposure that produces a net film density of 1.0 O.D. above the base plus fog. Because film is utilized in combination with intensifying screens, the speed commonly quoted is actually the speed of the screen–film system. Although with a faster film, the speed of screen–film system may be increased accordingly, the determining factor for the speed of a screen–film system is in the screen not in the film. For a screen–film system, the speed can be compared by examining the location of the H&D curve along the horizontal logarithmic exposure scale at a net density of 1.0 O.D. above base plus fog. In general, when a horizontal line is drawn at 1.0 O.D. above the base plus fog, the curve that appears toward the left on the exposure scale has faster speed than those that appear toward the right (see Fig. 8). Apparently, this speed depends on a number of variables, especially kVp and film processing conditions. The American National Standards Institute (ANSI) has published guidelines to standardize the method for measuring the speed of screen–film systems (22). The standard method is rigid in defining the exposure and the controlled film-processing conditions. Unfortunately, such absolute measured results are not widely available. In practice, a concept of “speed class system” has been commonly accepted for comparing the relative speeds among various screen–film systems (2). The specified speed class differs in value by $\sim 25\%$, which is similar to the camera speed class adapted in photography. For example, the sequence of the numbers in speed class starting from 100 is 100, 125, 160, 200, 250, 320, 400, 500. A great number in speed class means a fast screen–film system.

A faster screen–film system is mainly due to a greater intensification factor in the screen, which is often achieved by utilizing a thicker screen phosphor layer that leads to a reduction in resolving power. A general comparison of screen speed and resolving power is shown in Table 3.

Reciprocity Law

The relationship between the optical density and the X-ray exposure depicted in the H&D curve is usually not dependent on the exposure rate. The film density remains the same as long as the amount of the X-ray exposure received by the screen–film system is the same for a wide range of exposure rates. This phenomenon is known as reciprocity law. It allows the X-ray technique to vary in tube current

milliamperes (mA) and exposure time as long as the product of milliamperes and exposure time is constant.

However, the reciprocity law has been known to fail at very short (i.e., the exposure rate extremely high) or very long exposure (i.e., the exposure rate extremely low) with screen–film systems. For example, the film speed is observed to be reduced in mammography where very long exposure time (>2 s) has to be utilized for large or dense breasts (23). The reciprocity law failure can be explained by the latent image formation mechanism described earlier as the Gurney and Mott theory (18). In the case of extremely low exposure rate, the rate of photon absorption is too low to allow the gathering of enough silver atoms at the sensitivity speck to make it stable. Only one or two silver atoms at a sensitivity speck may disintegrate; thus the latent image may disappear before enough silver atoms can be gathered. Therefore, the speed is reduced at the extremely low exposure rate. In the opposite case, where the exposure rate is extremely high, the production of free electrons is too fast for all the silver ions to migrate to the proper sensitivity specks. Therefore, a fraction of the silver ions may not be utilized in forming silver atoms. This also leads to the reduction in film speed. For X-ray units, the automatic brightness control (ABC) on the system, or sometimes referred to as the phototimer, is a feature designed to maintain the consistency in film density regardless of the patient thickness. It is also designed to compensate for the reciprocity law failure.

The reciprocity law failure applies only to screen–film systems, but not to direct exposed film systems.

THE SCREEN–FILM CASSETTE

Figure 11 shows images of two screen–film cassettes. The double emulsion film is sandwiched between the two intensifying screens. In some applications, such as mammography, a single emulsion film is utilized with only one intensifying screen in the cassette. The front of the cassette is X-ray transparent usually made of materials with low atomic numbers, while the back of the cassette often includes a sheet of lead to reduce backscattering. The cassette is light tight so that the film, once being shut into the cassette, will not be exposed to ambient light. Opening or closing the cassette has to take place in a darkroom or other dark environment. The screens are usually mounted on layers of compressible foams so that when the two cassette halves close, air is expelled from the space between the screens and the film. This is to ensure good image quality by having a firm film screen contact without air trapping in between.

FILM PROCESSING

After a latent image is formed on the film by X-ray exposure, the film is subjected to a sequence of chemical processes in order to make the latent image visible to human eyes. This is film processing or film development (21).

Films had been developed manually in the darkroom until the first automatic film processor was introduced in 1956 (4). An automatic film processor (see Fig. 12) consists

Table 3. Physical Characteristics of Screen–Film Systems.

Type of screen–film system	Speed Class	Resolving Power, lp–mm ⁻¹
Film alone		~ 100
Fine detail	<100	$\sim 10\text{--}15$
Par-speed	100	$\sim 8\text{--}10$
Regular	200–400	$\sim 6\text{--}8$
High speed	~ 800	$\sim 4\text{--}6$

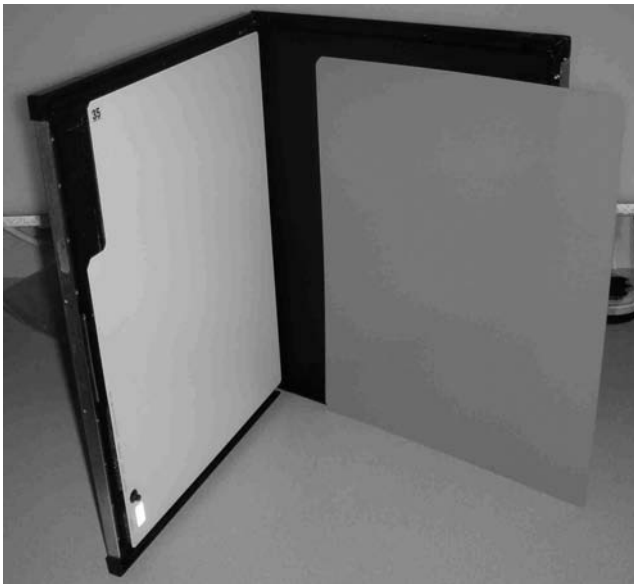
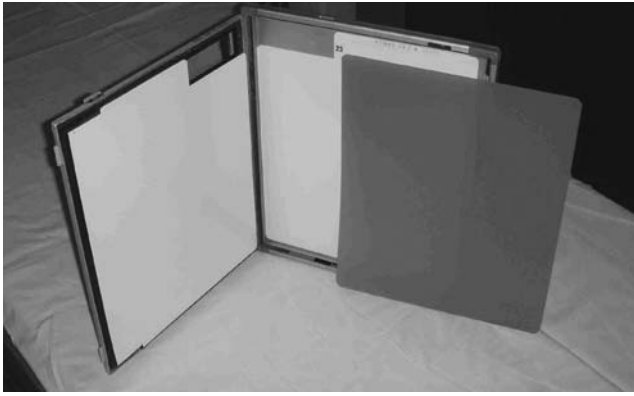


Figure 11. Pictures of the screen–film cassettes: (a) a double-screen double-emulsion cassette; (b) a single-screen single-emulsion cassette.

of three tanks: developer, fixer, and water to wash off the chemicals from the film. The film is then dried in a drying chamber and ready for reading at the drop slot. A standard cycle takes ~ 90 s for film processing. There are rapid cycles that take less time and extended cycles that need longer time. Rollers are utilized to transport the film along the circuitous path through each of these tanks. These rollers need to be serviced periodically to prevent roller transport processing artifacts (24). Film artifacts refer to those features on the image that do not reflect the true subject being imaged. The film artifacts are highly undesirable because they distract or even mislead readers. Although artifacts may be generated by factors other than the film processing such as X-ray equipment, patient positioning, film storage and handling, and so on, the film-processing artifacts are among the most commonly seen artifacts in routine clinical images. Details of the film artifacts, and how to battle against these artifacts, can be found in Ref. 21 Chapt. 6.

As the film enters the developer tank, the developer acts as the reducing agent that selectively reduces those



Figure 12. A picture of an automated daylight film processor.

exposed grains to silver. For those silver halide crystals that have been exposed to photons, latent image centers have been formed, each center having at least 3–5 silver atoms, and typically having 10 or more silver atoms. These silver atoms act to catalyze the further reduction of the silver in the presence of the reducing agent which, in this case, is the developer. After development, those exposed silver halide crystals become dark silver grains. One sensitivity speck can be converted to $\sim 10^9$ silver atoms, which gains an enormous amplification (16). They contribute to the O.D. of the film. A great O.D. region has a high concentration of such silver grains; a light O.D. region has a low concentration of such silver grains. An ideal developer should have only the exposed grain developed and unexposed grain washed out. But realistically, unexposed silver halide grains that do not contain latent images may also be developed. This causes “fog”.

The film developing time and temperature of the developer must be optimized because the film speed and the film contrast are directly affected by both factors. Usually, the speed and the contrast increase, up to a certain degree, by the increase of the development time. Then the speed and the contrast reach a plateau or may even decrease with further increase of the development time. Similar effect is observed for the development temperature on the speed and the contrast. Also, note that film fog increases with the increasing development time or temperature, which is not desirable. Therefore, it is important to optimize these film-processing conditions in order to achieve a reasonable compromise among film speed, contrast, and fog.

A fixer is used to terminate the development process and dissolves unexposed silver halide grains without damaging the image formed by metallic silver already developed by

the developer. An incompletely fixed film is easily recognized because it has a "cloudy" appearance. This is a result of the dispersion of transmitted light by those very small silver bromide crystals that have not been removed by the fixer.

Washing the film thoroughly with water is necessary. Otherwise the chemical residues left on the film will turn the film into brown as it ages. This is the function of the water tank after the developer and the fixer.

The last step for film processing is drying. Warm air is blown over both surfaces of the film when it is transported by rollers through the drying chamber. Finally, the dried film is delivered at the drop slot of the automated film processor.

QUALITY ASSURANCE AND QUALITY CONTROL (QA/QC) FOR SCREEN-FILM SYSTEMS

There are many aspects of QA/QC protocols for screen-film system maintenance and film processing. We will briefly discuss several important tests for routine QA/QC in the clinical environment.

The film processor needs day-to-day quality control in order to ensure consistent performance. The major concern in film processing is the instability due to wet chemistry variation and temperature fluctuation. The film sensitometry is a method to evaluate and monitor the performance of a film processor on a daily basis. As an important step in film processor QC, a light sensitometry strip (see Fig. 13) is made by a sensitometer and charted in the morning before any patient images are processed. Needed for film sensitometry are a sensitometer, a box of control films that are identical so that variations among films are negligible, and a densitometer to measure the film density. In addition, a thermometer is utilized to measure the developer temperature to ensure the optimal temperature according to the manufacturer's guideline. Three basic concerns for film sensitometry are (1) the base plus fog; (2) the speed

index; (3) the contrast index. The base plus fog is often measured at an unexposed area of the sensitometry film. For the speed index, a step is often designated with a mid-gray film density such as a film density of no less than, but close to 1.2 O.D. for mammography (19). For the contrast index, two steps are designated with one step of high film density, such as a film density ~ 2.20 O.D. and the other of low film density such as a film density of no less than, but close to 0.45 O.D. (19). The difference in density between these two steps is called the contrast index. The aim values of these three indexes are established over a period such as 5 consecutive days to smooth out normal day-to-day variation at the initial film processor installation or after a thorough preventative maintenance and calibration on the film processor. Day-to-day observation compares the measured indexes with the aim values. Figure 14 demonstrates sensitometric data for a film processor within one month. Records like these allow the tracking of the performance trends. The purpose is to eliminate problems in processor performance before those problems affect patient images. In mammography, 0.10 O.D. variation from the aim value triggers the action level for further monitoring and correction, and 0.15 O.D. variation demands immediate corrective actions before patient films can be processed (19).

There are certain screen-film cassette maintenance issues. First, the intensifying screens need to be kept clean. Artifacts may appear on images in the presence of dust on the screens. In mammography, screen clean-up is required at least once a week (19). In general radiography, screens are recommended to be cleaned at least on a quarterly basis. The cleaning procedure involves a solution containing an antistatic compound and a detergent. With a soft lint-free cloth, the screen surface should be wiped very gently with the solution. Then the cassette should be left open to air-dry. Another important issue of the screen-film cassette maintenance regards screen-film contact. A good screen-film contact is essential to prevent loss of spatial resolution. A simple method for testing the screen-film contact is to image a piece of wire pattern such as a mesh phantom placed on the top of the cassette. The sharpness of the wire pattern in all regions of the filmed image should be examined. If the screen-film contact is poor in certain areas, fuzziness or slightly increased density will be observed at those locations.

Finally, an important routine QA/QC test in a film-based imaging environment is related to darkroom fog. The darkroom fog degrades the image quality unnecessarily and should be prevented. Remedies to eliminate darkroom fog are straightforward, involving sealing the darkroom from light leaks (e.g., the door frame, the film processor, the film pass-box and any openings into the room) and following instructions on safelight type, bulb wattage and safelight positioning (19,25,26). Such easily preventable problems are often unnoticed or neglected (26). Therefore, it needs to be emphasized to have the darkroom fog tested during the initial film processor installation and monitored thereafter periodically (e.g., semiannually). To measure the darkroom fog, we may expose a film with a phantom that would produce a mid-gray film density when the film is developed (19). In the darkroom before

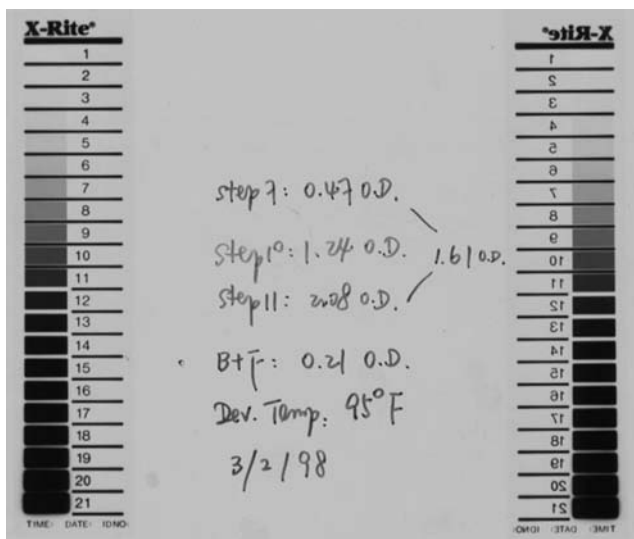


Figure 13. An example of a film sensitometry strip made by a sensitometer.

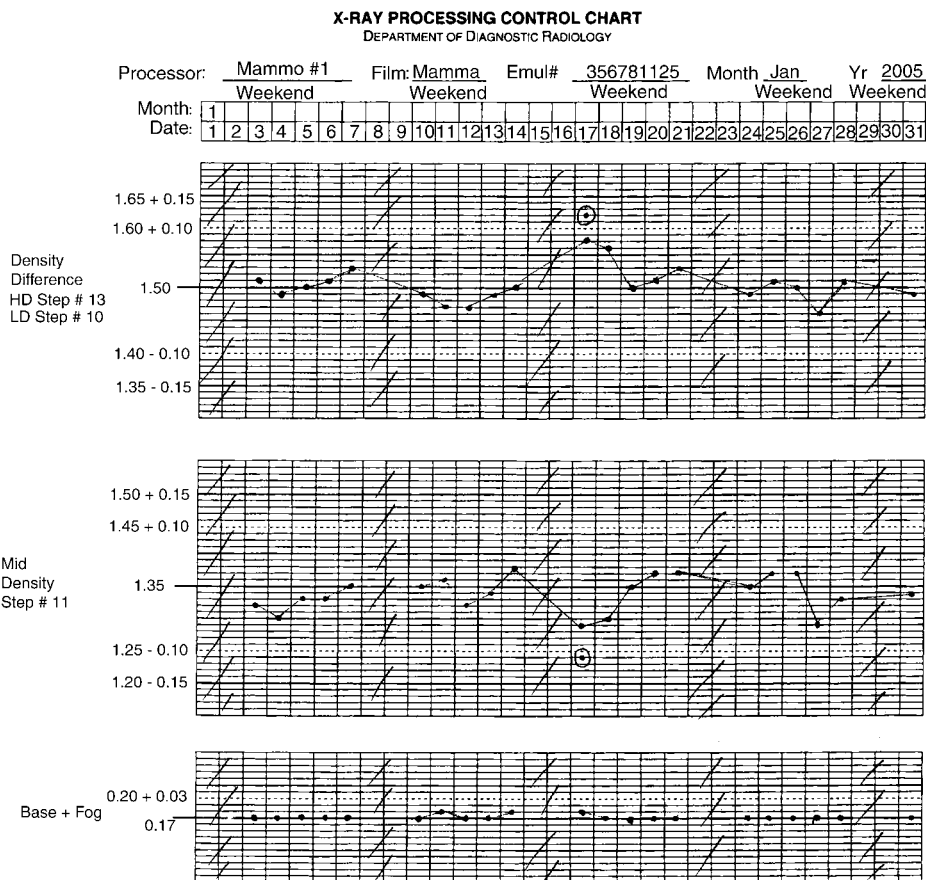


Figure 14. A chart of a film processor through a period of 1 month using the sensitometry method.

processing the film, place the film on the darkroom counter and bisect the latent image by using an opaque paper so that one-half of the film is protected from and the other exposed to any possible fogging sources in the darkroom. The covering line should be parallel to the anode-cathode direction of the X-ray tube in order to prevent any density difference due to the heel effect. Turn on the safelight in the darkroom. After 2 min of such fogging exposure, develop the film. The density difference of the two portions of the film near the covering line determines the darkroom fog. Usually, the darkroom fog should not exceed 0.05 O.D. (19).

ACKNOWLEDGMENTS

The author is grateful to Dr. Robert Dickerson at Eastman Kodak Company for many insightful suggestions and to Ms. Maryellen Peinelt for assistance in preparation of the article.

BIBLIOGRAPHY

1. Glasser O. Wilhelm Conrad Röntgen and the Early History of the Roentgen Rays. San Francisco: Norman Publishers; 1989.
2. Curry III TS, Dowdey JE, Murry Jr RC. Christensen's Physics of Diagnostic Radiology. 4th ed. Philadelphia: Lea & Febiger; 1990. Chapt. 9-11. p 118-164.
3. Ter-Pogossian MM. The Physical Aspects of Diagnostic Radiology. New York: Harper & Row; 1967. Chapt. 6. p 185-240.

4. Bushberg JT, Seibert JA, Leidholdt Jr EM, Boone JM. The Essential Physics of Medical Imaging. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 2002. Chapt. 6-7. p 145-189.
5. Buchanan RA. An improved X-ray intensifying screen. IEEE Trans Nucl Sci 1972;NS-19:81-86.
6. Buchanan RA, Finkelstein SI, Wickersheim KA. X-ray exposure reduction using rare earth oxysulfide intensifying screens. Radiology 1972;105:185-190.
7. Stevels AN. New phosphors for X-ray screens. Medica Mundi 1975;20:12.
8. Huang HK. PACS Basic Principles and Applications. New York: John Wiley & Sons Inc.; 1999. Chapt. 4. p 63-90.
9. Sanada S, Doi K, Xu X, Yin F, Giger ML, MacMahon H. Comparison of imaging properties of a computed radiography system and screen-film systems. Med Phys 1991;18: 414-420.
10. Samei E, Flynn MJ. An experimental comparison of detector performance for direct and indirect digital radiography systems. Med Phys 2003;30:608-622.
11. Pisano ED, Yaffe MJ. Digital mammography. Radiology 2005;234:353-362.
12. Samei E, Seibert JA, Andriole K, Badano A, Crawford J, Reiner B, Flynn MJ, Chang P. AAPM/RSNA tutorial on equipment selection: PACS equipment overview: general guidelines for purchasing and acceptance testing of PACS equipment. Radiographics 2004;24(1):313-334.
13. Arnold BA. Physical characteristics of screen-film combinations. In: Haus AG, editor. The Physics of Medical Imaging, Recording System Measurements and Techniques. New York: American Institute of Physics; 1979. p 30-71.
14. Thomason C. Screen-film systems. In: Webster JG, editor. Encyclopedia of Medical Devices and Instrumentation. New York: John Wiley & Sons Inc.; 1988. p 2599-2609.

15. Wayrynen RE. The photographic process. In: Haus AG, editor. *The Physics of Medical Imaging, Recording System Measurements and Techniques*. New York: American Institute of Physics; 1979. p 1–15.
16. Dickerson R. Fundamental of silver halide film design. In: Haus AG, editor. *Advances in Film Processing Systems Technology and Quality Control in Medical Imaging*. Medical Physics Publishing; 2000. p 73–84.
17. Pizzutiello Jr RJ, Cullinan JE. *Introduction to Medical Radiographic Imaging*. Rochester (NY): Kodak Publication M1-18; 1993. Chapt. 4–5. p 71–122.
18. Mees CEK, James TH. *The Theory of the Photographic Process*. New York: Macmillan; 1966. Chapt. 5. p 87–119.
19. *Mammography Quality Control Manual*, by ACR Committee on Quality Assurance in Mammography chaired by Hendrick RE, ACR; 1999.
20. Meeson S, Young KC, Rust A, Wallis MG, Cooke J, Ramsdale ML. Implications of using high contrast mammography X-ray film-screen combinations. *Br J Radiol* 2001;74:825–835.
21. Haus AG, Jaskulski SM. *The Basics of Film Processing in Medical Imaging*. Medical Physics Publishing; 1997. Chapt. 3, 5, 6.
22. American National Standards Institute: *Method for the Sensitometry of Medical X-Ray Screen-Film-Processing Systems*. New York: (ANSI PH2.43-1982); 1982.
23. Almeida A, Sobol WT, Barnes GT. Characterization of the reciprocity law failure in three mammography screen-film systems. *Med Phys* 1999;26:682–688.
24. Widmer JH, Lillie RF. Roller transport processing artifact diagnosis. In: Haus AG, editor. *Film Processing in Medical Imaging*. Medical Physics Publishing; 1993. p 115–129.
25. Suleiman OH, Showalter CK, Gross RE, Bunge RE. Radiographic film fog in the darkroom. *Radiology* 1984;151(1): 237–238.
26. Gray JE. Mammography (and radiology?) is still plagued with poor quality in photographic processing and darkroom fog. *Radiology* 1994;191:318–319.

See also X-RAYS: INTERACTION WITH MATTER.

SENSORS, GLUCOSE. See GLUCOSE SENSORS.

SENSORS, OPTICAL. See OPTICAL SENSORS.

SENSORS, PIEZOELECTRIC. See PIEZOELECTRIC SENSORS.

SENSORY AIDS. See BLIND AND VISUALLY IMPAIRED, ASSISTIVE TECHNOLOGY FOR.

SEXUAL INSTRUMENTATION

KIRK A. BRUNSWIG NEWRING
 CRISSA DRAPER
 WILLIAM O'DONOHUE
 University of Nevada
 Reno, Nevada

INTRODUCTION

From the annals of measurement comes an item of lore on the origins of measurement, occurring when two male

hominids were urinating side by side arguing over whose was bigger. More recently, it is told in the pubs of Ireland that the yardstick was invented in Dublin to settle such disputes.

As Semmlow (1) stated in the first edition of this Encyclopedia, sexual instrumentation is a rapidly evolving science. (*Note:* This article does not endorse an instrument or instruments. None of the authors have received compensation from a manufacturer of any of the products described in this work.) Many early developments in the arena of sexual instrumentation relied on “nonspecific responses associated with increased activity in the sympathetic nervous system”, such as heart rate and blood pressure. However, more recent investigations have turned to more specific physiological responses, generally focusing on the physiological responses of the genitals. While the scope of this entry prohibits a full description of the history and developments in the assessment and interventions related to the human sexual response, interested readers are welcome to refer to the works of Kinsey and Masters and Johnson for a more detailed review of the history of human sexuality research.

The need for and sophistication of measurement of human sexual behavior has improved dramatically over the years; however, there are still several scientific, methodological, and statistical challenges facing those researching the measurement of human sexual behavior. The current entry will review these measurements along with instruments and procedures related to female and male sexual behavior. Within each sex, the measures related to assessment and diagnosis of sexual concerns are reviewed first, followed by a review of devices used in treatment.

INSTRUMENTS AND MEASUREMENT OF FEMALE HUMAN SEXUAL BEHAVIOR

Formal research into physiological aspects of female sexual behavior was spurred in a large part by the work of Masters and Johnson in the 1960s. They highlighted the importance of genital vasocongestion as an indicator of sexual arousal. Later researchers explored the multidimensional nature of female arousal, often ascribing three necessary aspects: physiological, cognitive, and emotional. That is, physiological arousal is a necessary condition for sexual arousal, though it alone is often insufficient; however, early research into women's sexual behavior focused almost exclusively on the physiological aspects; detailing the physiological changes across the human female sexual response of arousal, plateau, orgasm, and resolution. The absence of an integrative research approach toward the assessment and treatment of female sexual response is discouraging, as this is a necessary condition for sexual arousal, and it seems to be a common problem among women. Estimates range from 1 in 3 to 1 in 5 women between 18 and 59 years of age complain of a lack of interest in sex.

A series of studies in the 1980s evaluated the normative aspects of human sexuality research volunteers. Through

these studies, it became clear that both male and female university-based research participants were decidedly different than their college peers. Typically, participants were more likely to participate in human sexuality research if they were to remain clothed. Also, those who were willing to participate were more likely to masturbate, to have had early and more frequent exposure to commercial erotica, to have less sexual fear, and (for female participants) to have a history of being sexually abused. Thus, for the early research studies listed, restraint should be used when extrapolating the findings discussed to dissimilar populations. More recent research has employed intent-to-treat and wait list control approaches in community-based research. These approaches have increased the ecological validity of this line of research, and this trend should continue as researchers increasingly use more ecologically valid methodology.

FEMALE SEXUAL BEHAVIOR ASSESSMENT

Internal Devices

In the late 1960s and 1970s, several devices were constructed to assess physiological components of female sexual response. One early device measured vaginal blood flow using two thermistors mounted on a diaphragm ring (please refer to later section on thermistor clip for the use of temperature measurement in female sexual behavior for more information). Other early devices included those intended to measure vaginal pH, temperature, and genital engorgement and blood flow properties via photoplethysmography. Common to all of these early instruments was a lack of standardization. Of these instruments, photoplethysmography received the most empirical attention.

Vaginal Photoplethysmography. Hatch's 1979 (2) review of vaginal photoplethysmography (VPPG) provides a description of the device as well as the issues and controversies surrounding its use at the time. The VPPG is described as a cylindrical probe approximately the size and shape of a menstrual tampon which is easily inserted into the vagina by the subject. A light source contained within the clear acrylic probe is directed at the wall of the vaginal lumen. A photoelectric transducer is also situated on the probe in such a position that it detects only that fraction of the incident light which is reflected from the vaginal tissue and the blood circulating within it. Changes in blood volume within the vagina produce changes in the amount of the incident light backscattered to the light detector, because of the large difference in transparency between blood and bloodless tissue. Changes in blood volume can therefore be easily recorded as changes in the output of photoelectric transducer (p. 358) (Figs. 1 and 2).

Hoon et al. (5) provide a description of a VPPG (including construction, instruction, recommended calibration procedures, and diagrams) and note there is a lack of standardization among VPPG, indicating that differences in probe size may have impacted cross-study comparisons, as larger probes are thought to reduce movement at the expense of



Figure 1. Vaginal photoplethysmograph similar to that originally designed by Sintchak and Geer (3) using a single light source and photodetector. (Courtesy of Farrall Instruments, Inc., Grand Island, NE.)

possible tissue stretching, blood supply occlusions, reduced sensitivity, and decreased acceptability by women.

One of the challenges with VPPG is the choice of current: alternating current (ac) or direct current (dc). Direct current coupling is used to assess slowly developing changes in vaginal blood volume (VBV), which is suggested as reflecting the pooling of blood in the vagina. Alternating current coupling is used to assess vaginal pulse amplitude (VPA), which is the observation of the arrival of the pulse wave at the observation point on each cardiac cycle. While some researchers approach the ac/dc issue as an either/or issue (2), more recent research (5) has measured both VBV and VPA. In reviewing the controversy surrounding the sensitivity of VBV and VPA, Hatch concludes, "there is some evidence suggesting that VPA is the more sensitive variable in the sense that it has been shown to significantly discriminate among responses in various types of erotica in some cases where VBV has not" (Ref. 2 p. 359).

Geer et al. (6) presented one of the earlier reports on the use of VPPG to demonstrate differential vaginal states following exposure to erotic and nonerotic visual stimuli. In their study, VPPG assessed VBV and VPA. However, they were one of many research groups to find that physiological arousal did not necessarily correlate with subjective ratings of arousal. Similarly, van Dam et al. (7) used VPPG to compare women with and without amenorrhea. The authors theorized and found that women with

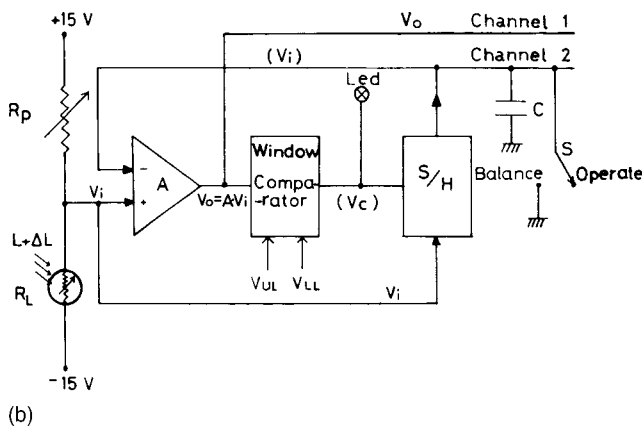
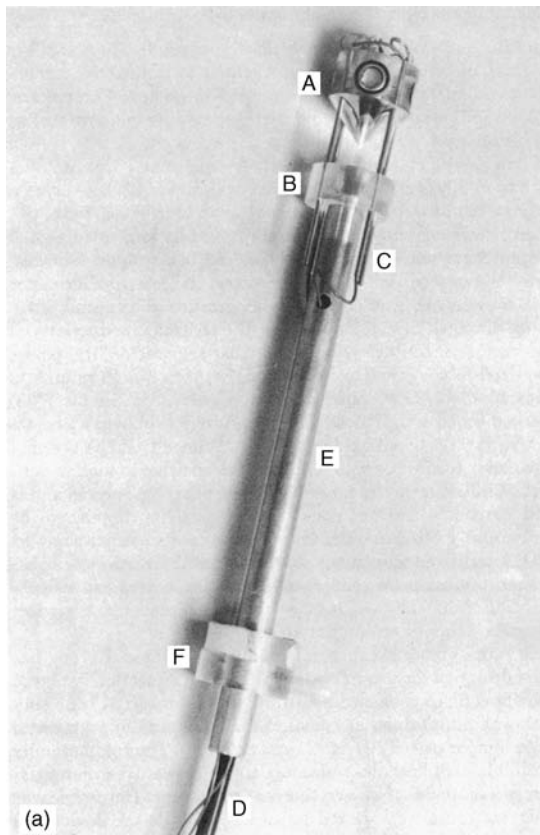


Figure 2. (a) Vaginal plethysmograph featuring a circularly symmetrical source/detector pattern. (b) Automatic baseline offset circuit. (From Ref. 4 © 1978, IEEE. Figures courtesy of the authors.)

amenorrhea had decreased vaginal blood flow as assessed via VPPG in contrast with “normals” under equivalently rated conditions of sexual arousal. Again, there was a negligible correlation between physiological and subjective ratings of arousal. Vaginal photoplethysmography VPPG has also been used to assess hematological function pre- and postoperatively for a variety of procedures.

Notably, both VBV and VPA are measured via polygraph. This results in the use of millimeters of deflection or change scores as the dependent measure in such measure-

ments. This often results in within-study consistency, with cross-study inconsistency. Furthermore, some researchers use deviation from baseline (average over the length of preintervention assessment), while others use deviations from time-specific baselines (e.g., average of 10 s prior to next condition, which can be erotic or nonerotic).

Other problems with the use of VPPG include the reliance of the measurement on blood oxygen saturation, temperature, and response time of the light indicator. Hoon et al. (5) also discuss the relative difference of pre- and postmenopausal status, as well as sleep cycle status, as it relates to basal blood flow. Each of these affects the ability of the VPPG to reliably detect a difference within and across subjects over time. Furthermore, genital engorgement during the time of ovulation has been reported as much greater than during the 10th day of the cycle (2), suggesting hormonal levels may influence baseline genital hematology through the menstrual cycle. Notably, variations in hormone levels through both menstrual cycles and entire lifespans are related to vaginal blood flow. Similarly, increases or decreases in naturally occurring agents, such as prolactin and androgens, can change the basal levels of vaginal blood flow and alter the rapidity and duration of genital engorgement. Finally, external agents, such as SSRIs and oral contraceptives can interfere with genital blood flow. However, arousal responses appear to be minimally reliant upon the woman's hormonal levels relative to other variables, such as type of stimulation (e.g., fantasy, film, audiotape) or activity (e.g., watching a film versus masturbating).

Beck et al. (8) highlighted some of the above-mentioned problems with VPPG, emphasizing the measurement concerns related to temperature, response consistency, and drift. They assert that, at the time of publication, no devices seemed sufficient in measuring female sexual arousal.

Several recent studies have explored the relationship among parasympathetic nervous system arousal, genital arousal, and subjective arousal. Early returns suggest parasympathetic arousal may facilitate sympathetic activation in female sexual arousal. Vaginal photoplethysmography has been used to distinguish between vaginal blood flow related to sexual arousal versus sexual anxiety. The ability to differentiate anxiety-induced blood flow versus arousal-based blood flow can be quite helpful, depending on the issue being assessed. Further, sexually dysfunctional women often report significantly less autonomic arousal. One possible extension of these data would be to include treadmills and exercise bikes as potential sexual response enhancement devices, as physical exercise often leads to parasympathetic arousal.

Perhaps surprisingly, the gross differences within and across vaginas has not been shown to introduce confounds in experimental research. Given the changing shape of the vagina during sexual response cycle, researchers have found a fair degree of reliability is possible, even when the VPPG shifts or is repositioned within the vagina. However, researchers and practitioners need to heed the inherent measurement artifact introduced due to movement within subjects and differences across subjects.

Another problem in comparing VPPG study is the use of arousing stimuli, and the induction of arousal in a

laboratory setting. Several researchers have commented on the challenge to ecological validity when having volunteer undergraduates insert an acrylic tampon in a university research laboratory to assess sexual response. One improvement on this variable has been the development of a portable data collection augment to the VPPG to enhance validity. This provides for a greater sense of ecological validity, in that the research volunteers are able to pursue their sexuality in more naturalistic settings, such as their bedroom or dormitory room.

Further confounding is the type of activity employed to induce arousal in these settings. Some researchers use visual material, such as video depictions of heterosexual and homosexual congress, including oral, anal, manual, and vaginal. There is empirical, marketing, and anecdotal evidence that women are less responsive to visual erotica than to other media. Some researchers have instead employed audiotaped erotica, while others direct the subjects to engage in fantasy.

Some have used the VPPG to assess responding to male-versus female-produced sexually explicit videos. In addition to these variations, other studies have included the request to digitally masturbate, use a vibrator on the vulva and clitoris (without penetration), and to attempt or achieve orgasm, or some combination of all of these. Thus, it is difficult to compare the findings across studies using the VPPG to assess arousal, as there has been so little uniformity in the experimental procedures.

Several researchers have noted that physiological arousal and subjective arousal are not synonymous. One of the few studies to demonstrate a link among physiological and subjective arousal involved the use of VPPG in postoperative male-to-female transsexuals (9). Notably, they found postoperative male-to-female transsexuals demonstrated male-typical category-specific sexual arousal patterns following sex reassignment surgery. Their study included both genital and physiological responding, and discussed some of the physiological differences among natal women and sex-reassigned women, as the later may also include penile erectile tissue.

While relatively few in comparison to male sex offenders, female sex offenders appear to be receiving more attention in judicial and forensic settings. Due to their small numbers, there are few studies assessing female sexual arousal patterns using the VPPG. However, there are case studies in this area, with results likening the utility of the VPPG with the penile plethysmograph (described below).

The VPPG has its strengths and weaknesses relative to other assessment devices. Compared to self-report measures, it is more invasive, cumbersome, and difficult to use. However, it typically provides reliable and valid data that is otherwise unknown. It appears to have fallen out of favor in large scale randomized clinical trials, though is still popular for use in smaller scale and individually based studies.

Vaginal Electromyography. Engman et al. (10) used pelvic floor muscle surface electromyography (EMG) to assess for partial vaginismus (involuntary contractions of the muscles of the outer third of the vagina such that



Figure 3. The EMG transducer with associated detector and recording instrumentation for monitoring the contractile activity of the pubococcygens muscle. (Courtesy of Farrall Instruments, Ins., Grand Island, NE.)

intromission is painful or impossible) and vulvar vestibulitis (a specific form of vulvar pain). They conclude that EMG is not a useful method to distinguish between asymptomatic women and women with partial vaginismus and vulvar vestibulitis. However, other researchers have found EMG to differentiate women with vaginismus from those without during basal and induced vaginismus conditions. Others have used EMG to demonstrate differences between ejaculatory and nonejaculatory orgasmic women. Electromyography has been used successfully in several studies researching etiology, treatment, and bio-feedback for urinary incontinence. While this approach appears promising, there does not appear to be a consensus on the utility of this approach relative to other approaches (e.g., Vulvalgesimeter, discussed below) (Fig. 3).

Thermistor Clip and Vaginal Temperature. In a series of studies, Henson et al. (11) tested for a relationship between labial temperature and subjective ratings of arousal. They began by developing a thermistor-clip to measure changes in labial temperature, and found that labial temperature rose in response to erotic stimuli, while other measures of body temperature were unaffected. In assessments throughout the sexual arousal cycle, VPPG data and labial temperature variation were found to show corresponding variations up to the point of orgasm. The temperature remained relatively constant during orgasm, but began to rapidly decrease during the resolution phase. A follow-up study by another group using a portable vaginal temperature gauge found variation in vaginal temperature over the course of the day. These researchers hypothesized that the variations may be concomitant with circadian temperature changes, with temperature decrease in arousal being perhaps related to vaginal wall edema or a change in the position of the uterus. While Henson and colleagues made strong arguments for the use of temperature as a meaningful measure, later researchers have not followed-up on their early works.

External Measurement Devices

Vulvalgesiometer. Pukall et al. (12) introduced a new instrument for the assessment of pain in vulvar vestibulitis syndrome (VVS), which is a common form of dyspareunia in premenopausal women. Given the lack of standardization in the cotton-swab test for diagnosing VVS, Pukall et al. (12) developed a simple mechanical device, the Vulvalgesiometer, to standardize genital pain assessment. The device consists of a set of pen-like devices, each holding springs with varying resistance. The bottom of the device holds a Q-tip-type cotton swab, which is pushed down on the area being tested for pain. The pain threshold is measured by how much pressure can be exerted (or by how far the spring is compressed). The researchers reported that subjects endorsed similar pain experiences with the Vulvalgesiometer in comparison to intercourse pain. Furthermore, the authors contend that practitioners will be able to assess change over time more reliably with the Vulvalgesiometer. While still relatively new to the field, this approach shows promise as it appears relatively easy to use for both practitioner and client, as well as both reliable and valid for the assessment of vulvar pain (Fig. 4).

Vaginal Fluid Production. Levin (13) reviews the early work in the assessment of human female genital function. In his review, he discusses the measurement of vaginal pH, pO_2 , blood flow, motility, fluid and its ionic concentrations, electrical activity, and amino acid concentrations in both arousal and basal states. Other researchers have evaluated the prevalence and elements of female ejaculate. The measurement of vaginal lubrication has been assessed through the weight gain of preweighed filter papers. Research in this area has demonstrated the correlation between vaginal lubrication and other physiological measures of arousal. Thus, while not a device *per se*, it is a measurement approach to an aspect of female sexuality. However, the lack of uniform methods of assessment and relative difficulty in obtaining the data appear to have impeded additional research in this area.

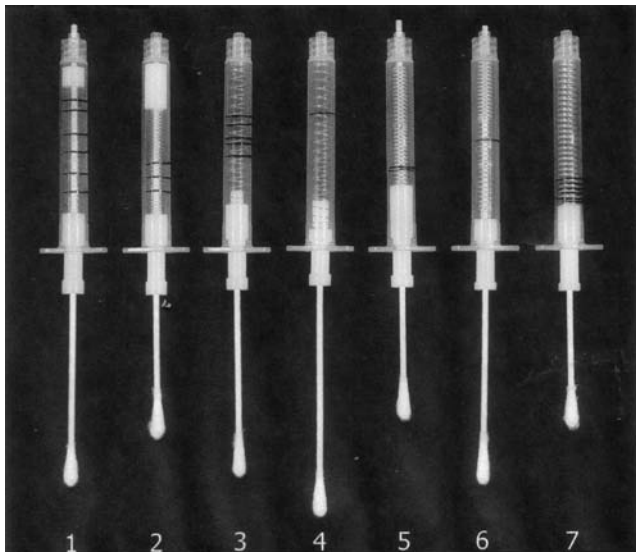


Figure 4. Figure of vulvalgesiometer.

Imaging

Duplex Doppler Ultrasonography. Munarriz et al. (14) reported on the use of duplex doppler ultrasonography (DDU) to assess genital engorgement. They used ultrasonography combined with the Doppler effect to assess clitoral and corpus spongiosum diameter, as well as clitoral and corpus spongiosum peak systolic and end-diastolic velocity. The Doppler effect can show whether blood is flowing toward or away from the device, while the ultrasound helps for visualization. They concluded the use of ultrasonography appears to be a less invasive and a consumer-friendly approach to assess genital engorgement in comparison with VPPG. Others have used DDU to assess changes in clitoral engorgement in a study of Alprostadil. Relative to other approaches, such as self-report, DDU is more time and cost-intensive and less invasive than VPPG.

Magnetic Resonance Imaging. Rosen (15) reports on the promising developments in the use of magnetic resonance imaging (MRI) to assess genital engorgement and response in human sexuality. However, there are few published studies in this promising area in the assessment of human female sexuality. Like the DDU, it is anticipated that with additional research, the MRI will be rated as more consumer-friendly than VPPG, and will likely be more cost-intensive than self-report measures of sexual arousal and genital engorgement.

FEMALE SEXUAL BEHAVIOR TREATMENT

Devices and Instruments

Vaginal Dilators. Vaginal dilators are generally a set of graduated glass tubes or cylinders, 2–5 cm in diameter and 15 cm in length. Typically, vaginal dilators are used for the treatment of vaginismus, which is the involuntary muscle contraction of the outer third of the vagina.

With the advent of plastics, nonbreakable dilators are preferable. Furthermore, while vaginal dilation had historically been conducted under general anesthesia, it is now being offered as an outpatient referral for self-administration at home. A recent study using vaginal dilators for in-home use lists their instructions as: These dilators are to relax the muscles around the entrance to the vagina and to gently stretch the area. You should start using the smallest of the dilators, which should be inserted for 10–15 min·day⁻¹ and preferably for two episodes per day of 10–15 min, passing the dilator downward and backward into your vagina. Light lubrication with K-Y jelly or similar is advised. After 1 week you should increase the size of the dilator to the next biggest and increase the size by one every week until you are using the largest of the dilators or you feel comfortable to resume intercourse, whichever happens first. The time that the vaginal dilator should spend in the vagina is the same on each occasion (i.e., 10–15 min).

They also note: “Sometimes it was necessary to help the patient insert the first dilator in the clinic explaining the need to relax her pubococcygeal muscles while advancing the tube during muscle relaxation” (16) (Fig. 5).



Figure 5. Vaginal dilator (17). (Courtesy of www.vagenemus.com.)

Several researchers have noted the benefit of pairing muscle relaxation exercises, and specifically pelvic floor muscle relaxation, as a beneficial adjunct to vaginal dilator therapy. Weiss (18) describes the use of dilators in the treatment of vaginismus. She presents a case study in which, due to economic circumstances, an assortment of vegetables was substituted successfully in the treatment protocol, with positive long-term results. Dilators appear to be one of the most commonly used treatment approaches for vaginismus. For more on the etiology, diagnosis, and interventions related to vaginismus, please refer to Koehler (19). Dilators have also been used in several cases of vaginal construction or reconstruction. However, there is some variability in practice using dilators in this way.

Clitoral Vacuum. Billups et al. (20) introduced the Eros clitoral vacuum for the treatment of female sexual dysfunction. The device is a small battery-powered device designed to enhance clitoral engorgement, increase blood flow to the clitoris, and ultimately to work toward increased sexual arousal and response. It looks a bit like a computer mouse and has a small suction cup to be placed over the clitoris. In follow-up studies, this device was shown to increase genital and clitoral engorgement. Participants self-reported improved libido, arousal, lubrication, orgasm, and satisfaction, as well as decreased pain. Gynecological exams revealed improved mucosal color and moisture, vaginal elasticity, and decreased ulceration and bleeding (Fig. 6, Ref. 21).

Clitstim and Vibrotactile Stimulation. Riley and Riley (2003) (23) present a small study on the use of a finger-cot

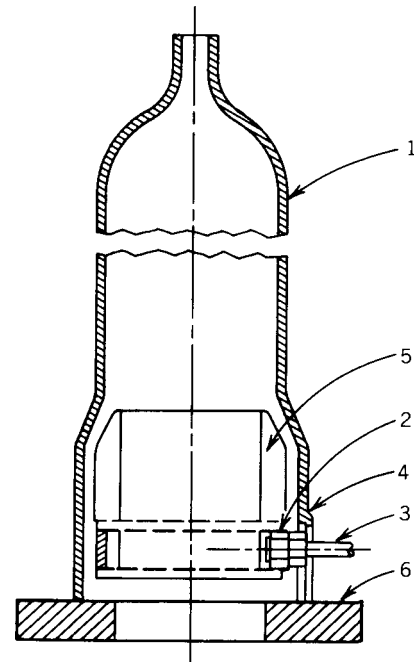


Figure 6. The Freund phalloglethysmograph: (1) glass cylinder, (2) plastic ring, (3) metal tube with threads, (4) lock nut, (5) rubber cuff, (6) flat, soft sponge rubber ring. (From Freund et al. (22). Copyright 1965 by the Society for the Experimental Analysis of Behavior, Inc. Reprinted by permission.)

shaped device (sometimes called a vibrotactile genital stimulator) designed to increase digital clitoral stimulation when worn on a finger. The device is similar in nature to a nonmedical clitoral vibrator, but its unobtrusive design was created with the hopes that it would be more socially acceptable. Based on their results, the researchers conclude that the device may be of assistance in anorgasmic or slow-to-orgasm women.

Over-the-counter vibrotactile devices (e.g., vibrators) are the most commonly used sexual instruments for anorgasmic and slow-to-orgasm women. Vibrotactile devices have also been developed for use with men, though the literature in this area is sparse.

IUD with Danazol. Cobellis et al. (24) report on the successful use of an IUD loaded with danazol for the treatment of dysmenorrhea, pelvic pain, and dyspareunia associated with endometriosis. Danazol had previously been used as an oral treatment for these ailments, however, this route of administration led to side effects. Cobellis et al. (24) found that the release of the danazol through the IUD eliminated these side effects (only adding a common IUD side-effect of first-month spotting). They reported 6 month duration of benefit with generally favorable consumer satisfaction.

Treatment Articles. There are several articles reviewing the treatment of sexual dysfunction in the human female without the use of specific instruments or devices. These include the use of botulin toxin for vaginismus, the role of fantasy training, androgen therapy, and cognitive-behavioral therapy including progressive muscle

Table 1. Female Sexual Behavior Assessment

Device	Function	References
VPPG	Measures changes in blood volume	2,5
EMG	Tests for vaginismus and vulvar vestibulitis	10
Thermistor clip	Measures labial temperature	11
Vulvalgesiometer	Measures pain associated with vulvar vestibulitis syndrome	12
DDU	Measures genital engorgement	14
MRI	Measures genital engorgement	15
Vaginal dilators	Treats vaginismus	16
Clitoral Vacuum (EROS therapy)	Increase genital engorgement	20
IUD with danazol	Treats dysmenorrhea	24

relaxation and meditation. Many of the more recent treatment outcome studies limit their dependent measures to self-report, often ignoring the physiological assessment that some researchers see as necessary and vital. Beck (25) reviewed the theories of etiology, prevalence estimates, and theories related to hypoactive sexual desire disorder. She concluded that a lack of valid, reliable, and consumer-satisfactory dependent measures impede scientific progress in this area.

Self-Report. For a review of self-report measures of female sexual function and behavior, please refer to Althof et al. (26) (Table 1).

INSTRUMENTS AND MEASUREMENT OF MALE HUMAN SEXUAL BEHAVIOR

Erectile Dysfunction and Sexual Deviance

The measurements of and devices related to male sexual behavior typically fall into one of two categories: erectile dysfunction and deviant sexual arousal. Coleman (27) reviews the scientific literature on the etiology and intervention for Erectile Dysfunction (ED), which is defined as the persistent inability to achieve or maintain an erection sufficient for satisfactory sexual activity. It is estimated that > 50% of men of 65 years-of-age experience ED. Meuleman and Van Lankveld (28) note that with the increasing availability of pharmacological interventions for ED, male hypoactive sexual desire disorder may be commonly misdiagnosed. They conclude, "HSDD is more common in men than in women. In public opinion and in medical practice, HSDD is often misinterpreted as ED, and treated as such. There is a need for physicians and patients to be educated, and for the development of reliable clinical tools to assess this aspect of male sexual function (294). The tools used for the assessment and treatment of ED are described below.

The second area of assessment and intervention is toward deviant sexual interest and arousal: most notably, pedophilia. For the assessment of sexual interest, researchers have often turned away from face-valid self-reports and

looked to physiological data. However, as with the women discussed above, it has been shown that physiological arousal is not uniformly correlated to emotional or cognitive arousal.

An area of concern in this topic is the limits to ecological validity inherent in university- or medical-center-based research on male sexuality. Just as the women above are presumed to rarely masturbate with photoplethysmographs inserted, it is presumed that most men do not wear penile Strain Gauges or plethysmographs when engaging in sexual activity (see Ref. 29 for more in this area). Again, more recent researchers have adopted intent-to-treat or wait-list controls to enhance the ecological validity of their research.

MALE ERECTILE DYSFUNCTION

Assessment

Circumferential versus Volumetric Assessment. The early research on male erectile response typically used volumetric assessment. Volumetric assessment usually involves the placement of a flaccid penis in an airtight cylinder with a monitored release valve. Engorgement leads to displacement of air out the valve, leading to an indication of volume displacement due to erection.

Circumferential assessment often involves the placement of a mercury-filled rubberized band around the shaft of the penis near the base. Engorgement typically leads to expanding girth, resulting in millimeter displacement as measured by an accompanying graphing device (Figs. 7 and 8).

In a comparison of volumetric and circumferential measures of penile erection, Wheeler and Rubin (30) found that the circumferential method was preferable for several reasons. While both measures showed a correlation, the authors contend the volumetric approach produced more artifacts, was more difficult to use, and was no more sensitive than the easier-to-use circumferential method. However, a rapid change in length may result in a net decrease in girth, at least temporarily. Taken together, information on the length and width of penile tumescence are consistent with recommendations to take measures over time to assess the level, course, and trend of the observed response.

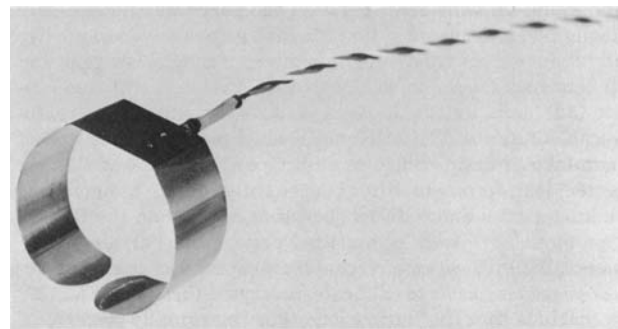


Figure 7. Barlow-type strain gage for monitoring penile tumescence. (Courtesy of Farrall Instruments, Inc., Grand Island, NE.)

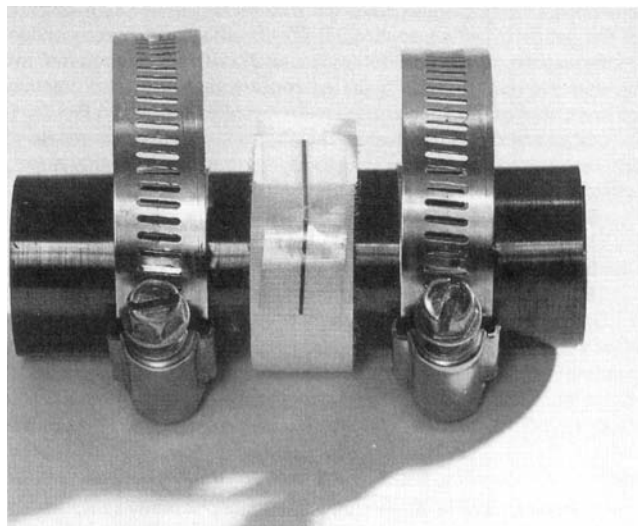


Figure 8. Measuring graphing device for expanding girth.

The assessment of ED often involves the use of a snap gauge, which is an application of multiple circumferential measures (there are several brands and types available, e.g., Dacomed's Snap Gauge, the RigiScan). In essence, this measures a sample of volumetric expansion, which allows for the assessment of both rigidity and circumference. Snap gauges often involve the placement of a band or sleeve around the shaft of the penis, with two or more expansion-specific snaps placed along the length of the gauge. Thus, as the penis expands from the base to tip, the snaps are presumed to snap in sequence dependent on the length of turgidity.

A common assessment avenue for ED is the use of a snap gauge during sleep. Typically, adult males enter REM stage three to four times per night, achieving at least one erection during each REM stage. The sleep-related erections are termed nocturnal penile tumescence (NPT). Presence of NPT is suggestive of psychogenic ED, while an absence of NPT is often inferred as organic ED.

Libman et al. (31) discussed the issues and problems with the reliance on NPT and snap gauges to assess for erectile dysfunction. These authors emphasize the growing body of research indicating waking erectile capacity is not directly related to NPT, concluding there may be different processes underlying NPT and erotically induced erection. Further, Libman noted several subjects balked when presented with the notion of "snap" and "penis" in the same sentence. They altered the description of the procedure to impotence testing and expansion tape. In their study, they confirmed that NPT, as assessed by a snap gauge, did not constitute a valid measure of daytime erectile ability in their sample. However, they provide recommendations to enhance the utility of the assessment, including assessment during waking and sleeping states, as well as a multimodal approach.

The Snap Gauge (Dacomed Inc.) is described by Diedrich et al. (32): A commercial device, which, according to the manufacturer, can be used for the determination of organic impotence. The Snap Gauge consists of nonstretch-

able fabric band that is fastened around midshaft of the penis by Velcro straps. Three plastic snap elements are attached to the device such that they break in sequence if an erection of sufficient circumferential expansion and local hardness is obtained. A male using the device may break none, one, two, or three of these elements corresponding to the sufficiency of his erection.

It is possible to judge a person's ability to reach sufficient tumescence for intercourse by looking at how many snaps remained intact. With all three intact, no intercourse would be possible. With all three unsnapped, the erection would be insufficient for intercourse. If two snaps remain, it is questionable whether the subject has reached sufficient rigidity (Fig. 9, Ref. 32).

Nobre et al. (33) found a similar discordance between physiological and subjective ratings of arousal. Generally, they found men in their sample to underreport their arousal in relation to physiological measures of tumescence. Notably, in their study, participants were unable to manually or visually assess their erections. The researchers noted that positive affect facilitated arousal and was predictive of subjective and physiological arousal estimate concordance.

Strain Gauge. The measurement of human penile tumescence was one of the first review articles describing the construction and measurement issues associated with the armamentarium for the assessment of male sexual arousal. Rosen and Keefe (34) recommend volumetric assessment when precision is important (e.g., specific research question) and circumferential assessment for efficiency and ease of use. They cite the mercury-in-rubber as an optimal tool for the measurement of circumference.

Strain Gauge measurement (a component in PPG) is often conducted via a method involving millimeter displacement as a percentile of maximum erection. In this procedure, the participant places the gauge on the flaccid penis. They are then directed to self-stimulate to a full erection. The point of maximum tumescence is then taken as 100% erection, with later erections being taken as a proportion. Notably, this procedure allows for erections of >100%. There is also an inherent floor effect, in that the lab setting allows for a measurement of flaccidity that does not measure any aversive or deroxing responding (33).

The Dacomed Snap Gauge was evaluated by Diedrich et al. (32) using both mechanical and human subject tumescence. In the biomechanical procedure, a device was manufactured that allowed for the incremental increase in circumference (1mm intervals). The Snap Gauge was placed on the mechanical device, as well as a mercury Strain Gauge. With the human participants, a mercury gauge was placed near the base of the penis, with the Snap Gauge at mid-shaft. Participants were able to manually and visually assess their erections, and were asked to report as the Snap Gauge snaps broke, as well as to report their sufficiency for penetration following the presentation of sexually explicit videotape. The authors noted that the Snap Gauge appears to be lacking in validity based on three problematic deficiencies: the lack of erection uniformity, lack of placement uniformity, and lack of uniform rigidity. The authors direct practitioners to proceed

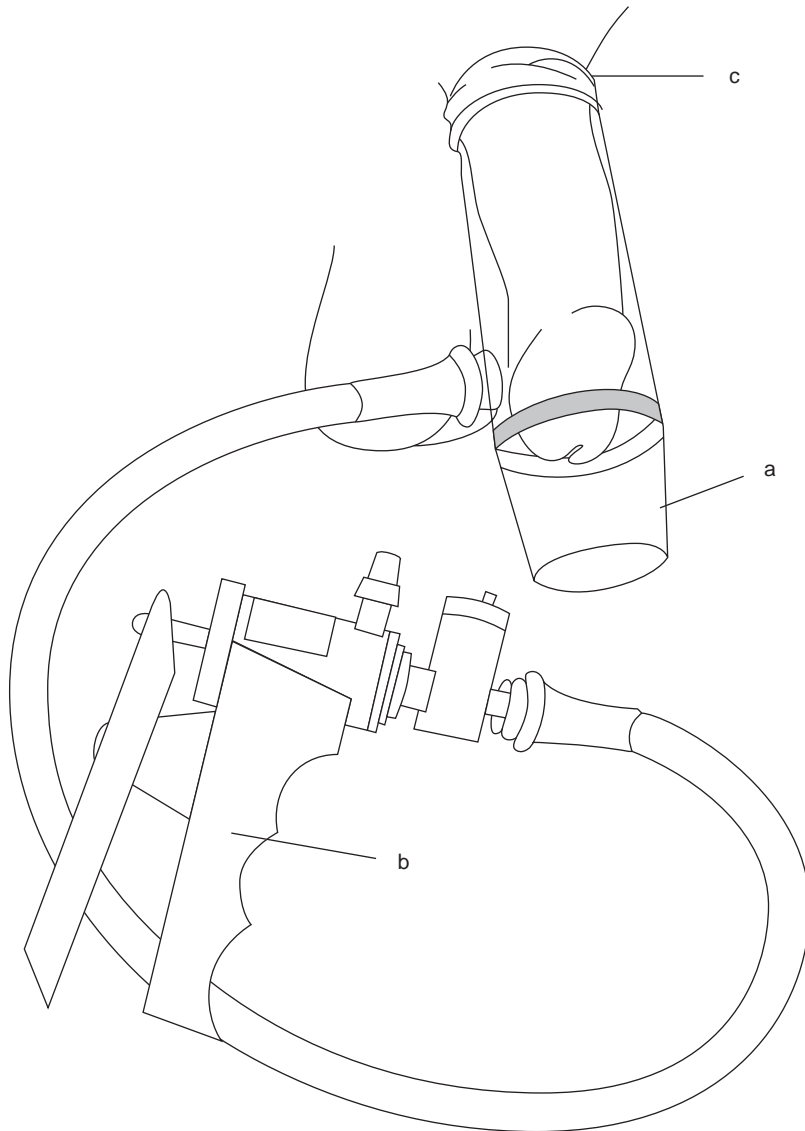


Figure 9. Snap Gauge (Dacomed Inc.).

with caution when diagnosing ED based upon Snap Gauge data, as in their analysis, the Snap Gauge was subject to both false positives and negatives.

In comparison studies of the RigiScan and Snap Gauge in the assessment of penile rigidity, researchers have concluded the Snap Gauge is more cost-effective than and as reliable as the more complicated RigiScan. They recommend the use of The Snap Gauge primarily, with the RigiScan being recommended for clients when The Snap Gauge is inconclusive or when assessment that is more detailed is necessary (e.g., per research protocol).

Stamp Test. One of the simplest, least expensive, and perhaps oldest tests for ED is known as the stamp test. The stamp test is used to assess NPT. In this procedure, the subject fastens postage stamps around the base of the penis, similarly to where the snaps would be found on a snap gauge. The NPT is likely to displace the stamps. The stamp test was developed before the widespread use of self-adhesive stamps and instead used the moisture-catalyzed

gumming adhesive. Moreover, while this is relatively inexpensive, the published reports provide two false positives and one false negative in a rather small sample.

Electromyography. Da Silva et al. (35) found that EMG could be useful in the differential diagnosis of ED. In this clinician-administered procedure, EMG inserts electrode needles into the muscle, and when the subject is asked to move that muscle, the EMG gives a general picture of the muscle activity by showing the action potentials occurring in the specific surrounding muscle cells. This makes it possible to see whether the dysfunction is indeed physiological. However, few researchers have completed follow-up studies on this promising work.

Imaging

Near-Infrared Spectroscopy. Burnett et al. (36) compared near-infrared spectroscopy (NIS) with color duplex ultrasonography, Strain Gauge circumference measure,

penile tonometry, and clinical assessments. They found NIS to be a safe, inexpensive, and easy-to-use device that provides for quantitative measurements of vascular physiology in erectile assessment, by measuring the percentage of blood volume reaching the penis, and monitoring the circulation. Again, few researchers have continued in this line of research.

Self-Report. A review of self-report measures for the assessment of male sexual behavior can be found at Berman et al. (37) and Kalmus and Beech (38).

THE TREATMENT OF ED

Devices and Instruments

With the advent and marketing of several pharmacological agents for the treatment of ED, some authors have questioned the utility of mechanical devices and surgical interventions. There have been several cases of adverse responses from pharmacological interventions, and there are several conditions for which these medications are contraindicated. For these individuals, surgical or mechanical intervention may be their only hope for a reliable and useful erection.

Vacuum Constrictive Device. The vacuum constrictive device (VCD) typically includes an acrylic cylinder into which the flaccid penis is inserted. The VCD has a manually squeezed bulb pump for the forced suction of air out of the cylinder (up to 250 mmHg, 33.33 kPa), which in turn leads to increased engorgement. A constrictive band is then placed at the base of the penis, with the intent of maintaining the engorgement through the completion of the sexual act by reducing, not preventing, penile venous outflow. Some pumps allow for the placement of the constrictive band (similar to what is sometimes called a “cockring”) while the pump is in place; others involve the removal of the pump and the placement of the band prior to detumescence. A device similar to the VCD (albeit satirized) played a supporting role in the first of the Austin Powers films. Wylie and Steward describe a homemade device for ED, for use when pharmacotherapy is contraindicated (39). In their case study, a 65 year old client fashioned his own VCD for the effective creation of erections (Fig. 10, Ref. 40).

Several researchers have commented on the utility of the VCD, with successes reported with diabetic and neuropathic populations, as well as with psychogenic ED.

Researchers have also found that the device is generally more effective when it is combined with couples’ therapy. However, several complications have also been discussed. Some men have complained of feelings of demoralization, discouragement, and confusion with the device, while others reported frustration when their spouses were not supportive or helpful when using VCD. Furthermore, some men report that even though they can create an erection, they are not necessarily emotionally aroused. Lastly, as the VCD involves the restriction of blood flow out of the penis, possible problems include hemotoma, skin irritation, and in at least one reported case, penile gangrene. This com-

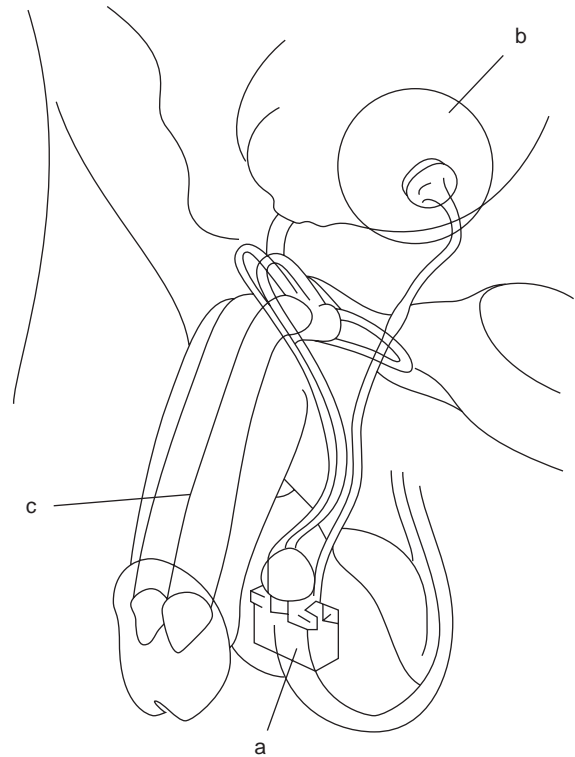


Figure 10. The vacuum constrictive device (VCD) sucks air out of a cylinder to increase penile engorgement.

plication underscores the need for proper instruction and compliance with the device.

Penile Prosthetics. Penile implants were some of the earliest treatments for ED. While practitioners now recommend nonsurgical interventions (e.g., pharmacological or the VCD treatment), there are still some clients for whom a prosthetic implant is preferable. Penile prosthetics have also been used successfully in the creation of neophallus in sex reassignment surgery, and in reconstructive surgery following injury.

Penile prosthetics are typically either malleable or inflatable. Malleable implants are basically permanent, semirigid tubes placed within the penis. When an erection is desired, the man (or partner) positions the penis into the desired direction. Inflatable prosthetics often involve two or more components. A fluid-filled pump reservoir is implanted in the scrotal sac, with hollow tubes (initially one, now two or more in connection) placed within the penile shaft. When an erection is desired, the man or partner squeezes the pump, leading to an outflow of fluid from the reservoir into the tubes, resulting in an erection. Surgically, the procedure can involve the removal of cavernous tissue. With the placement of prosthetic, it may result in fibrous involvement, scarring, or both. These complications can result in difficulty changing implants if needed, and preventing future “natural” erections. A fenestrated implant has been successful in the sparing of cavernous tissue and spontaneous erections.

Follow-up studies on penile prosthetics show high rates of consumer satisfaction, with occasional problems with

Table 2. Erectile Dysfunction

Device	Function	Reference
Snap Gauge	Measures penile rigidity and circumference	32
Strain Gauge	Measures penile circumference via displacement	34
EMG	Tests for physiological penile dysfunction	35
NIS	Measures penile blood volume	36
VCD	Treats ED by increasing engorgement	39
Penile Prosthetics	Surgical treatment for ED	

leakage, breakage, infection, and stretching due to over-use. A common partner complaint is the lack of girth; other complaints have included spontaneous deflation, a cold penis, and that intercourse felt unnatural. Malleable implants are reputed to be easier to install, with inflatable prosthetics typically rated higher by consumers and their partners (Table 2).

Summary. Several researchers have explored the effectiveness, advantages and disadvantages of pharmacological, psychotherapeutic, surgical, and mechanical treatment of ED, and in combinations of these. Taken together, there is no clear gold standard for the treatment of ED; rather, there are several tools available to the clinician and consumer that may be appropriate given the client's physiology, psychology, and context.

DEVIANT PHYSIOLOGICAL AROUSAL

Assessment

The second major area of inquiry into male sexual behavior relates to physiological changes related to deviant sexual arousal. The following section describes the use of the penile plethysmograph (PPG), visual tracking (VT), and pupillometry (PM) to assess differential arousal in response to presentation of varied stimuli. The bulk of this work centers on deviant sexual arousal, typically with adult males being shown or presented videotapes, still images, or audio recordings. The stimuli typically depict age-appropriate as well as age-inappropriate stimuli, and may include other challenges, such as sadistic themes, adult rape themes, fetishes, bondage or sadomasochistic themes, or both, and control conditions with no intended sexual content.

While in Freund's early phallometry work (41) PPG typically assessed volume, the bulk of the last half-century's research using PPG seems to be reliant upon circumferential assessment. Bancroft et al. (42) describe the construction, provide schematics, and a photograph of a simple transducer for measuring penile tumescence. The device is a mercury and rubber strain gauge, which they describe is inexpensive, portable, easy to apply, and unobtrusive in use. They contrast the simple strain gauge with the more complex volumetric PPG used by Freund

(described earlier, please refer to earlier figures of penile plethysmography devices).

Penile Plethysmograph. A PPG is a strain gauge (a stretchable band filled with mercury that is fitted around a subject's penis, discussed earlier in this article), connected to a video screen and data recorder, which records changes in the penis with different stimulus. While PPG can assess both volumetric and circumferential changes, PPG using measures of circumference changes appears to be more widely used. Advances in technology have led to the advent of computer-based assessment and scoring in penile plethysmography. O'Donohue and Letourneau (43) concluded, "there does not appear to be a standardized penile tumescence assessment, but rather there is a family of procedures that share some common aims and features (p. 126)." The 17 potential sources of variation in the assessment that O'Donohue and Letourneau described have not been satisfactorily addressed in most recent reviews. Still, some researchers use an average of pen deflection; some use a sampled interval average, while some use maximum response and percentage of maximum response. Further, as measurement is often less reliable at the ends of the metric, there was some discussion about the reliability and validity of PPG at large expansions.

Kalmus and Beech (38) reviewed the self-report and physiological assessment of male sexual response. They noted that PPG is still the most common method for such assessment, though to acknowledge it is prone to a number of possible artifacts, including faking or intentional suppression. Other clinicians have noted a possible retest habituation effect, which may be an artifact of attempts at standardization. However, some authors contended the PPG is useful in repeated assessments, provided the stimuli are varied and content matched to reduce habituation.

Golde et al. (44) assessed participants' ability to suppress penile responding under audio-only and audiovisual stimulus conditions. They also included measures common to the polygraph (discussed in a later section polygraph), galvanic skin response (GSR), and finger pulse amplitude (FPA). The authors found participants were able to suppress penile arousal while not providing indicators of deception as measured through GSR and FPA. Both naïve and experienced participants were able to suppress, and arousal was less pronounced in audio-only conditions. These data suggest the PPG is susceptible to faking through suppression, especially in the audio-only presentation.

Other evidence of altered penile tumescence comes in treatment studies using the PPG as a dependent measure. Rosen and Kopel (45) demonstrated that with multiple sessions of biofeedback, penile tumescence to undesired sexual stimuli was reduced with lasting effects. The authors contended the result was not due to habituation or generalized suppression. Follow-up data indicated the case study subject had been deceptive in his self-report and had renewed the targeted sexual behavior. This research vignette highlights the oft-noted weakness of deception when using self-report data.

In a study comparing adult males alleged to have sexually offended against children and "normal controls",

Haywood et al. (46) assessed the relationship between physiological arousal as measured by PPG and subjective arousal assessed through self-report. The alleged sex offenders showed more subjective arousal to children, with nonincest offenders reporting higher subjective arousal than incest offenders. Notably, both alleged offenders and normals reported subjective arousal without physiological responding, and subjective repulsion, despite showing penile responding. The authors recommend for the continued use of subjective and physiological measures in the assessment of sexual response.

Other concerns with the PPG include the ethical use of such a device in correctional or forensic settings, including the use of PPG as a treatment device for the challenging of sexual offenders reported responses and arousal patterns. There are also cautions against the use of PPG for predictive purposes, or for reasons beyond which it has been demonstrated to be reliable and valid.

Recent work has explored the use of portable circumferential PPG. Rea et al. (47) found the data taken inside the lab to generally correspond with data taken outside the lab. Not surprisingly, there appeared to be an inverse relationship between penile tumescence and proximity of researcher.

Finally, Mathews et al. (48) reviewed the arguments and problems with the use of PPG in legal settings. They discussed the Frye and Daubert standards of evidence as it relates to the PPG. These standards describe requirements of a measure before it is considered acceptable in evidentiary purposes. A description of these standards can be found in the O'Donohue and Levensky edited collection, *The Handbook of Forensic Psychology* (49). Given the above findings indicating the PPG can be suppressed, and be subject to false positives and negatives, Mathews et al. (48) contended the PPG should not be admitted as evidence for forensic purposes.

Pupillometry. Pupil dilation is indicative of arousal, including sexual arousal. Pupillometry uses a magnified video recorder to track a subject's pupil changes when presented to different stimuli. While the results using pupillometry originally seemed promising, the research has been criticized for a variety of methodological problems, and appears to have lost favor as an avenue of inquiry.

Viewing Time. In viewing time studies, participants are presented visual stimulus materials and allowed to view the materials. The instructions vary across studies as to how long and how much liberty the participant has to view the materials (e.g., in some, the participants are allowed to choose display times, while in others, visual field is tracked). The principle behind this approach is that participants will view images that hold more sexual interest to them for longer periods than those without such appeal.

Abel et al. (50) reported on the development of the "Abel Screen", also known as the Abel Assessment for Sexual Interest (AASI). This device is reported to correspond well with the more intrusive volumetric assessments, and to be more efficient. Laws and Gress (51) review the multiple criticisms of this specific instrument, although they also

acknowledge the benefit of VT-based assessments. The Affinity [Glasgow et al. (52); Kalmus and Beech (38)] is a VT device designed for use with learning-disabled offenders. While there are minimal data to support its use, the existing data are promising.

The Abel Screen [cf. Gaither (53)] was created in response to some of the criticisms of plethysmography. The Abel Screen is different in that it measures attention (as opposed to tumescence or vasocongestion) as measured by tracking visual focus on stimulus objects. The Abel Screen was developed in part to eliminate the need for nude slides as stimulus materials. By using the Abel Screen, clinicians are able to assess the focus of the respondent's attention, and the duration of that attention. Further, the Abel Screen is less time intensive, and less intrusive than plethysmography. While Abel and colleagues report that visual reaction time is a reliable (alphas 0.84–0.90 across stimuli) and valid means of assessing interest, Gaither's data was mixed in an analogue study using undergraduate students. Other researchers have criticized VT studies for their clear face validity, which makes the VT vulnerable to faking. The device consists of a questionnaire, which tests for things such as deviant and inappropriate sexual behaviors, and a program on a computer, in which the participant is to rate the slides on a seven-point Likert scale from "highly sexually disgusting" to "highly sexually arousing". The computer tracks how long it takes for the subject to rate the slide and advance to the next slide, thus tracking how long the subject looks at the slide. Abel and colleagues also state that measures have been taken to identify offenders attempting to conceal their offenses (abelscreen.com) (54), and even with deniers Abel and colleagues contend device is able to detect 88% of sexual abusers.

Laws and Gress (51) reported on the development of computer-generated images in a standardized assessment using PPG, VT, and Choice Reaction Time. They posited this approach as a stopgap measure in the advancement of the inclusion of new technologies to assist in the assessment and monitoring of deviant sexual arousal. They argued that developments in virtual reality might provide avenues of worthy inquiry by posing more realistic challenges to relapse prevention plans while addressing community safety and privacy. With increases in technology and privacy regulations, many researchers are turning to computer-generated composite images to further enhance and clarify sexual arousal patterns. In doing so, researchers may be able to provide for individually based stimulus materials that would maximize the likelihood of deviant arousal and more realistic virtual reality scenarios to provide therapeutic challenges for sex offender treatment skills application.

In a comparison of Abel's visual reaction time (VRT) and PPG with audiostimuli, Letourneau (55) revealed interesting data: both identified offenders against young boys, and neither reliably identified offenders against adult women. Visual reaction time identified offenders against young girls (although not reliably), and surprisingly, PPG data indicated men with female child victims produced significantly lower levels of arousal than men in other offence categories.

Thermistor, Temperature, and Photoelectric Surface Blood Volume. Temperature measurements, indicative of penile and groin blood flow, have not been supported as valid and reliable measures of arousal due to the latency involved in tumescence and detumescence (38). Previous reviews discussed the possibility of telemetric temperature dissipation assessment as an avenue of physiological arousal assessment, however, there are few publications promoting its use in this way. While photoelectric blood volume research is the basis for VPPG in women, it has not born fruit in assessment of male sexual response (38).

Electroencephalogram Movement. Cohen (56) presented one of the few research projects in which the sex organ between the ears was assessed. Their preliminary results suggested an electroencephalogram (EEG) can be used to assess sexual arousal, and that responding was different based on stimulus modality. Although limited, a promising area of inquiry is the EEG measurement of contingent negative variation. Researchers in this area noted a presumed difficulty in faking EEG response. However, given the limited data in this area, caution is warranted in using this approach in the assessment of sexual preference (38).

Galvanic Skin Response. The relatively few studies in this area suggest galvanic skin response (GSR) may be a useful augment to PPG in assessment in which the respondent is showing minimal response or is hypothesized to be faking or suppressing a response. Typically, research in this area is based on the concurrent use of PPG and polygraph.

Polygraphy. Oksol and O'Donohue (57) provided a thorough review of the use of the polygraph in a forensic setting. In their review, they described polygraphy as the evaluation of physiologic reactions that purportedly occur in response to the emotions of fear or conflict, or are in some other way associated with lying. The polygraph measures a number of subtle and involuntary changes in physiological functions, such as heart rate, skin resistance, and blood pressure. By amplifying and recording autonomic functions on a multichannel instrument, the polygraph detects the changes in these functions. It is so named because it has many (poly) pens, with each pen measuring and recording (graphing) a different physiological response. Some of these physiologic responses are recorded on a polygraph chart and the polygrapher interprets these changes to be indicative of truthfulness or deception.

Over the years, there have been many settings in which polygraph examinations have played an important role. These settings include criminal investigations of suspects accused of theft, rape, murder, or lesser crimes. Recently, a number of criminal justice and treatment professionals have been advocates of the increased use of polygraph tests to assess child sexual abuse allegations and have implemented polygraph testing in their practices or treatment programs. For example, polygraphers have been called to assess the verity of a subject's self-reported masturbation patterns, fantasy content, and truthfulness in treatment.

Contrary to lay opinion, there is no such thing as a "typical polygraph test". In reality, the polygraph exam-

ination varies immeasurably from operator to operator. From formulating the questions that will be asked of the suspect to scoring the results of the physiological responses, polygraph techniques are not standardized. Rather, each polygraph administration is a complex and highly variable conglomerate of choice points and interview situations. Polygraphers choose from at least eight different question formats of lie detection, with each having its own psychometric properties (e.g., accuracy rates, reliability).

Oksol and O'Donohue (57) highlighted the multitude of problems facing the forensic use of the polygraph. Most problematic are the lack of uniformity in the procedures, the lack of reliability, and the lack of validity, in that the polygraph has not been able to reliably demonstrate that it assesses what its proponents argue the polygraph assess.

TREATMENT OF SEXUAL DEVIANCE

There are not any accepted devices or instruments for the treatment of sexual deviance, other than the assessment devices described above. For a review of the issues, controversies, and methodology in the treatment of sexual deviance, please refer to Sbraga (58) (Table 3).

Summary. There are several devices and instruments available to the clinician assessing and treating human sexual function. A common theme throughout the measures of human sexuality is the lack of concurrence among physiological and self-report data. It seems as though self-report measures are more satisfactory to the consumer, and perhaps have better psychometric properties. As a result, more and more outcome studies are relying on self-report data for their ease of use and psychometric superiority to physiological devices and instruments.

While there are several reliable, valid, and useful devices and instruments for the assessment and treatment of human sexuality issues, there are some shortcomings that limit the use of these tools. A common theme in review articles is a lack of uniformity in the psychophysiological assessment of human sexual behavior. This lack of uniformity is likely one component in the relatively lower reliability in comparison with self-report data. In addition, many of the research studies in this area have used

Table 3. Deviant Sexual Arousal

Device	Function	Reference
PPG	Measures circumference changes with different stimuli	43
Pupillometry	Tracks pupil dilation with different stimuli	
Abel Screen	Tracks visual tracking with different stimuli	59
EEG Movement	Assesses sexual arousal stimuli	56
GSR	Measures skin response with different stimuli	
Polygraphy	Measures subtle physiological changes associated with lying	57

volunteers, which is problematic due to the demonstrated differences among human sexuality research volunteers and normals. However, some gains have been made in increasing the ecological validity of human sexuality research. Taken together, the above data are consistent with the recommendation for a multimodal assessment, addressing the physiological and psychological (cognitive and emotional) components of human sexuality. Consumer satisfaction with medical devices and instruments appears to be driven by ease of use, partner support, and effectiveness. Technological advances may increase the precision with which the assessment and treatment of the above concerns can be made; however, a lack of cohesiveness in the field may serve to perpetuate the aforementioned problems.

BIBLIOGRAPHY

- Semmlow JL. Sexual instrumentation. In: Webster JG, editor. *Encyclopedia of Medical Devices and Instrumentation*. New York: John Wiley & Sons, Inc.; 1988.
- Hatch JP. Vaginal photoplethysmography: Methodological considerations. *Arch Sexual Behav* 1979;8(4):357–374.
- Sintchak G, Geer JH. A vaginal plethysmography system. *Psychophysiology* 1975;12:113.
- Ormon H, Weinnian J, Weinstein D. A vaginal photoplethysmography transducer. *IEEE Trans Biomed Eng BME* 1978; 25:434.
- Hoon PW, William D, Laughter JS. Infrared vaginal photoplethysmography: Construction, calibration, and sources of artifact. *Behav Assess* 1984;6(2):141–152.
- Geer JH, Morokoff P, Greenwood P. Sexual arousal in women: The development of a measurement device for vaginal blood volume. *Arch Sexual Behav* 1974;3(6):559–564.
- Van Dam FS, Honnebier WJ, van Zalinge EA, Barendregt JT. Sexual arousal measured by photoplethysmography. *Behav Eng* 1976;3(4):97–101.
- Beck JG, Sakheim DK, Barlow DH. Operating characteristics of the vaginal photoplethysmograph: Some implications for its use. *Arch Sexual Behav* 1983;12(1):43–58.
- Lawrence A, Latty E, Chivers M, Bailey J. Measurement of sexual arousal in postoperative male-to-female transsexuals using vaginal photoplethysmography. *Arch Sexual Behav* 2005;34(2):135–145.
- Engman M, Lindehammer H, Wijma B. Surface electromyography diagnostics in women with partial vaginismus with or without vulvar vestibulitis and in asymptomatic women. *J Psychosomat Obs Gyn* 2004;25(3–4):281–294.
- Henson DE, Rubin HB, Henson C. Labial and vaginal blood volume responses to visual and tactile stimuli. *Arch Sexual Behav* 1982;11(1):23–31.
- Pukall C, Binik YM, Khalif S. A new instrument for pain assessment in vulvar vestibulitis syndrome. *J Sex Marital Ther* 2004;30(2):69–78.
- Levin RJ. A journey through two lumens. *Inter J Impot Res* 2003;15(1):2–9.
- Munarriz R, et al. A prospective duplex doppler ultrasonographic study in women with sexual arousal disorder to objectively assess genital engorgement induced by EROS Therapy. *J Sex Marital Ther* 2003;29(Special Issue): The Annual Meeting of the Female Sexual Function Forum: 85–94.
- Rosen RC. Assessment of female sexual dysfunction: Review of validated methods. *Fertility Sterility* 2002;77(4):89–93.
- Idama TO, Pring DW. Vaginal dilator therapy—an outpatient gynaecological option in the management of dyspareunia. *J Obst Gyn* 2000;20(3):303–305.
- Vaginismus.com (No date). Dilator set. [Online]. Available at <http://www.vaginismus.com/products/dilatorset/>. Accessed 2005, August.
- Weiss JC. Treating Vaginismus: Patient without partner. *J Sex Educ Ther* 2001;26(1):28–33.
- Koehler JD. Vaginismus: Diagnosis, etiology and intervention. *Cont Sexuality* 2002;36(9):9–17.
- Billups KL, et al. A new non-pharmacological vacuum therapy for female sexual dysfunction. *J Sex Marital Ther* 2001;27(5).
- UroMetrics (No date). Photo Downloads. [Online]. Available at <http://urometrics.com/pressroom/photodownloads.cfm>. Accessed 2005, August.
- Freund K, Sedlacek F, Kanob K. A simple transducer for mechanical plethysmography of the male genital. *J Exp Anal Behav* 1965;8:169.
- Riley A, Riley E. The effect of Clitstim (Vielle™) on sexual response induced by masturbation in female volunteers. *Sexual Relationship Ther* 2003;18(2):45–52.
- Cobellis L, et al. A donazol-loaded intrauterine device decreases dysmenorrhea, pelvic pain, and dyspareunia associated with endometriosis. *Fertility Sterility* 2004;82(1):239–240.
- Beck JG, Sakheim DK, Barlow DH. Operating characteristics of the vaginal photoplethysmograph: Some implications for its use. *Arch Sexual Behav* 1983;12(1):43–58.
- Althof SE, et al. Outcome measurement in female sexual dysfunction clinical trials: Review and recommendations. *J Sex Marital Ther* 2005;31(2):153–166.
- Coleman E. Erectile dysfunction: A review of current medical treatments. *Can J Human Sexuality* 1998;7(3):231–244.
- Meuleman EJH, van Lankveld JJDM. Hypoactive sexual desire disorder: An underestimated condition in men. *BJU Inter* 2005;95(3):191–196.
- Rowland DL. Issues in the laboratory study of human sexual response: A synthesis for the nontechnical sexologist. *J Sex Res* 1999;36(1):3–15.
- Wheeler D, Rubin HB. A comparison of volumetric and circumferential measures of penile erection. *Arch Sexual Behav* 1987;16(4):289–299.
- Libman E, et al. Sleeping and waking-state measurement of erectile function in an aging male population. *Psycholog Assess* 1989;1(4):284–291.
- Diedrich GK, Stock W, LoPiccolo J. A study on the mechanical reliability of the Dacomed Snap Gauge: Implications for the differentiation between organic psychogenic impotence. *Arch Sexual Behav* 1992;21(6):509–523.
- Nobre PJ, et al. Determinants of sexual arousal and accuracy of its self-estimation in sexually functional males. *J Sex Res* 2004;41(4):363–371.
- Rosen RC, Keefe FJ. The measurement of human penile tumescence. *Psychophysiology* 1978;15(4):366–376.
- Da Silva JP, Santiago L, Goncalves JC. The corpus cavernous electromyography in the erectile dysfunction diagnosis. *Acta Chirur Hung* 1994;34(3–4):243–252.
- Burnett AL, et al. Near infrared spectrophotometry for the diagnosis of vasculogenic erectile dysfunction. *Inter J Impot Res* 2000;12(5):247–254.
- Berman L, Berman J, Zierak M, Marley C. Outcome measurement in sexual disorders. In: IsHak WW, Burt T, editors. *Outcome Measurement in Psychiatry: A Critical Review*. Washington (DC): American Psychiatric Publishing; 2002.

38. Kalmus E, Beech AR. Forensic assessment of sexual interest: A review. *Aggression Violent Behav* 2005;10(2):193–217.
39. Wylie KR, Steward D, A Homemade Device for Erectile Dysfunction. *J Sex Marital Ther* 2000;26(4):335–339.
40. Brosman SA. (Dec. 14, 2004). Erectile Dysfunction Resource Center. [Online]. eMedicine. Available at <http://www.emedicine.com/rc/rc/pimages/i8/s11/ed.htm>. Accessed August 2005.
41. Freund K, Charles U. A laboratory method for diagnosing predominance of homo- or hetero- erotic interest in the male. *Behav Res Ther* 1963;1(1):85–93.
42. Bancroft JJJ, Jones HG, Pullan BR. A simple transducer for measuring penile erection, with comments on its use in the treatment of sexual disorders. *Behav Res Ther* 1966;4: 239–241.
43. O'Donohue W, Letourneau EY. The psychometric properties of the penile tumescence assessment of child molesters. *J Psychopathol Behav Assess* 1992;15(3):259–274.
44. Golde JA, Strassberg DS, Turner CM. Psychophysiological assessment of erectile response and its suppression as a function of stimulus media and previous experience with plethysmography. *J Sex Res* 2000;37(1):53–59.
45. Rosen RC, Kopel SA. Penile plethysmography and biofeedback in the treatment of a transvestite-exhibitionist. *J Consulting Clin Psychol* 1977;45(5):908–916.
46. Haywood TW, Grossman LS, Cavanaugh JL. Subjective versus objective measurements of deviant sexual arousal in clinical evaluations of alleged molesters. *Psycholog Assess* 1990;2(3):269–275.
47. Rea JA, DeBriere T, Butler K, Saunders KJ. An analysis of four sexual offenders' arousal in the natural environment through the use of a portable penile plethysmograph. *Sexual Abuse: J Res Treatment* 1998;10(3):239–255.
48. Mathews C, Hartsell JE, Kohn M. Debunking penile plethysmograph evidence. *Reporter* 2001;28(2):11–14.
49. O'Donohue W, Levensky E. *Handbook of Forensic Psychology*. New York: Academic Press; 2004.
50. Abel GG, et al. Screening tests for pedophilia. *Criminal Justice Behav* 1994;21(1): Special issue: The assessment and treatment of sex offenders 115–131.
51. Laws RD, Gress CLZ. Seeing things differently: The viewing time alternative to penile plethysmography. *Legal Criminologist Psychol* 2004;9(2):183–196.
52. Glasgow DV, Osborne A, Croxen J. An assessment tool for investigating paedophile sexual interest using viewing time: An application of single case methodology. *Br J Learning Disab* 2003;31(2):96–102.
53. Gaither GA. The reliability and validity of three new measures of male sexual preferences. *Dissertation Abs Inter: Sec B: Sci Eng* 2001;61(9-B):4981.
54. Web brochure. (No date) Abel Assessment for Sexual Interest Brochure. [Online]. Abel Screening, Inc. Available at <http://www.abelscreen.com/asipdfs/webbrochure.pdf>. Accessed August 2005.
55. Letourneau EJ. A comparison of objective measures of sexual arousal and interest: Visual reaction time and penile plethysmography. *Sexual Abuse: J Res Treat* 2002;14(3): 207–223.
56. Cohen AS, Rosen RC, Goldstein L. EEG hemispheric asymmetry during sexual arousal: Psychophysiological patterns in responsive, unresponsive, and dysfunctional men. *J Abnormal Psychol* 1985;94(4):580–590.
57. Oksol EM, O'Donohue W. A critical analysis of the polygraph. In: O'Donohue W, Levensky ER, editors. *Handbook of Forensic Psychology* New York: Elsevier; 2004.
58. Sbraga TP. Sexual deviance and forensic psychology: A Primer. In: O'Donohue W, Levensky ER, editors. *Handbook of Forensic Psychology* New York: Elsevier; 2004.
59. Abel, Saether 2001.

See also ANALYTICAL METHODS, AUTOMATED; PERIPHERAL VASCULAR NON-INVASIVE MEASUREMENTS; STRAIN GAGE; TEMPERATURE MONITORING.

SHAPE MEMORY ALLOYS. See ALLOYS, SHAPE MEMORY.

SHOCK, TREATMENT OF

SABIN DECAU
COLIN F. MACKENZIE
University of Maryland,
School of Medicine
Baltimore, Maryland

INTRODUCTION

Assessment and treatment of shock is based on understanding of circulatory system physiology and cellular metabolism. Shock is defined as inadequate supply of oxygen and nutrients due to inadequate capillary perfusion to the cells. This definition is not always correct, as in severe shock some cells cannot metabolize oxygen even with adequate perfusion. In addition, the removal of metabolites is equally important or even more crucial over time, since accumulated metabolites will cause cell injury.

Deficient capillary perfusion triggers a host of metabolic changes in every organ and tissue, which affects whole body homeostasis and circulation. Ideally, one should evaluate the cellular metabolism, but this assessment can be done only indirectly by measuring the acid–base balance in the form of blood gas and pH. This arterial or venous blood test does not give any information about the regional metabolic alterations. For evaluation of the circulatory function, indirect measurements are used including heart rate, blood pressure, and urinary output. Invasive hemodynamic monitoring devices, such as the pulmonary artery catheter are also used for cardiac function assessment. To effectively reverse shock requires monitoring of the pathophysiological state of hypoperfusion. Various devices are used for this state to be monitored directly or indirectly, continuously and intermittently, so that the patient response to therapy can be determined and treatment adjustments made.

ETIOLOGY

There are three common types of shock: loss of blood volume (hypovolemic shock), abnormal blood distribution (e.g., septic shock), and cardiac pump failure (e.g., cardiogenic shock). In practice, these three types of shock do not occur independently, but are frequently mixed with a similar final pathway, irrespective of the circulatory or cardiac state that may have been the precipitant (1).

Heart

The critical event in the development of cardiogenic shock is severe impairment of heart muscle contractile performance. Cardiac performance is primarily determined by four factors: preload, afterload, strength of contraction, and heart rate. Preload is defined as the force exerted on myocardium at the end of ventricular filling. If the filling is inadequate (low preload), cardiac output will be reduced. This finding is best exemplified by extracardiac obstruction, such as pericardial tamponade (blood in the pericardial sac that surrounds the heart) or by the reduction of venous return to the heart caused by high thoracic pressure (such as occurs with a pneumothorax, which is air under tension between the ribs and lung). Afterload is the resistance to emptying of the ventricles with myocardial contraction after the opening of the pulmonary and aortic valves. A rapid increase in afterload as in valvular stenosis leads to decreased volume of blood ejected from ventricles and decreased cardiac output. Contractility is most affected by loss of heart muscle as a result of myocardial infarction (heart attack). In the acute setting of myocardial ischemia, loss of at least 40% of left ventricular heart muscle results in severe depression of cardiac performance and shock.

A similar picture may also result from myocarditis (an inflammation of the heart muscle) and with prolonged cardiopulmonary bypass during cardiac surgery. In addition, myocardial stunning may occur following reversible myocardial ischemia. If the heart rate is too fast as in ventricular dysrhythmias, this may compromise ventricular filling and decrease cardiac output. If the rate is too slow, cardiac output may be insufficient, and shock may ensue. Both high and low heart rates are common complications of myocardial infarction (1–3).

Loss of Volume

Hypovolemic shock is caused by a reduction in intravascular circulating volume to a point where compensation is no longer possible, by constriction of venous capacitance vessels, to maintain cardiac filling. The loss of circulating volume may result from hemorrhage, dehydration, or leakage from the circulation into the body tissues. Trauma and gastrointestinal bleeding are common causes of rapid blood loss and lead to a reduction in preload and cardiac output. However, the oxygen (O_2) carrying capacity of blood is not severely impaired except in massive blood loss causing a decrease in hemoglobin concentration. Dehydration results in intravascular volume reduction with an increase in O_2 carrying capacity, since the number of red blood cells per unit of volume will increase, but the cardiac output is decreased.

Abnormal Distribution

Maldistribution of blood flow occurs with widespread vasodilation usually caused by infectious agents (septic shock) but vasodilation with low systemic vascular resistance may be caused by other mechanisms including endocrine disease, anaphylaxis (severe systemic allergic reaction), and neurogenic shock.

PATHOPHYSIOLOGY

The underlying result of shock of any etiology is hypoperfusion at the cellular level due to an inability to provide the cell with adequate oxygen, glucose, and other nutrients necessary to maintain normal functions. Consumption is calculated by the Fick equation, where cardiac output = O_2 consumption % (arterial-mixed venous O_2 content). If the cardiac output is low, the tissue blood flow is reduced.

Hypoperfusion and hypovolemia due to blood loss decrease the blood pressure and cardiac output. With the loss of up to 15% of total blood volume, no detectable changes in heart rate and blood pressure may be found because venous capacitance vessel constriction compensates for hypovolemia and maintains cardiac filling pressures. If fluid loss is 15–30% of blood volume, that is, 750–1500 mL in a 70 kg male, heart rate will increase tending to maintain cardiac output. Pulse pressure (the difference between systolic and diastolic blood pressure) decreases due to a rise in vascular resistance mediated by circulating catecholamines. The systemic vascular resistance (SVR) increases by constriction of arterioles, which tends to maintain an adequate arterial pressure. However, an increase in SVR raises arterial pressure, but unless it is accompanied by an increase in cardiac output, it has no effect on tissue blood flow. The relationship between flow and pressure is as follows: flow = pressure/resistance, and can be applied to the entire circulatory system, or to a single organ or even an electric circuit (Ohms law).

If the heart produces a constant flow per unit time (cardiac output), then the tissue blood flow (perfusion) will be inversely proportional to the vascular resistance. Therefore, in the case of shock, capillary perfusion should be globally reduced. However, this is not the case, as different organs have variable blood flow and oxygen utilization. The kidneys have high blood flow and little O_2 extraction, whereas the heart has relatively low blood flow (because it is contracting) and high O_2 extraction (Fig. 1) (4). In shock, vasoconstriction occurs preferentially in certain regions (skin, muscle, viscera), diverting the blood flow from these high vascular resistance organs to more vital organs with low resistance, such as heart and brain. If blood loss continues beyond 30–40% of total blood volume, the compensatory mechanisms fail and the blood pressure will fall, accompanied by marked tachycardia, changes in mental status, and decreased urinary output (1).

In the first stage of shock, arterioles are constricted and the amount of oxygen available to tissues may be insufficient for their metabolic needs. The global oxygen consumption will decrease after O_2 extraction from hemoglobin has been maximized (from a normal of 25% of the O_2 carriage to up to 70%). When O_2 consumption falls below a critical level, local metabolism becomes anaerobic, which leads to an increased production of lactic acid by the cells. Later on, this metabolic acidosis causes relaxation of the precapillary sphincters while the postcapillary venules are still constricted. Therefore, the capillaries become overfilled with blood and the hydrostatic pressure

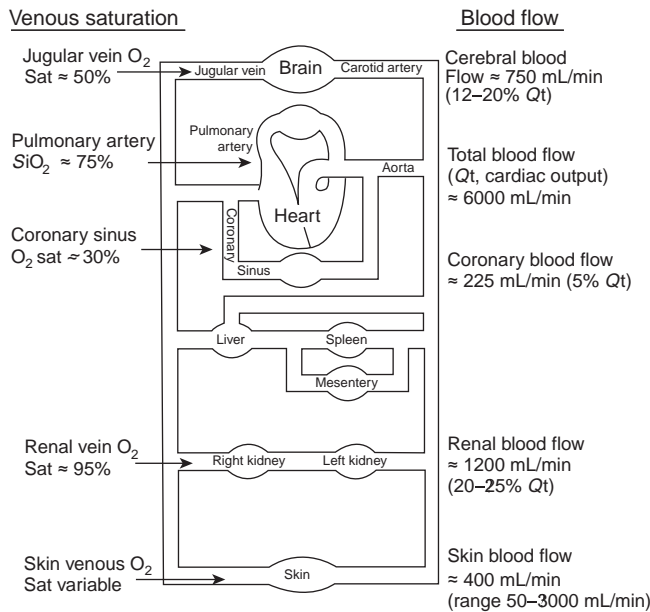


Figure 1. Venous oxygen saturations of blood leaving various organs are shown on the left side and blood flow expressed in mL/min and as a % of total cardiac output (Q_t) are shown on the right side of this figure. Note that the heart and brain with low blood flow relative to their oxygen requirements (e.g., coronary blood flow and carotid artery) have low venous oxygen saturations. In comparison, an organ such as the kidney has high blood flow, but little oxygen requirement and contributes relatively more to the final mixed-venous O₂ saturation of 75% in the pulmonary artery than does the much smaller venous blood flow from the heart and brain. For this reason, a normal pulmonary artery oxygen saturation is not a good indicator of adequate shock resuscitation of the brain or heart.

increases such that there will be loss of plasma into the interstitial space, further depleting the circulating volume. Hemoconcentration occurs and the blood viscosity is increased, causing slowing of the blood flow through the microcirculation. A vicious circle is established, where slow flow causes an increase in viscosity, which in turn leads to a decrease in flow, and so on.

In the latter stages, capillaries are filled with a sludge of red blood cells, which impairs the local flow. In addition, there is further reduction in tissue exchanges by shunting of blood through arteriovenous anastomoses so the functional flow is nearly zero. Therefore, cells lack O₂ and anaerobic metabolism will proceed to a point where cells are no longer able to survive. If a significant number of cells die, the organ function will be compromised. The lung is the first organ to fail, approximately 3–7 days after the primary shock event. The condition that results is called adult respiratory distress syndrome (ARDS) and is characterized by an increase in physiologic shunting and refractory hypoxemia. Kidney involvement is apparent within 2–5 days with acute renal failure due to ischemic tubular necrosis, followed by electrolyte disturbances and metabolic acidosis. Clearance of creatinine by the kidney approximates glomerular filtration rate and when this falls <25 mL/min, there is early onset of renal failure that is still

potentially reversible (5). If liver failure occurs, the first sign is jaundice, but the most significant evidence of failure is abnormal hepatic metabolism, with impaired protein synthesis and an inability to process available energy substrates. Gastric hemorrhage may occur as a late manifestation and is usually precipitated by coagulopathy, a common event in shock. The immunologic response is depressed leading to an increased susceptibility to infections. The syndrome in which all these events occur—multiple organ system failure—is a terminal event common to all types of shock (3).

SHOCK ASSESSMENT

Assessment and management of a patient in shock are accomplished simultaneously. Since evaluation of cellular metabolism cannot be done directly, the physician must rely on surrogate clinical findings such as blood pressure (BP), heart rate, skin temperature, urinary output and mental status, and on data obtained by using various monitoring devices.

Measurement of BP is routine in shock patients. The systolic blood pressure measure (SBP) is not a good indicator of blood loss, as up to 30% of circulating volume may be lost without any change in SBP. Instead the diastolic blood pressure (DBP) is more sensitive and is usually elevated in shock due to peripheral vasoconstriction. Therefore, mean arterial pressure defined as $MAP = DBP + 1/3 \text{ pulse pressure (SBP - DBP)}$ seems to be more useful for BP monitoring. Blood pressure measured by auscultation is inaccurate in patients with low peripheral blood flow. Invasive measurement using an intra-arterial catheter inserted into radial, brachial, axillary, or femoral arteries is more precise and provides continuous monitoring and easy access for blood gas and pH analysis.

Change in heart rate occurs with increasing blood loss, values >100/min in adults are detectable before any change in SBP. Pulse pressure is usually low due to increased DBP. Another clinical monitor is capillary refill time, which is the time for return of blood flow after compression of the nail beds until blanching occurs. It is prolonged >2 s in severe peripheral vasoconstriction. Skin temperature correlates well with peripheral blood flow and a difference of 3–6 °C between toe and rectal temperature reflects severe peripheral vasoconstriction (2). The ability of commonly used clinical parameters to quantify acute hemorrhage is shown in Fig. 2 (6). Base deficit, mean arterial pressure, serial hemoglobin, and serum lactate are related to blood loss.

Preload is assessed by measuring central venous pressure (CVP), which correlates with right atrial pressure, with a catheter inserted in the superior vena cava via the jugular or subclavian veins. Normal CVP is 5–12 cm of H₂O and values <5 are generally found in hypovolemic states and indicate the need for assessment of reserve cardiac function by rapid fluid administration together with assessment of changes in CVP. However, CVP does not always correlate with fluid requirement because an increase in pulmonary vascular resistance (hypoxia, acidosis, increased intrathoracic pressure) may be associated

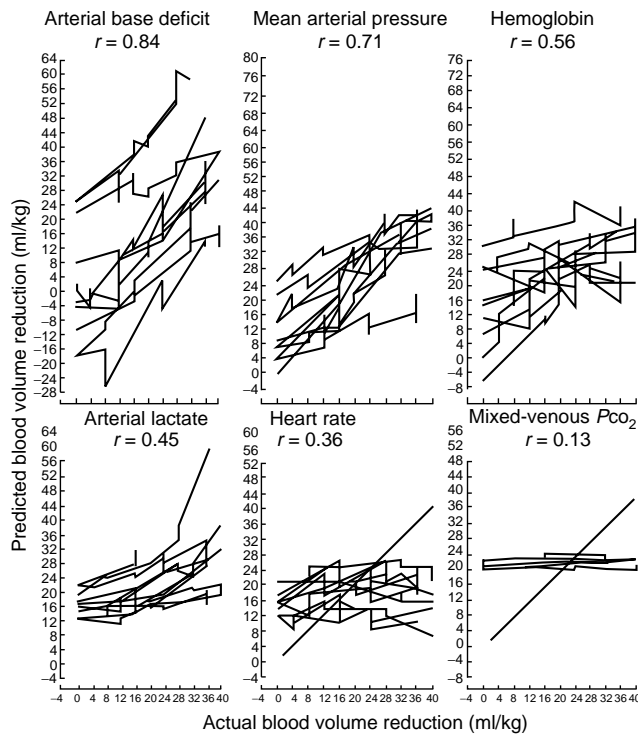


Figure 2. Commonly used clinical and laboratory values compared to quantity of acute hemorrhage. Graphs show predicted versus actual blood volume reductions for six representative parameters. Solid black lines represent an individual animal ($n = 10$). Gray bar represents the ideal in which predictions equal actual blood volume reductions. For mean arterial pressure, predictions at large volume hemorrhage were more accurate than predictions with small volume bleeds. Models such as heart rate and lactate both showed significance variability before hemorrhage among animals and flat slopes (e.g., mixed-venous PCO_2) indicated fixed-volume predictions regardless of actual degree of hemorrhage. (Reproduced with permission from Waisman Y et al. *J. Appl. Physiol.* 1993;74:410–519.)

with high CVP, reflecting right ventricular failure, even when there is considerable blood loss.

Fluid Challenge

Reserve cardiac function is assessed by means of a fluid challenge until CVP pressures are elevated at least 2 mmHg above the baseline for 10 min after fluid infusion ceases. There are four possible outcomes when 250 mL boluses of fluid are given over 5 min (Fig. 3). Outcome No. 1: Filling pressures rise with the challenge and continue to rise even after fluid infusion ceases. If cardiac output is measured (see below), there is no increase with elevation of filling pressures and the heart has limited reserve function and is not able to deal with the increased fluid load by increasing contractility by the Frank–Starling mechanism (this mechanism describes the property of heart muscle increasing its contraction in proportion to fiber length, up to a maximum point when contractility decreases and cardiac failure occurs). Further fluid infusion (when this response occurs) is expected to produce cardiac failure. The therapeutic indication this response dictates is to restrict

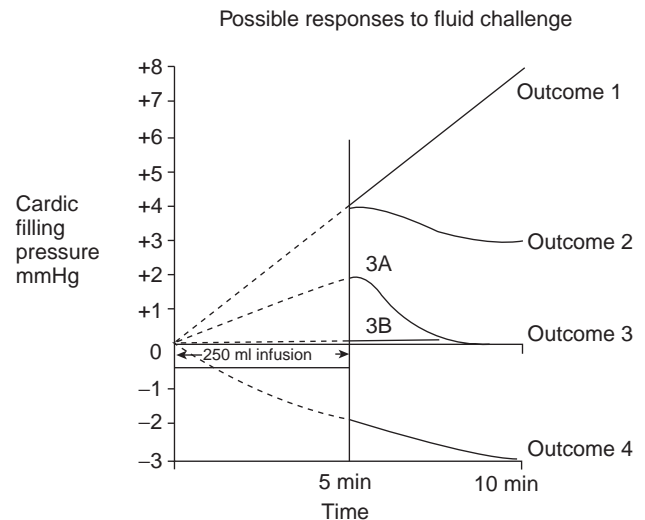


Figure 3. The possible responses to rapid (50 mL/min for 5 min) administration of a fluid is shown schematically. The vertical axis shows the change in cardiac filling pressure (could be either CVP or PAQP) in response to fluid. Time is on the horizontal axis. Fluid infusion ceases after 5 min and the filling pressures are remeasured (and cardiac output measured if available). Fluid challenge in outcomes 3A and 3B should be repeated until filling pressures increase and remain elevated for 5 min after ceasing fluid infusion (outcome No. 2). (Reproduced with permission from Mackenzie CF et al. *J. Neurosurg.* 1985;62:843–849.)

fluid infusion and reduce myocardial depressant agents. If the trend continues, inotropic agents are required to increase cardiac contractility and reverse myocardial depression. Outcome No. 2: Central venous pressure or pulmonary capillary wedge pressure (PCWP) rise 3–4 mmHg, but then falls to a level 2 mmHg above baseline, indicating myocardial contractility is adequate for the increase in cardiac preload. Management should be to infuse fluids and maintain cardiac filling pressures within this range for optimum cardiac output and oxygen transport for the prevailing vascular tone. In outcome No. 3, filling pressures either rise briefly (3A) or do not rise at all (3B) with the 250 mL fluid challenge indicating that the patient has considerable reserve cardiac function and a greater fluid load could be tolerated. If the patient has low urine flow, has evidence of poor tissue perfusion, such as acidosis or low mixed venous oxygen tension, inotropic agents, or diuretics should not be given. Rather, fluid infusion should continue at the same rate until outcome No. 2 is obtained. In fact, because filling pressures have not increased in outcome No. 3, it is unlikely that myocardial fiber length would increase and, therefore, cardiac contractility would not have changed by the Frank–Starling mechanism. Outcome No. 4 is seen in ~5% of cases and results in a fall in filling pressures when fluid is infused rapidly. The two most likely explanations are either that rapid fluid infusion has a vasodilator effect on peripheral vasculature and reduces left ventricular afterload, increasing cardiac output, or alternatively, it may be related to a reduction in heart rate seen when fluid is infused in the hypovolemic patient. The decreased heart rate allows more

time for myocardial perfusion, which occurs mostly during diastole or cardiac relaxation. Cardiac output increases and the Frank–Starling function curve shifts to the left due to improve myocardial oxygenation (7).

Cardiac Catheterization

Catheterization of pulmonary artery (PA) is a method of hemodynamic monitoring that can be used to assess right and left ventricular function as well as quantitate the proportion of blood shunting and to calculate tissue delivery and O_2 extraction. Catheterization is usually reserved for chronically ill patients with heart disease or shock refractory to conventional therapy, because it requires an invasive flow directed, balloon tipped catheter to be floated in the blood stream through the right side of the heart into the pulmonary artery. Such heart catheterization has complications including infection and cardiac rhythm irregularities. It provides data on PA pressure, PCWP, vascular resistances, cardiac output (commonly by a thermodilution technique), and also allows blood sampling from PA, for measurement of mixed-venous saturation (SvO_2) and calculation of intrapulmonary shunting of blood.

Equations

Normal SvO_2 is ~75% (Fig. 1), and it reflects the ratio between oxygen delivery [arterial O_2 content (CaO_2) \times cardiac output (Qt)] and oxygen extraction [ratio of O_2 consumption (VO_2) and O_2 delivery]. Consumption is calculated by the Fick equation, where cardiac output = O_2 consumption % (arterial mixed-venous O_2 content). If the cardiac output is low, the tissue blood flow is reduced.

Blood gas and pH analysis give a rough estimate of oxygen utilization and cellular metabolism by calculation of bicarbonate and base deficit. A low arterial pO_2 (hypoxemia) may accompany shock and be found before any clinical sign due to ventilation–perfusion inequality and increased venous admixture by shunting (see above). The causes of hypoxemia include hypoxic hypoxemia (low inspired O_2), anemic hypoxemia (low hemoglobin for carrying O_2) stagnant hypoxemia (low cardiac output for delivery of O_2), and histotoxic hypoxemia (poisoning of the enzyme systems used to offload O_2 from hemoglobin at the tissues). Blood lactate is an indicator of tissue hypoperfusion and anaerobic metabolism and is elevated in patients with low cardiac output. Lactate is a clinically useful marker of the amount and duration of shock. Sustained reduction in elevated lactate is an important clinical marker of recovery. Arterial pH can be high or normal despite metabolic acidosis, because of low pCO_2 due to increased respiratory rate and alveolar ventilation, common in patients with low cardiac output (1,3,8).

Real-Time Noninvasive Measures of Systemic Perfusion

Sublingual capnometry (measurement of sublingual pCO_2 – $PsICO_2$) is a new technique for assessment of systemic perfusion failure. It is based on elevated pCO_2 in tissues with low blood flow due to intracellular buffering of

hydrogen ions by bicarbonate. Elevated $PsICO_2$ correlates well with increased blood lactate and low mean arterial pressure (MAP), markers of tissue hypoperfusion. $PsICO_2$ has the advantage of prompt indication and continuous monitoring of tissue flow reversal, unlike lactate whose clearance presents significant delay. A threshold value of 70 mmHg for $PsICO_2$ has been identified that is predictive of both the severity state and survival. A $PsICO_2$ >70 mmHg is highly predictive of circulatory failure whereas readings <70 mmHg are highly predictive of survival (9). A similar technique to sublingual capnometry is gastrointestinal tonometry, which measures gut mucosal pCO_2 . Calculation of intramucosal gut pH is possible (pHi) using the Henderson–Hasselbach equation: $pHi = 6.1 + \log (HCO_3^- / \alpha^* \text{ tonometer } pCO_2)$, where α is the solubility of CO_2 in plasma ($\alpha = 0.03$) Values of pHi <7.32 define mucosal hypoperfusion and are associated with high mortality rates (10).

Brain Perfusion

Brain perfusion monitoring is technically difficult and inaccurate. Clinically, signs of confusion, altered sensorium, agitation, and decreased consciousness give a rough idea about cerebral hypoperfusion. Jugular venous oxygen saturation ($SjvO_2$), transcranial cerebral oximetry, and brain tissue oxygen tension ($PbtO_2$) monitoring are the methods of monitoring brain oxygenation. Measurement of $SjvO_2$ is performed using a catheter inserted in the jugular bulb, the upper part of the internal jugular vein. Continuous monitoring of venous saturation without blood sampling is possible by using intravenous oximetry. This type of device has been used in patients with head injury and during anesthesia for neurosurgery. It provides only global brain oxygenation monitoring and is susceptible to errors. Cerebral oximeters using near-infrared spectroscopy seem to be a promising alternative. They can evaluate regional ischemia, hemoglobin saturation, and even concentration of oxygenated and reduced hemoglobin. However, these monitors can only assess trends, where each patient is their own control. There are no normative data for comparison and the boundaries of monitored brain tissue cannot be precisely defined. Brain tissue oxygen tension is measured with small catheters inserted directly into the brain tissue during a craniotomy or via a burr hole. These catheters measure pO_2 , pCO_2 , pH, and temperature. Some studies suggest a normal value for $PbtO_2$ of ~35 mmHg, whereas cerebral ischemia is usually defined as a $PbtO_2$ < 8 mmHg. These data were obtained in patients with traumatic brain injury, but their usefulness should be confirmed by further studies (11,12). Extra cellular glutamate and aspartate measures (obtained by microdialysis) are closely related to outcome after head injury (13). These markers were also related to the type of head injury and suggest that excitatory amino acids play a role in the evolution of brain injury.

SHOCK MANAGEMENT

Treatment of shock should be directed to its underlying cause. However, establishing an exact diagnosis can be

time consuming, so management is focused on simultaneously stabilizing the patient and proceeding with diagnostic tests to identify the cause of shock.

For cardiogenic shock, the goal is an increase in cardiac output by acting to change preload, afterload, contractility, or heart rate. Therapy is tailored using information provided by a PA catheter. Various drugs are given to increase contractility and to relieve pulmonary congestion. In unresponsive cases, additional measures may be considered, such as urgent myocardial revascularization in acute myocardial infarction, intraaortic balloon counterpulsation, and anatomic defects repair (ruptured valves). In extracardiac compression, relieving the pressure of pericardial tamponade by pericardial puncture or insertion of a chest tube for increased intrathoracic pressure due to pneumothorax is the treatment of choice, when these are the causes of impaired cardiac filling and decreased cardiac output.

In case of septic shock (the most common form of distributive shock), large quantities of fluids are administered to fill the vascular bed and maintain perfusion pressure. Cardioactive drugs are used only if cardiac output declines. At the same time steps are taken to identify and control the source of infection (3).

Hypovolemic shock requires initial rapid expansion of circulating volume. Fluid resuscitation is initiated with administration of crystalloid or colloid solutions through large-bore intravenous lines. Rapid infusion devices can be used to pump large amounts of fluids in <10 min. A potential adverse effect resulting from resuscitation with large amounts of fluid, when using rapid infusion devices, is a drop in body core temperature. Levels <35 °C are associated with impaired coagulation and depressed cardiac contractility (14). Covering the patient with inflatable warming blankets can prevent this complication, but the most effective method for rewarming is an extracorporeal countercurrent warmer through femoral artery and vein cannulation, which can elevate temperature ~6 °C in <30 min. During fluid infusion, hemodynamic parameters are continuously monitored and signs of instability (persistent SBP <90 mmHg) imply there is ongoing blood loss or shock has not been reversed. Classically, a hemoglobin (Hb) level <10 g/dL with continuous loss requires blood transfusion, but recent studies have demonstrated that this level can be as low as 7 g/dL without adverse effects in the majority of the population (15). Those with cardiac or cerebrovascular disease should be transfused at higher hemoglobin concentrations.

THE FUTURE OF SHOCK DIAGNOSIS AND MANAGEMENT

Future trends in shock diagnosis and management include identification of mediator's released in shock states and blockage of the release of harmful mediator factors while facilitating release of those with benefits. The field of proteomics, defining protein expression with shock, will

provide many future treatment and diagnostic opportunities. Genomics may identify some individuals or disease states susceptible to adverse outcomes from shock. These future findings could lead to improved outcome, particularly from septic shock, which has a high mortality and morbidity.

BIBLIOGRAPHY

1. Kelman GR. *Applied Cardiovascular Physiology*. London: Butterworths; 1977.
2. Hardaway G. *Shock. The Reversible Stage of Dying*. Littleton: KPGS Publishing Company; 1988.
3. Shoemaker WC, Appel PL, Kram HB. Role of oxygen transport in the pathophysiology, prediction of outcome and therapy of shock. In: Bryan-Brown CW, Ayres SM, editors. *Oxygen Transport and Utilization*. The Society of Critical Care Medicine 1987. p 65–92.
4. Mackenzie CF. Anesthesia in the shocked patient. *J Eur Emer* 1994;4:163–172.
5. Shin B, Mackenzie CF, Helrich M. Creatinine clearance for early detection of post-traumatic dysfunction. *Anesthesiology* 1986;64:605–609.
6. Waisman Y, Eichacker PO, Banks SM, Hoffman WD, MacVittie TJ, Natanson C. Acute hemorrhage in dogs: construction and validation of models to quantify blood loss. *J Appl Physiol* 1993;74(2):510–9.
7. Mackenzie CF, Shin B, Krishnaprasad D, Illingworth W. Assessment of cardiac and respiratory function during surgery on patients with acute quadriplegia. *J Neurosurg* 1985;62: 843–849.
8. Darovic GO. *Hemodynamic Monitoring: Invasive and Non-invasive Clinical Application*. Philadelphia: WB Saunders; 1995.
9. Weil MH, Nakagawa Y, Tang W, Sato Y, Ercoli F, Finegan R, Grayman G, Bisera J. Sublingual capnometry: a new non-invasive measurement of diagnosis and quantitation of severity of circulatory shock. *Crit Care Med* 1999;27(7):1225–1230.
10. Gutierrez G. *Gastrointestinal Tonometry*. Oxford Textbook of Critical Care. Oxford: Oxford University Press; 1999.
11. Symthe PR, Samra SK. Monitors of cerebral oxygenation. *Anesth Clin N Am* 2002;20(2):293–313.
12. Valadka AB, Gopinath SP, Contant CF. Relationships of brain tissue PO₂ to outcome after severe head injury. *Crit Care Med* 1998;26:1576–1581.
13. Gopinath SP, Valadka AB, Goodman JC, Robertson CS. Extracellular glutamate and aspartate in head injured patients. *Acta Neurochir Suppl* 2000;76:437–438.
14. Dunham CM, Belzberg H, Lyles R, Weireter L, Skurdal D, Sullivan G, Esposito T, Hamini M. Resuscitation of hypovolemic traumatic patients. *Resuscitation* 1991;Apr 21(2–3), 207–229.
15. Herbert PC, Wells G, Blajchman MA. A multicenter randomized controlled clinical trial of transfusion requirements in critical care. *N Eng J Med* 1999;340:409–417.

See also BLOOD PRESSURE MEASUREMENT; CARDIOPULMONARY RESUSCITATION; PERIPHERAL VASCULAR NONINVASIVE MEASUREMENTS.

SHUNT, FOR HYDROCEPHALUS. See HYDROCEPHALUS, TOOLS FOR DIAGNOSIS AND TREATMENT OF.

SIMULATION. See PHYSIOLOGICAL SYSTEMS MODELING.

SINGLE PHOTON EMISSION COMPUTED TOMOGRAPHY. See COMPUTED TOMOGRAPHY, SINGLE PHOTON EMISSION.

SKIN SUBSTITUTE FOR BURNS, BIOACTIVE

ROBERT H. DEMLING
 LESLIE DeSANTI
 Harvard Medical School
 Boston, Massachusetts

INTRODUCTION

A body burn is a complex injury process resulting in local changes in skin integrity as well as profound systemic changes in fluid and electrolyte balance, metabolism, and immune defenses. Severe psychosocial changes also occur in addition to long-term, often permanent changes in skin function. Major advances in care have resulted in a marked decrease in mortality and also morbidity, especially with massive burns. In addition to survival, the current focus in burn care is on improving the long-term function and appearance of the healed or replaced skin cover as well as quality of life.

This focus on quality has generated a significant interest in the use of skin substitutes to be used to improve wound healing, to control pain, to more rapidly close a burn wound, to improve functional and cosmetic outcome, and, in the case of massive burns, to increase survival.

To more effectively address these new roles, the new generation of skin substitutes are developed to be biologically active. The concept behind providing bioactivity is that the wound healing process can be substantially improved as compared with a simple synthetic barrier-type dressing. It remains to be seen just how much better this new generation of products will impact the burn wound. To date, the new products to be discussed have not displaced the more inert standard burn wound dressings, but rather are used in conjunction and for quite specific indications.

The skin substitutes are initially classified according to whether they are to be used as a temporary wound covering to decrease pain and augment healing or a permanent skin substitute to add to or replace the remaining skin components.

The ideal properties and indications for these products will be better clarified after a discussion of the function of normal skin and the effect of a burn on skin integrity.

THE NORMAL PROPERTIES OF SKIN

Normal skin is a very complex bilayer organ with a wide variety of properties mainly protective barriers, which are critical to survival (1–5). Loss of these protective barrier functions occurs with a skin burn. Restoration of skin structure and function become the necessary properties

of temporary and permanent skin substitutes. The skin is a bilayer organ with each layer having specific functions. But both layers are needed for proper skin function. An outer thin epidermal layer covers an inner thicker dermal layer.

Epidermis

The outer thinner layer known as the epidermis is composed mainly of epithelial cells or keratinocytes (1–3). The deepest epidermal cells are immature cells and are continually dividing and migrating toward the surface, to replace lost surface cells; keratinocytes. The same types of regenerating epidermal cells are found in hair follicles and other skin appendages, which are anchored in the dermis. As the cells mature and migrate to the surface, they form keratin, which becomes an effective barrier to environmental hazards such as infection and excess water evaporation.

Stratum Corneum. The stratum corneum is the “outermost” layer of the epidermis consisting of several flattened layers of dead keratinocytes as well as keratin. This layer protects against entry of bacteria and toxins. The epidermal layer regenerates every 2–3 weeks, but regeneration requires the structure and functional components of the dermis.

Skin Functions: Epidermis (Outer Layer).

- Protection from drying.
- Protection from bacterial entry (infection).
- Protection from toxin absorption, like chemicals on the skin.
- Fluid balance: avoiding excess evaporative water loss that would cause dehydration.
- Neurosensory (touch, pain, pressure, sensation).
- Social-interactive (visible portion of the body covering).

The epidermis is firmly anchored to the dermis by the rete pegs, which are ingrowth of epithelial cells interdigitating into the upper dermis like the teeth on a saw.

Dermis

The dermis (1,4,5) is the deeper skin layer responsible for skin durability, and the barrier functions of controlling body temperature, and flexibility and barrier function. The nerves for touch and pain, blood vessels, and hair follicles are present in the dermis. The dermis is responsible for orchestrating the formation of the epidermis.

Skin Functions: Dermis (Inner Layer).

- Regulation of body temperature avoiding hypothermia and hyperthermia.
- Properties of elasticity and durability necessary for movement.
- Epidermal regeneration.

The imbedded epidermal cells can multiply and re-form an epidermal structure under the direction of growth

factors and cell signals found in the dermis (6). The dermal signals are richest in the upper third of the dermis known as the papillary dermis. The deeper dermal layer is less able to regenerate epidermis and itself, therefore the thinner the thickness of remaining dermis in a wound, the less likely it is that the skin can regenerate. More scar will develop to replace the lost skin. If the dermis is totally destroyed, a burn cannot heal by itself and a skin graft or permanent skin substitute is required.

Of extreme importance is the psychological impact in the quality of healed or replaced skin, as this organ has a major role in human communication and helps define the individual. One of the key objectives, of the newer skin substitutes, is to restore some normalcy to the new skin cover.

BURN INJURY

A skin burn is the damage to the structure and function of skin, caused by heat or other caustic materials (7,8). Severity is based on the degree to which the outer epidermis and the inner dermis are destroyed or damaged. The most immediate and obvious injury is one due to heat. Excess heat causes rapid protein denaturation and cell damage (9,10). The depth of heat injury is dependent on the depth of heat penetration. In addition, the body's response to the burn in the form of inflammation and inflammatory mediator release, especially oxidants, results in further cell damage. The damage to skin caused by a burn is therefore a very dynamic process starting with a heat or chemical insult, and then evolving with time, especially in the deeper higher risk burn insult. The initial thickness of the skin is also a major factor as to severity as the thinner the initial skin the more severe will be the burn for the same heat insult. Children and the elderly have very thin skin (7,8).

Burn severity is determined by the depth of the burn, the burn size relative to the percent of total skin burned, and the location. The greater the function (e.g., hands) or cosmetic importance (e.g., face) the more severe the burn. Other factors include age and status of overall health. The very young and very old are at a greater risk with a burn due to an impaired ability to tolerate severe bodily trauma as well as thinner skin. The body response and postburn complications, especially infection, add to severity (7–10).

Burn Size

Burn size is defined as the percent of the persons body skin burned. In the adult, the "Rule of Nines" assessment is commonly used. Each area is considered 9% of total body surface (TBS), each leg 18%, back 18%, front torso 18%, and head being 9% of total. The palm of the patients' hand is considered to be 1% of that persons body skin surface area.

Burn Depth

Burn depth is defined by how much of the two skin layers are destroyed (Fig. 1). Burns can be categorized by degree:

- First degree: confined to the outer layer only.
- Second degree: also involves part of the dermis.
- Third degree: destruction of both layers of skin.

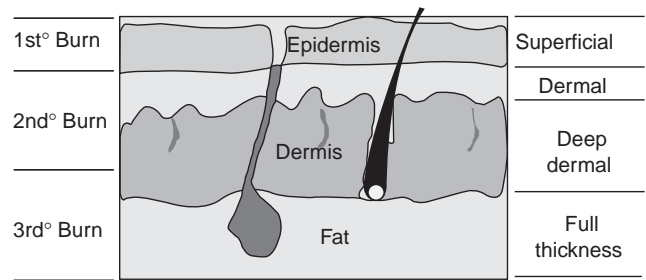


Figure 1. Schema describing burn depth with terminology.

More recent terminology is

Partial thickness: a second degree burn consisting of injury to part of the dermis.

Full thickness: a third degree burn consisting of necrosis to both layers.

Partial thickness or second degree is further divided into superficial-confined to epidermis and upper dermis and deep: when most of the dermis is destroyed.

Full thickness is further categorized as burns involving destruction of the skin alone and burn extending below skin into (e.g., muscle). The latter is often referred to as a fourth degree burn.

Superficial Second Degree

Involves the entire epidermis and no more than the upper third of the dermis is heat destroyed. Rapid healing occurs in 1–2 weeks because of the large amount of remaining skin and good blood supply. Scar is uncommon. However, in young children and the elderly, with an already thin dermis, a burn of this type actually extends to mid-dermis increasing the healing rate with an increasing risk of scarring, especially if re-epithelialization takes >3 weeks.

Initial pain is the most severe of any burn, as the nerve endings of the skin are now exposed to the air. Protection of the burn wound surface from desiccation, inflammation, and infection is necessary to optimize healing. Temporary skin substitutes are used, especially in children, to protect the wound and relieve the pain while it re-epithelializes beneath (11,12).

Deep Second Degree (Deep Partial Thickness) Burn

Most of the skin is destroyed except for small amount of remaining dermis. The wound looks white or charred indicating dead tissue. Pain is much less as the nerves are actually destroyed by the heat. Usually, one cannot distinguish a deep dermal from a full thickness (third degree) by visualization. The presence of sensation to touch usually indicates the burn is a deep partial injury rather than full thickness. There is a high risk of infection due to decreased blood flow and impaired local immune defenses (1,2). This burn typically takes several months to re-epithelialize due to few remaining epidermal cells. The quality of the epithelial covering is typically poor, being thin and friable. Wound scar is usually severe with healing (13).

Table 1. Mean Survival Rate after Burns in Burn Centers,^a age versus Burn Size

Burn size, % of total skin ^b	Age (years)						
	0–1	2–4	5–34	35–49	50–59	60–74	>75
0–10	>95	>95	>95	>95	>95	>95	90
10–20	>95	>90	>95	>90	>90	>70	>60
20–30	>90	>90	>90	>90	>75	50	35
30–40	75	80	90	80	70	40	<20
40–50	50	65	80	60	40	10	<10
50–60	50	60	70	60	40	<25	<10
60–70	40	50	60	40	25	<10	0
70–80	35	40	45	30	25	0	0
80–90	30	35	30	30	<20	0	0
90–100	20	20	20	15	0	0	0

^aSee Refs. (10) and (14).

^bThe body surface burn is the combined second and third degree area relative to total body surface (TBS) (adapted from current literature).

Typically this depth of burn is managed by early surgical excision of the dead tissue and coverage with a skin graft (or permanent skin substitute) (11,12,14).

Third Degree (Full Thickness) Burn

Both layers of skin are completely destroyed leaving no cells to heal. Any significant burn will require skin grafting. Small burns will heal with scar. There is typically no initial pain as nerve endings are destroyed. Because of the high risk of infection and inability to heal primarily third degree burns are typically managed by early excision of the dead burn tissue and skin grafting (or permanent skin substitute) (11–14).

Survival Rate

As can be seen from Table 1, many young patients with massive burns now survive in burn centers, increasing the need for both temporary and permanent skin substitutes.

Note that survival is very high for even massive burns (<75% of the body) in older children and in the young adult. Survival decreases in the elderly due to the presence of pre-existing disease and inability to withstand severe stress. Survival is also less for babies and toddlers for the same size (14).

Burn Scar and Pain Relative to Depth

The initial problems of pain and later problems of scar relate to burn depth and healing time (Table 2) (13,15,16).

Pain is a major problem with superficial burns. Temporary skin substitutes markedly decrease initial pain. Scarring is a result of loss of a large amount of dermis and a resulting prolonged wound healing period (13–17). The wound healing process includes an initial inflammation followed by an increase in wound fibroblasts, new vessels, and epithelial cell proliferation (if the burn is superficial). By 7 days, the fibroblasts are producing increased amounts of collagen, which persists until the wound has healed or is closed. Wound inflammation persists as long as the wound is open.

In superficial burns, epidermal regeneration will be relatively rapid, if the wound environment is optimized. The injured dermal elements are usually covered by new epithelium within 2 weeks if protected from environmental insults. Only modest amounts of collagen are deposited. The wound usually becomes relatively pliable with time and minimal to no wound contraction is seen. Cosmetically, the superficial second degree burn, which heals in 2 weeks, results in very minimal to no long-term scarring (15–17).

The histology of the wound bed, however, changes dramatically if it has not been re-epithelialized by 3 weeks, as would be the case with a deeper dermal burn (13–17). Fibroblasts and macrophages become the predominant cells. Large numbers of myofibroblasts also enter the wound. Besides leading to contraction, these cells continue to deposit large quantities of collagen and glycosaminoglycans. Later, closure of this wound by re-epithelialization or

Table 2. Etiology and Prognosis Relative to Burn Depth^a

Second degree (partial thickness)	Cause	Appearance	Pain	Healing	Scar
Superficial	Hot liquid, short exposure	Wet, pink blisters	Severe	10–14 days	Minimal
Mid-dermal	Hot liquid, hot grease, or flash flame with longer exposure	Less wet, red ± blisters	Moderate	2–4 weeks	Moderate
Deep dermal	Chemicals, direct contact with flames	Dry, white	Minimal	8–16 weeks	Severe, usually needs skin graft
Third degree (full thickness)	Chemicals, flames, explosion with very high temperature	Dry, white or char	None	Needs skin graft	Mild to severe, depending on timing and type of graft

^aThe short- and long-term impact of a burn is described relative to depth.

grafting does not eliminate the stimulus for ongoing scar formation. The components produce a harder, less pliable wound shortening the scar and causing fusion of the collagen fibers in the contracted site. The process of wound contraction will lead to joint contractures.

Also, there are typically no finger like structures called rete pegs produced if healing time exceeds 4–6 weeks and the epidermis is not well anchored such that blistering is common (13,16).

The risk of hypertrophic scar formation increases with the healing time of deeper burns. Hypertrophic scar is an excess scar formation leading to a red, raised, itchy, nonpliable skin cover (13,15,16).

ROLE OF BIOACTIVE SKIN SUBSTITUTES

Outcome is defined both in terms of survival, the quality of the healed or grafted skin, and the degree of morbidity from the burn. The final objective of burn management is achieving survival as well as minimizing the morbidity such as loss of muscle and strength, maintaining quality of life, minimizing scarring, and optimizing the healing process (7,8).

The major stimulus for advances in skin substitutes is to improve the quality of the closed burn wound and avoid poor quality skin (18–20). As can be seen in the table describing burn survival, there are now many massive burns that survive necessitating the use of skin substitutes. Remaining nonburn skin is not sufficient to close these massive burns (11,12,14).

A superficial burn involves the epidermis and very little of the dermis, and the dead tissue peels off, in the form of blisters, leaving a viable wound bed that must be protected. Temporary skin substitutes can improve the healing while decreasing pain (16,17). With deeper burns involving some or all of the dermis, the dead tissue, remains adherent to the wound. The dead tissue known as the burn eschar, will then cause an inflammatory response producing both local and systemic effects. The systemic effects include a profound increase in metabolic rate with a marked increase in muscle wasting, and impairment in immune defenses (1,2). Controlling this systemic response, by earlier removal of the dead burn tissue and closure of the wound, has markedly decreased overall mortality morbidity (11–13). Skin substitutes are used to temporarily or permanently close the excised burn, especially in large burns where there isn't enough remaining skin to use for skin grafts (18–21).

The addition of biological activity to the skin substitute, is intended to improve the healing process with the intention of more rapidly healing a superficial burn and restoring valuable dermal components in a deeper burn wound bed thereby minimizing scarring and optimizing function (18–21). As stated before, the impact of the added bioactivity in burns is yet to be firmly established. A list of the noncellular components of dermis, used in available skin substitute, is shown below (22–26) (Table 3). Epidermal and dermal cells are usually also added to dermal elements in permanent skin substitutes in addition to these dermal components.

Table 3. Components of Dermis that Are Involved in Healing and Used in Skin Substitutes

Dermal Components Stimulating Healing^a

Structural component or scaffolding
Biologically active component stimulating all phases of healing
Collagen (protein)
Scaffold for cell migration and matrix deposition
Cell guidance
Elastin (protein)
Tissue elasticity
Fibronectin (protein)
Cell-to-cell adherence
Contact orientation for cells
Increases epithelial cell division, migration
Chemo attractant for fibroblasts, macrophages
Growth factors (proteins)
Stimulate all phases of wound healing
Glycosaminoglycan (glycosylated protein)
Cell adherence properties
Conduit for healing factors
Deactivator or proteases
Scaffold or foundation for dermal elements
Hyaluronic acid (complex carbohydrate)
Maintaining matrix hydrated
Decreases inflammation
Stimulates healing
Proper cell alignment

^aSee Refs. (22–26).

AVAILABLE BIOACTIVE SKIN SUBSTITUTE

A list of skin substitutes, categorized by biologic make-up, is presented below (Table 4). All have some degree of biologic activity for improving the wound-healing environment. A disadvantage of all of these skin substitutes is the absence of active antimicrobial activity. However, early effective wound closure does decrease the risk of infection.

Skin substitutes can also be categorized as to use and indication into temporary or permanent.

Table 4. Available Skin Substitutes Are Categorized Based on Composition

Available Biologically Active Skin Substitutes

Naturally occurring tissues
Cutaneous allografts
Cutaneous xenografts
Amniotic membranes
Porcine small intestinal submucosa
Composite Synthetic-Biological
Collagen based dermal analogs
Integra
Culture-derived tissue
Bilayer human tissue
Cultured autologous keratinocytes
Fibroblast seeded dermal analogues
Epithelial seeded dermal analogue

Skin substitutes can also be categorized as to use and indication into temporary or permanent.

Table 5. Ideal Properties of a Temporary Skin Substitute^a

<i>Ideal Properties of a Temporary Skin Substitute</i>
Rapid and firm adherence properties for closure of the wound
Relieves pain
Easily applied and secured
Does not incite inflammation
Stimulates wound healing
Barrier to microorganisms
Avoids wound desiccation
Optimizes healing environment
Does not cause hypertrophic tissue response
Hemostatic
Prevents evaporative water loss
Flexible yet durable
Easy to remove when
Wound has re-epithelialized
Wound ready for grafting
Cannot transmit disease
Inexpensive
Long shelf life
Does not require refrigeration

^aThese properties are then sought when developing new skin substitutes.

Temporary skin substitutes are used to help heal the partial thickness burn (or donor site) and close the clean excised wound until skin is available for grafting. There are typically no living cells present.

Permanent skin substitutes are used to replace lost skin providing either epidermis or dermis, or both and to pro-

vide a higher quality of skin than a thin skin graft. Most permanent skin substitutes contain viable skin cells as well as components of the dermal matrix.

Temporary Bioactive Skin Substitutes

The purpose of a temporary skin substitute is twofold (Table 5). Temporary skin substitutes are typically a bilayer structure. There is an outer epidermal analog and a more biologically active inner dermal analogue (18–21). The first objective is to close the wound, thereby protecting the wound from environmental insults (18,19,27–29). The second objective is to provide an optimal wound healing environment by adding dermal factors that activate and stimulate wound healing (18–29). Biologically active dermal components naturally are typically provided to the inner layer, which is then applied to the remaining dermis in a partial thickness burn or to an excised wound. Below is a list of the commonly available dermal matrix elements present in these products, and their actions.

The currently available products are listed in Table 6.

Human Allograft (Cadaver Skin)

Human allograft is generally used as a split-thickness graft after being procured from organ donors (30–32). When used in a viable fresh or cryopreserved state, it vascularizes and remains the “gold standard” of temporary wound closures. It can be refrigerated for up to 7 days, but must be stored frozen for extended periods. It is also used in a

Table 6. Available Bioactive Temporary Skin Substitutes^a

Product	Company	Tissue of Origin	Layers	Category	Uses	How Supplied
Human allograft	Skin bank	Human cadaver	Epidermis and dermis	Split thickness skin	Temporary coverage of large excised burns	Frozen in rolls of varying size
Pig skin xenograft	Brennan Medical St. Louis, MO	Pig dermis	Dermis	Dermis	Temporary coverage of partial thickness and excised burns	Frozen or refrigerated in rolls
Human amnion	On site procurement	Placenta	Amniotic membrane	Epidermis Dermis	Same as above	Refrigerator
Oasis	Healthpoint, LTD San Antonio, TX	Xenograft	Extracellular wound matrix from small intestine submucosa	Bioactive Dermal like Matrix	Superficial burns Skin graft donor sites Chronic wounds	Room temperature storage Multiple sizes 3×3.5 cm 7×20 cm
Biobrane	Dow Hickam/Bertek Pharmaceuticals	Synthetic with added denatured bovine collagen	Bilayer product outer silicone Inner nylon mesh with added collagen	Synthetic epidermis and dermis	Superficial partial thickness burns, Temporary cover of excised burns	Room temperature storage 15×20 in. 10×15 cm 5×15 in. 5×5 in.
Transcyte	Smith and Nephew Wound Management Largo, FL	Allogenic dermis	Bilayer product Outer silicone Inner nylon seeded with neonatal fibroblasts	Bioactive dermal matrix components on synthetic dermis and epidermis	Superficial to mid-partial thickness burns Temporary coverage of excised burns	Frozen in 5×7.5 in. sheets

^aThe names and properties of available temporary skin substitutes, with some biological activity are described. Also listed are the indications of the various products.

Table 7. Advantages and Disadvantages of Allograft Skin as a Skin Substitute

Allograft Skin

Advantages
 A bilayer skin providing epidermal and dermal properties
 Revascularizes maintaining viability for weeks
 Dermis incorporates into the wound

Disadvantages
 Epidermis will reject
 Difficult to obtain and store
 Risk of disease transfer
 Expensive Need to cryopreserve

nonviable state after preservation in glycerol or after lyophilization: however, most existing data describe best results when it is used in a viable state. The epidermal component provides a barrier until rejected by the host in 3–4 weeks. The dermis revascularizes and incorporates.

Homograft, another term for human allograft, can only be obtained from a tissue bank as strict protocols are required for harvesting and storage. Donors must be rigidly screened for potential viral and bacterial disease to avoid any transmission of disease. The product is in limited supply and very expensive.

The primary indication for use is to cover a large excised burn wound until an autogenous skin or a permanent skin substitute becomes available. Allograft is also used to cover a wide meshed skin graft, sealing the interstices during the healing process (Table 7).

Xenografts

Although various animal skins have been used for many years to provide temporary coverage of wounds, only porcine xenograft is widely used today (33,34) (Table 8). The epidermis of the porcine xenograft is removed and the split thickness dermis is provided in rolls. Split-thickness porcine dermis can be used after cryopreservation, or after glycerol preservation. It effectively provides temporary coverage of clean wounds such as superficial second degree burns and donor sites (33,34). Porcine xenograft does not vascularize, but it will adhere to a clean superficial wound and can provide excellent pain control while the underlying

Table 8. Advantages and Disadvantages of Xenograft as a Skin Substitute

Xenografts

Advantages
 Good adherence
 Decreases pain
 More readily available compared to allograft
 Bioactive (collagen) inner surface with fresh product
 Less expensive than allograft

Disadvantages
 Does not revascularize and will slough
 Short term use
 Need to keep the fresh product frozen

Table 9. Advantages and Disadvantages of Human Amnion as a Skin Substitute

Human Amnionic Membrane

Advantages
 Acts like biologic barrier of skin
 Decreases pain
 Easy to apply, remove
 Transparent

Disadvantages
 Difficult to obtain, prepare and store
 Need to change every 2 days
 Disintegrates easily
 Risk of disease transfer

wound heals. In general, xenograft is not as effective as homograft but is more readily available and less expensive. Primary indications are for coverage of partial thickness burns during healing and used burn wounds prior to skin grafting.

Human Amnion

Human amniotic membrane is used in many parts of the world as a temporary dressing for clean superficial wounds such as partial-thickness burns, donor sites, and freshly excised burns (35,36). Amniotic membrane is generally procured fresh and used after brief refrigerated storage. It can also be used in a nonviable state after preservation with glycerol. Amnion does not vascularize but still can provide effective temporary wound closure. The principal concern with amnion is the difficulty in screening the material for viral diseases. The risks of disease transmission must be balanced against the clinical need and the known characteristics of the donor (Table 9). The primary indications are the superficial burn and the excised wound.

Oasis Wound Matrix

This product is made of the submucosa of the porcine small intestine found between the mucosa and muscularis, in the wall of the porcine small intestine (37,38). The freeze dried acellular natural matrix retains its natural collagen and matrix structure and contains most of the bioactive matrix proteins found in the human dermis (Table 10).

Table 10. Advantages and Disadvantages of OASIS as a Skin Substitute

Oasis Wound Matrix

Advantages
 Excellent adherence
 Decreased pain
 Provides bioactive dermal like properties
 Long shelf life, store at room temperature
 Relatively inexpensive

Disadvantages
 Mainly a dermal analog
 Incorporates and may need to be reapplied

The submucosal layer is ~0.2 mm in thickness, but is quite durable. The product is freeze-dried removing the cells. The product is sterile, porous, biocompatible and nonimmunogenic. It has a long shelf life and can be stored at room temperature. The OASIS is incorporated into the wound bed over ~7 days and needs to be reapplied if the wound has not yet healed. The outer-barrier function is diminished with incorporation.

The primary indication is for use in difficult to heal nonburn wounds. It's use in burns is for the partial thickness burn and the skin graft donor site.

Biobrane

This product is a two-layer membrane (39,40). The outer epidermal analog is constructed of a thin silicone film with barrier functions comparable to skin. Small pores present in silicone allow for exudates removal and has permeability to topical antibiotics.

The inner dermal analogue is composed of a three-dimensional (3D) irregular nylon filament weave upon which is bonded type I collagen peptides. The surface binding of inner membrane is potentiated by collagen-fibrin bonds as well as fibrin deposition between the nylon weave. A thin water layer is maintained at the wound surface for epidermal cell migration maintaining moist wound healing. Excellent adherence to the wound significantly decreases pain in the superficial partial thickness burns. The silicone and nylon weave provides flexibility. The biobrane is removed once the partial thickness wound has re-epithelialized or the covered excised burn wound is ready for grafting. However, if left in place for >2 weeks the product is difficult to remove as tissue grows into the inner layer. Biobrane L contains a nylon fabric woven from monofilament threads that provide a less dense matrix and less adherence, preferred (e.g., on a donor site). There is likely very little direct bioactivity from the collagen peptides (40). The product has a long shelf life and can be stored at room temperature. It is also relatively inexpensive (Table 11).

The primary indication is for closure of the clean superficial burn or the excised burn wound.

Transcyte

This product is also a bilayer skin substitute (41,42). The outer epidermal analogue is a thin nonporous silicone film

Table 11. Advantages and Disadvantages of the Use of BIOBRANE as a Skin Substitute

Biobrane

Advantages

- Bilayer analog
- Excellent adherence to a superficial burn
- Decreases pain
- Maintains flexibility
- Easy to store with long shelf life
- Relatively inexpensive

Disadvantages

- Has very little direct bioactivity
- Difficult to remove if left in place >2 weeks

Table 12. Advantages and Disadvantages of TransCyte as a Skin Substitute

Transcyte

Advantages

- Bilayer analog
- Excellent adherence to a superficial to middermal burn
- Decreases pain
- Provides bioactive dermal components
- Maintains flexibility
- Good outer-barrier function

Disadvantages

- Need to store frozen till use
- Relatively expensive

with barrier functions comparable to skin. The inner dermal analog is layered with human neonatal foreskin fibroblasts that produce products, mainly collagen type I, fibronectin, and glycosaminoglycans.

A subsequent cryopreservation destroys the fibroblasts, but preserves the activity of fibroblast derived products on the inner surface. These products are then anticipated to stimulate the wound healing process (Table 12). A thin water layer is maintained at the wound surface for epidermal cell migration.

The nylon mesh provides flexibility and excellent adherence properties significantly decrease pain in the partial thickness burn. The product is peeled off after the wound has re-epithelialized.

The Transcyte must be stored at -70°C in order to preserve the bioactivity of the dermal matrix products. The primary indication is for closure of the clean superficial to mid-dermal burn, especially useful in children. Transcyte is also indicated for the temporary closure of the excised wound prior to grafting. Tissue ingrowth tends to be less of a problem even if the product is kept in place for >2 weeks.

Permanent Skin Substitutes

The purpose of these products is to replace full thickness skin loss as well as to improve the quality of the skin, which has been replaced after a severe burn (20–23).

As opposed to the bilayer concept of the ideal temporary skin substitute, permanent skin replacement is much more complex.

This area can be arbitrarily divided into two approaches (21–23). The first approach is the use of a bilayer skin substitute, with the inner layer being incorporated into the wound as a neodermis, rather than removed like a temporary product. The outer layer is either a synthetic to be replaced by autograft (epidermis) or actual human epithelial cells. If the outer layer is composed of epithelial cells that will form epidermis barrier function is not sufficiently developed at placement to act immediately as an epidermal barrier.

The second approach is the provision of either just an epidermal or a dermal analog or simply a coculture of cells containing elements of both. These products are technically not permanent skin substitutes (Table 13) upon initial placement as there is no bilayer structure until the product

Table 13. Components of Permanent Skin Substitutes

<i>Permanent Skin Replacement</i>
Bilayer structures with biologic dermal analog and either synthetic or biologic epidermal analog
Skin components
Epidermal cells alone
Dermis alone
Coculture of epidermal cells and fibroblasts

Table 14. Ideal Properties of a Permanent Skin Substitute^a

Rapid and excellent adherence properties
Easily applied and secured to an excised wound
Minimum wait period from time of burn to availability of skin substitute
Bilayer tissue containing both epidermal and dermal eliminates to best replicate normal skin
Rapid incorporation
Cannot transmit disease
Good functional and cosmetic result
Inexpensive

^aAs yet, the ideal skin does not exist.

evolves once placed on the wound. Both approaches will be discussed as will the ideal properties (Table 14).

The ideal property would be that of a bilayer structure.

The currently available clinical products are listed below (Table 15). There are a number of permanent skin substitutes in the development stage, which will not be listed.

Table 16. Advantages and Disadvantages of the Use of Apligraf As a Skin Substitute

<i>Apligraf</i>
Advantages
Bilayer skin containing human neonatal cells
Not requiring patients skin biopsy
No lag time for production
Contains epidermal and dermal functions (eventually)
Does not need to be frozen
Disadvantages
Made in small pieces and not practical for large burn
Cannot be stored for over 24 h
Relatively expensive

Apligraf

The dermal analog is made of fibroblasts from neonatal foreskin populated in a bovine type I collagen matrix. This layer incorporates into the excised full thickness wound adding a dermal component. Epithelial cells (keratinocytes) are also obtained from neonatal foreskin. No Langerhans cells at present so rejection does not occur. The epithelial cells divide, migrate and form an epidermis, which will eventually provide biologic barrier function. Donor foreskin is screened for viruses, which could be transmitted. The product is indicated mainly for chronic wounds. It's use in burns is currently off label (44,45). Apligraf is typically provided as a 7.5 cm diameter disk that is 0.75 mm thick. The advantages are that the product is already made and does not depend on obtaining the patients own cells (Table 16). However, the product must be shipped the night before use in a polyethylene bag in agar nutrient medium and a 10% CO₂ content and stored at room temperature until used (within 24 h).

Table 15. Properties and Uses of the Currently Available Permanent Skin Substitutes

<i>Available Permanent Skin Substitutes</i>						
Product	Company	Tissue of Origin	Layers	Category	Uses	How supplied
Apligraf	Organogenesis, Inc and Novartis Pharmaceuticals Corp	Allogenic Composite	Collagen matrix seeded with human neonatal keratinocytes and fibroblasts	Composite: Epidermis and dermis	Chronic wounds, often used with thin STSG Excised deep burn	7.5 cm diameter disk 1/pack
OrCel	Ortec International Inc.	Allogenic Composite	Collagen sponge seeded with human neonatal keratinocytes and fibroblasts	Composite: Epidermis and Dermis	Skin graft donor site, chronic wounds	6 × 6-cm sheets
Epicel	Genzyme Tissue Repair Corp	Autogenous keratinocytes	Cultured autologous keratinocytes	Epidermis Only	Deep partial and full thickness burns >30% TBSA	50 cm ² sheets in culture medium
Alloderm	Life Cell	Allogenic dermis	Acellular Dermis (processed allograft)	Dermis only	Deep partial and full thickness burns, Soft tissue replacement, Tissue patches	1 × 2–4 × 12 cm
Integra	Integra Life Science Corp	Synthetic	Silicone outer layer on collagen GAG dermal matrix	BioSynthetic Dermis	Full thickness soft tissue defects definitive "closure" requires skin graft	2×2 in. 4×10 in. 8×10 in. 5/pack

Table 17. Advantages and Disadvantages of Orcel as a Skin Substitute

<i>Orcel</i>
Advantages
Eventual bilayer skin
Not requiring patient's cells
Does not need to be frozen
Disadvantages
Bilayer structure requires 14 days to develop once applied
Not indicated for excised burns
Short storage time
Relatively expensive

Orcel

The product is produced as a coculture of keratinocytes and fibroblasts (neonatal foreskin) in a cross-linked bovine collagen sponge (46). The donor tissue is screened for viruses. The nonporous side of the sponge is seeded with the keratinocytes and the porous side with fibroblasts. After application, a neodermis forms. Once applied the epidermis and an epidermal barrier requires 14 days to develop as the keratinocytes migrate and divide on the surface (Table 17).

The product is indicated mainly for chronic wounds but is also indicated for coverage of split thickness skin graft (STSG) donor sites. This product is not currently indicated for use on excised burn. The purpose for the STSG is to provide a better functional and cosmetic outcome compared to a healed donor site. It is shipped in a package filled with its culture medium which is stored at room temperature until used, usually within a few days.

Epicel

This product, used mainly for very large burns is composed of the patients skin epithelial cells and referred to as cultured epithelial autograft (CEA). Therefore, only the epithelial layer is provided (47–49). The product is made from a small biopsy of normal skin (2×2 cm) from the burn patient. The epithelial cells are extracted and cultured. Use of a cell culture technique allows the keratinocytes to be grown in a thin sheet 10,000 times larger than the initial

Table 18. Advantages and Disadvantages of Epicel as a Skin Substitute

<i>Epicel</i>
Advantages
Patients own keratinocytes expanded several thousand fold
Small skin biopsy required
Can cover very large surfaces with reasonable graft take
Used in large burns
Disadvantages
2–3 week lag time for production
Provides only the epidermal layer
Epithelial layer can be quite fragile for some time
Needs to be used immediately on delivery
Very expensive

biopsy. This process does require 2–3 weeks from the time of biopsy. Often the burn wound is excised and covered with homograft (allograft) until the cells are ready to be transplanted. The CEA is then applied to the clean excised (or allograft covered) wound.

The CEA is supplied in sheets 2–6 cells thick on small pieces of petroleum gauze (50 cm²), which are bathed in culture medium. Immediate application is necessary. The CEA grafts are very fragile and easily rubbed off for at least several weeks. The backing is removed in several weeks as the CEA thickens and adheres. Graft take ranges from 30 to 75% of total epithelium applied. The epithelium gradually thickens but has a low resistance to shear forces for some time. Application of allograft dermis, prior to CEA grafting appears to improve skin quality.

The primary indication is for very large burns.

Alloderm

This product is basically treated human allograft with the epidermis removed (50–52). The dermis is treated to produce a copreserved lyophilized allodermis, which incorporates. The product is used as a dermal implant. Therefore application of a thin epithelial autograft is required.

Primary indication is for use in the replacement of soft tissue defects. This product is not commonly used in large burns. Typically the alloderm is applied to an excised wound and then a split thickness skin graft is placed on top of the alloderm (Table 19). The product has a long shelf life in its lyophilized form. It requires rehydration prior to use.

Integra

This product is composed of a dermal analog made of a biodegradable bovine collagen-glycosaminoglycan copolymer matrix. The collagen and glycosaminoglycan is cross-linked to attempt to maximize ingrowth of the patients own cells (53–55) (Table 20).

The epidermal analog is a thin silicone elastomer providing temporary barrier protection. After the dermal analog incorporates and the surface revascularizes, at ~2–3 weeks, the silicone layer is removed and replaced with a very thin skin graft from the patient (or CEA cells). The Integra needs to be carefully immobilized for the first 2 weeks as movement will cause devascularization and loss of the product. The primary indication is the treatment of large deep burns as well as reconstruction procedures. The

Table 19. Advantages and Disadvantages of Alloderm as a Skin Substitute

<i>ALLODERM</i>
Advantages
Easy to store, an off the shelf product
Does not require skin bank
Comes in large and small pieces
Disadvantages
Requires thin skin graft to provide epidermis
Two procedures required to achieve bilayer skin
Relatively expensive

Table 20. Advantages and Disadvantages of Integra as a Skin Substitute**Integra****Advantages**

- Provides thick dermal analog
- Reasonable shelf life
- No risk of transmitting viruses
- Relatively inexpensive
- Used in large burns

Disadvantages

- Need to provide epidermis from the patient
- Dermal cells must come from the patient requiring product incorporation
- Two procedures required to achieve bilayer skin

incorporated neodermis appears to improve the function of the final skin once the epithelial graft is applied. The product is provided in a number of sizes and sheets stored in 70% isopropyl alcohol. Shelf life is very good.

SUMMARY

The scientific principles and practical approaches, to replacing skin either temporarily or permanently are advancing at a rapid rate. Much of these advances can be attributed to both advances in the field of bioengineering as well as increasing interest in optimizing the outcome of the burned skin.

The ideal properties of a bioactive temporary and a permanent skin substitute have been well defined. As expected, the properties of temporary skin substitutes are more concrete, easier to categorize and determine efficacy. A bilayer structures is the current standard with the dermal component being bioactive. Permanent skin replacement on the other hand is much more complex. A variety of approaches are being used which can be loosely categorized as either use of bilayer products (usually the outer layer to be replaced by epidermal autograft) or replacement of either dermal or epidermal elements separately. The terminology of the latter approach is difficult because these component products are really not permanent skin substitutes on initial application but become so only when all the elements are in place.

An understanding of the properties of each product is essential for the user to optimize outcome.

BIBLIOGRAPHY

1. Mast B. The skin. In: Cohen K, Diegelmann I, editors. *Wound Healing*. Philadelphia: WB Saunders; 1992. p 344–355.
2. Wright N, Allison M. The Biology of Epithelial Cell Populations. Clarendon Press; 1984. p 283–345.
3. Stenn S, Malhotra R. Epithelialization. In: Cohen C, editor. *Wound Healing. Biochemical and Clinical Aspects*. Philadelphia: WB Saunders; 1992. p 115–127.
4. Grillo H. Origin of fibroblasts in wound healing. *Ann Surg* 1963;157:453–467.
5. Karasck M. Mechanism of angiogenesis in normal and diseased skin.
6. Raghov R. The role of extracellular matrix in post-inflammatory wound healing and fibrosis. *FASEB J* 1994;8:823–850.
7. Demling R. Burn care. In: Wilmore D, editor. *ACS Surgery*. New York: Web MD; 2002. p 479.
8. Herndon D. *Total Burn Care*. Philadelphia: WB Saunders; 2002.
9. Neely A, Brown R, Chendening C, et al. Proteolytic activity in human burn wounds. *Wound Rep Regen* 1997;5:302–309.
10. Muller M, Pegg S, Rule R. Determinants of death following burn surgery. *Br J Surg* 2001;88:583–587.
11. Komgova R. Burn wound coverage and burn wound closure. *Acta Chir Plast* 2000;42:64–68.
12. Sheriden R. Management of burn wounds with prompt excision and immediate closure. *J Inten Care Med* 1994;9:6–17.
13. Demling R, DeSanti L. Scar management strategies in wound care. *Rehab Manage* 2001;14:26–32.
14. Spres M, Herndon D, et al. Prediction of mortality from catastrophic burns in children. *Lancet* 2003;361:989–994.
15. Ladin D, Garner W, Smith D. Excessive scarring as a consequence of healing. *Wound Repair Reg* 1994;3:6–14.
16. Scott P, Ghabary A, Chambers M, et al. Biologic basis of hypertrophic scarring. *Adv Struct Biol* 1994;3:157–165.
17. Erlich HP, Krummell T. Regulation of wound healing from a connective tissue perspective. *Wound Repair Reg* 1995;4: 203–210.
18. Badylak S. The extracellular matrix as a scaffold for tissue reconstruction. *Cell Develop Biol* 2002;13:377–383.
19. Jones L, Currie L, Martin R. A guide to biological skin substitutes. *Br J Plast Surg* 2002;55(3): 185–193.
20. Gallico GG. Biologic skin substitutes. *Clin Plast Surg* 1990; 512–520.
21. Sheridan R, Tompkins R. Alternative wound coverings. In: Herndon D, editor. *Total Burn Care*. Philadelphia: WB Saunders; 2003. p 712.
22. Clark R, Folkvard J, Wortz R. Fibronectins, as well as other extracellular matrix proteins mediate human keratinocytes adherence. *J Invest Dermatol* 1985;84:378–383.
23. Clore J, Cohan I, Diegelmann R. Quantitation of collagen types I and III during wound healing. *Proc Soc Exper Biol Med* 1979;161:337–340.
24. Doillon C, Dunn M, Bender E, et al. Collagen fiber formation in repair tissue: Development of strength and toughness. *Collagen Relat Res* 1985; 481–485.
25. Takashima A, Grinnell F. Human keratinocytes adhesion and phagocytosis. Prompted by fibronectin. *J Invest Derm* 1984;83:352–358.
26. Miller E, Gay S. Collagen structures and function in wound healing: biochemical and clinical aspects. In: Cohen K, editor. Philadelphia: Saunders; 1992. p 130.
27. Nowicki CR, Sprenger C. Temporary skin substitutes for burn patients: a nursing perspective. *J Burn Care Rehab* 1988;9(2):209–215.
28. Shakespeare P. Survey: use of skin substitute materials in UK burn treatment centers. *Burns* 2002;28(4):295–297.
29. Smith K, Rennie MJ. Management of burn injuries: A rationale for the use of temporary synthetic substitutes? *Prof Nurse* 1991;5:71–574.
30. Bondoc CC, Burke JF. Clinical experience with viable frozen human skin and a frozen skin bank. *Ann Surg* 1971;174:371–382.
31. Herndon DN. Perspectives in the use of allograft. *J Burn Care Rehab* 1997;18:56.
32. May SR, Still JM Jr., Atkinson WB. Recent developments in skin banking and the clinical uses of cryopreserved skin. (Review) *J Med Assoc GA* 1957;73:233–236.

33. Song IC, Bromberg BE, Mohn MP, Koehnlein E. Heterografts as biological dressings for large skin wounds. *Surgery* 1966;59:576–583.
34. Elliott RA Jr., Hoehn JG. Use of commercial porcine skin for wound dressings. *Plast Reconstr Surg* 1973;52:401–405.
35. Ramakrishnan KM, Jayaraman V. Management of partial thickness burn wounds by amniotic membrane: A cost-effective treatment in developing countries. *Burns* 1997;23 (Suppl. 1): 533–536.
36. Ganatra MA, Durrani KM. Method of obtaining and preparation of fresh human amniotic membrane for clinical use. *J Pakistan Med Assoc* 1996;46:126–128.
37. Brown-Estris M, Cutshall W, Hiles M. A new biomaterial derived from small intestinal submucosa and developed into a wound matrix device. *Wounds* 2002;14:150–166.
38. Demling R, Niezgodka J, Haraway G, Mostow E. Small intestinal submucosa wound matrix and full thickness venous ulcers. *Wounds* 2004;16:18–23.
39. Smith DJ Jr. Use of biobrane in wound management. *J Burn Care Rehab* 1995;16:317–320.
40. Yang J, Tsai Y. Clinical comparison of commercially available Biobrane preparations. *Burns* 1989;15:197–203.
41. Purdue G, Hunt J, Still M, et al. A multicenter clinical trial of a biosynthetic skin replacement, Dermagraft-T compared with cryopreserved human cadaver skin for temporary coverage of excised burn wounds. *J Burn Care Rehab* 1997;18:52–57.
42. Demling RH, DeSanti L. Management of partial thickness facial burns (comparison of topical antibiotics and bio-engineered skin substitutes). *Burns* 1999;25:256–261.
43. Bell YM, Falabella AF, Eaglstein WH. Tissue engineered skin. Current status in wound healing. *Am J Clin Dermatol* 2001;2:305–313.
44. Folangi V, Sabolinski M. A bilayered living skin construct (APLIGFAP) accelerates complete closure of hard-to-heal venous ulcers. *Wound Repair Regen* 2000;7:201–207.
45. Fivenson DP, Scherschun L, Choucair M, Kukuruga D, Young J, Shwayder T. Graftskin therapy in epidermolysis bullosa. *J Am Acad Dermatol* 2003;48:886–892.
46. Still J, Glat P, Silverstein P, Griswold J, Mazingo D. The use of a collagen sponge/living cell composite material to treat donor site burn patients. *Burns* 2003;29(8):837–841.
47. Sheridan RL, Tompkins RG. Cultured autologous epithelium in patients with burns of ninety percent or more of the body surface. *J Trauma* 1995;38:48–50.
48. Rue LW III, Cioffi WG, McManus WF, Pruitt BA Jr. Wound closure and outcome in extensively burned patients treated with cultured autologous keratinocytes. *J Trauma* 1993; 34:662–667.
49. Loss M, Wedler V, Kunzi W, Meuli-Simmen C, Meyer VE. Artificial skin, split-thickness autograft and cultured autologous keratinocytes combined to treat a severe burn injury of 983% of TBSA. *Burns* 2000;26:644–652.
50. Buinewicz B, Rosen B. Acellular cadaveric dermis (Allo-Derm): a new alternative for abdominal hernia repair. *Ann Plast Surg* 2004;52:188–194.
51. Wax MK, Winslow CP, Andersen PE. Use of allogenic dermis for radial forearm free flap donor site coverage. *J Otolaryngol* 2002;31:341–345.
52. Druecke D, Steinstraesser L, Homann HH, Steinau HU, Vogt PM. Current indications for glycerol-preserved allografts in the treatment of burn injuries. *Burns* 2002;28 (Suppl. 1): S26–30.
53. Wisser D, Rennekampff HO, Schaller HE. Skin assessment of burn wounds covered with a collagen based dermal substance in a 2-year follow-up. *Burns* 2004;30:399–401.
54. Navsaria HA, Ojeh NO, Moiemem N, Griffiths MA, Frame JD. Re-epithelialization of a full-thickness burn from stem cells of hair follicles micrografted into a tissue-engineered dermal template (Integra). *Plast Reconstr Surg* 2004;113:978–981.
55. Frame JD, Still J, Lakhel-LeCoadou A, Carstens MH, Lorenz C, Orlet H, Spence R, Berger AC, Dantzer E, Burd A. Use of dermal regeneration template in contracture release procedures: A multicenter evaluation. *Plast Reconstr Surg* 2004;113:1330–1338.

See also **BIOCOMPATIBILITY OF MATERIALS; ENGINEERED TISSUE; SKIN, BIOMECHANICS OF; SKIN TISSUE ENGINEERING FOR REGENERATION.**

SKIN TISSUE ENGINEERING FOR REGENERATION

BRENDAN A. HARLEY
IOANNIS V. YANNAS
Massachusetts Institute of
Technology
Cambridge, Massachusetts

MAMMALIAN RESPONSE TO INJURY

Scale of Functional Deficit

Medical treatment options for the loss of normal organ or tissue function depend heavily on the scale of the defect, either macromolecular- or organ-scale. Since antiquity, macromolecular-scale defects have been treated with chemical therapeutics such as herbs and potions. More recently, pharmaceuticals, vitamins, hormones, and antibiotics have been used to treat a vast array of medical maladies that are caused by a macromolecular defect; these treatment regimens have been used successfully to replace or correct a missing function on the molecular scale. Organ-scale defects present a significantly larger wound site and require considerably different treatment practices.

Large-scale failures of a tissue or organ are created primarily by disease or by an acute or chronic insult; injury or damage of this type typically results in wound sites on the scale of a millimeter or centimeter, as opposed to the nanometer scale of a molecular defect. This type of damage cannot be treated with drugs because the problem is the failure of a mass of tissue, including a large number of cells, soluble proteins and cytokines, and insoluble extracellular matrix. The typical organismic response to an injury at the organ-scale is cell-mediated wound contraction and synthesis of nonphysiologic tissue (scar); this process is termed repair. Regeneration of lost or damaged tissue describes a process marked by synthesis of physiologic (normal, functional) tissue in the wound site. There are a few notable exceptions to the rule that the mammalian response to organ-scale injuries is repair. The epithelial tissues of the skin, genitals, cornea, and gastrointestinal tract all regenerate spontaneously, and the liver has shown the ability to synthesize a substantial organ mass without recovery of the original organ structure, but also without contraction or scar synthesis (1). Despite these few notable exceptions, the mammalian response to the loss or damage of a tissue or organ is almost exclusively repair, an irreversible

response resulting in the formation of scar tissue that lacks the structure and function of the damaged organ or tissue.

Regeneration Versus Repair

As previously noted, there are two possible outcomes of the mammalian healing process following acute or chronic injuries: regeneration or repair. Regeneration is characterized by synthesis of a replacement tissue in the anatomical wound site that is structurally and functionally similar to the original tissue. Repair is characterized by synthesis of scar tissue (nonphysiological tissue) to replace the normal tissue lost due to injury. In addition to synthesis of new, nonphysiologic tissue, contraction of the wound site is also observed during repair. Cell-mediated contraction of the wound site has been observed in many different species to varying degrees at different wound sites (1).

Wound closure following severe injury occurs by one or more of three distinct processes. The initial wound area can close by contraction (C), scar formation (S), and/or regeneration (R). The process of wound closure can be represented quantitatively by the defect closure rule: $A_C + A_S + A_R = 100$ ($A_X \equiv$ percentage wound area closed by process X); these three processes are the only mechanisms by which wound closure takes place. In adult mammals, chronic and acute injuries have a common clinical outcome because the repair processes responsible for wound healing close the wound through contraction and scar formation only ($A_R = 0$). With the exception of a certain class of injuries, such as injuries to epithelial tissues in select organs, the ability to regenerate tissues and organs is lost in mammalian adults. Spontaneous regeneration (regeneration without external stimulation) is observed in very specific tissues in adult mammals following minor injuries such as a small skin scrape or a first or second degree burn, while more severe injuries such as a deep skin wound or third degree burn result in contraction and scar formation. A review of the available data comparing cases of regeneration with those of repair has led to the proposal of antagonistic relationship between contraction and regeneration; in cases of adult healing where contraction occurs, regeneration is not observed, and regeneration is observed when wound contraction is blocked (1–3).

Mammalian Response to Injury

While the mammalian adult responds to severe injury by repair processes resulting in the formation of scar ($A_C + A_S = 100$; $A_R = 0$), the mammalian fetus is able to regenerate the lost tissue spontaneously ($A_C, A_S \ll A_R$) (4,5). For the sake of this article, the fetal classification refers to all mammalian fetuses that respond to injury with regeneration processes while the adult classification refers to all mammals (adult as well as juvenile) that respond to injury with repair processes. Modifying the adult mammalian wound healing response to more closely resemble that observed with the fetus has been an area of extensive study. In addition to understanding the differences between the fetal and adult healing processes, analogs of the extracellular matrix (ECM) have been used as templates for a variety of tissue engineering related disciplines

toward the goal of inducing regeneration of tissues damaged by severe injury where the organism would normally respond to injury with repair processes.

The extensive study of the mammalian response to injury has focused on understanding the mechanism and timing of the transition from regeneration to repair processes during the fetal–adult transition of development and whether regeneration can be induced in the adult mammals once the initial transition has taken place. For mammals, the transition from wound closure by regeneration to wound closure by scar synthesis and contraction takes place during the third trimester of gestation (1,4,5). While certain classes of amphibians have been studied extensively throughout metamorphosis from tadpole to young adult (6), there is for the most part much less information available about higher mammals. A detailed study of wound closure mechanisms of the amphibian frog during larval development was based on measurement of the fractional extent of wound contraction, scar formation, and regeneration during development. With increasing fetal development, wound closure depended increasingly less on regeneration and correspondingly more on contraction. Once metamorphosis to the young adult frog was complete, regeneration was not observed while contraction and synthesis became the predominant mechanisms for wound closure (7). These and several other observations support the conclusion that there is an antagonistic relationship between regeneration and wound contraction. To date, while the causes for the transformation in the mode of mammalian response to injury from the fetus to the adult are not known, adult mammals are known to be unable to regenerate tissue lost due to severe injury and close severe wounds by contraction and scar formation.

There are three tissue layers that are grouped together in sequence in all organs, namely, epithelia, the basement membrane, and the stroma (1,8). This sequence has been termed the tissue triad. A pattern has been observed as a result of examinations of the structure and tissue-specific response to injury of the three tissues that to date have been induced to regenerate: skin, peripheral nerves, and conjunctiva. Understanding of the response of each member of the tissue triad to injury aids in shedding light on the process of regeneration and has suggested methodology for inducing regeneration in tissues that do not regenerate spontaneously. A layer of epithelial cells (epithelia) covers all surfaces, tubes, and cavities of the body; this layer is totally cellular with no ECM component. The epithelia is completely cellular, avascular, and is the only member of the tissue triad that does not contain an extracellular matrix. Developmental and functional similarities between epithelial cells and tissues from a variety of different tissues and organs such as the skin and peripheral nerves have been observed; these observations have suggested that the tissue triad can be used as a more general tool to understand organismic response to injury (1).

The basement membrane (also termed basal lamina) is a continuous layer of tissue separating the epithelial layer from the stroma. In all tissues, the basement membrane is acellular, and no blood vessels pass through the basement membrane layer. The stroma contains connective tissues as well as the blood supply, and provides a reservoir for

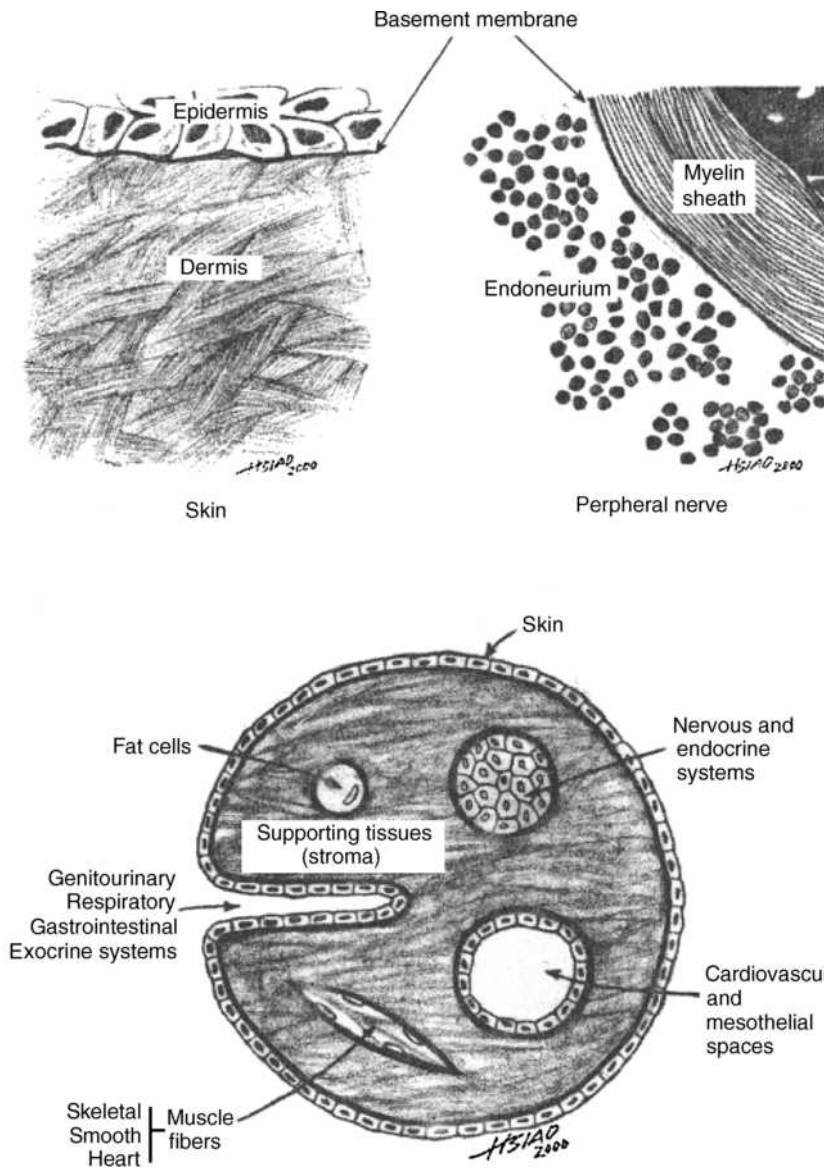


Figure 1. Schematics of the tissue triad structure observed in mammalian tissue. Top: Tissue triad of skin and peripheral nerves. The basement membrane is a thin extracellular matrix tissue located between the cellular and nonvascular epithelia (epidermis, myelin sheath) and the cellular, vascularized stroma (dermis, endoneurium). Bottom: Diagram of the distribution of epithelial, basement membrane, and stromal tissues in the mammalian system. Examples of stromal tissues are bone, cartilage, and their associated cell types as well as elastin and collagen. Examples of epithelial tissues are those covering the surface of the genitourinary, respiratory, and gastrointestinal tracts as well as surfaces of the mesothelial cells in body cavities, muscle fibers, fat cells, and endothelial cells in the cardiovascular system (1).

nutrient uptake to and waste removal from the basement membrane and epithelia. Figure 1 provides a basic diagram of the organization of the tissue triad in the adult mammalian system.

Using the tissue triad as a guide, we can identify similarities among the three tissues. In skin, peripheral nerves, and the conjunctiva, there are tissues that spontaneously regenerate (the epithelia and basement membrane) and a tissue that does not (the stroma). The regenerative capacity of epithelial tissue and the basement membrane as well as the irreversible (repair) nature of stromal wound healing have been extensively reported (1,9–13). Specifically, the stroma has been repeatedly observed as nonregenerative in skin, peripheral nerves, blood vessels, lung, kidney, and pancreas (1,3,14–18). It has been suggested that the mechanism for the irreversibility of injury (nonregenerative response) is entirely dependent on disruption of the stromal architecture, and

that proper replacement of the stromal layer is critical for any regeneration to occur.

Methods to Treat Loss of Organ Function

While a molecular scale defect can often be treated with the use of pharmaceuticals, an organ-scale defect requires more extensive treatments. Significant loss of function in the affected tissue or organ, also referred to as the “missing organ” (19), can lead to a number of significant consequences such as lack of social acceptance in the case of severe burns and facial scars, loss of mobility and sensory function in the case of neuroma, or life-threatening symptoms in the case of a cirrhotic liver, large-scale severe burns, and ischemic heart muscle.

Six basic approaches have been used to treat the problem of the missing organ: transplantation, autografting, implantation of a permanent prosthetic device, use of stem

cells, *in vitro* synthesis of organs, and induced regeneration. The last three techniques have been grouped together and are known by the moniker “tissue engineering” (20). All six techniques will be discussed in the following sections.

Transplantation. Transplantation, the transfer of an organ or a fraction thereof from a donor to a host, is widely utilized as a therapeutic strategy to replace complex tissues and organs. Since the introduction of transplantation in the early twentieth century (21), it has been used in increasingly complex organ systems such as the skin, cornea, kidney, liver, lungs, and heart. Patients have exhibited extraordinary survival characteristics even with the simultaneous transplantation of multiple organs (22–27).

While transplantation of tissues at a few immune-privileged sites such as the eye and testis can occur without rejection, the most significant challenge facing modern transplantation is the immunological barrier for transplantation of tissues from donor to host, where the donor organ is attacked and rejected by the host’s immune system upon transplantation; this response is termed host-versus-graft disease where the host’s (patient’s) immune system recognizes the antigens expressed on the graft tissue as foreign and attacks and destroys the tissue. The primary clinical method for preventing rejection of the transplanted tissue is the use of drugs to suppress the immune system of the host. Immunosuppression therapy often is necessary for the remainder of the host’s life to prevent transplant rejection; however, immunosuppression also presents a significant hazard to the host, as the immunosuppressed host becomes vulnerable to infections (28). Significant efforts have been mounted to develop technologies that allow for local rather than whole-body immunosuppression, such as the development of cells that can express a protein (i.e., Fas ligand) to induce immune cell apoptosis and the use of natural and synthetic polymers to encapsulate heterologous cells to mask their antigens (24,28–31).

The major obstacle in using human donors for transplantation has been the difficulty in finding immunocompatible donors and the shortness of supply of suitable organs because the supply is greatly exceeded by the demand (32). A more recent area of research has focused on developing a transgenic pig model to be used as an immunocompatible donor for humans for a procedure known as xenotransplantation. Development of this area of transplantation has been slowed by evidence that pig viruses are capable of infecting human cells and producing unique viral infections in the host (33–35).

Autografting. With autografting, the donor and the recipient are the same individual; a fraction of the tissue or organ is harvested from an uninjured site and grafted at the nonfunctional site (25). Autografting offers a technology that removes the danger of organ or tissue rejection due to a host-versus-graft response, but is relatively limited in its scope of application. Autografting necessitates the creation of a second wound site (donor site), subjecting the patient to a second severe trauma. Therefore, autografting is utilized only when the loss of functionality at the sec-

ondary wound site is outweighed by the loss of functionality or morbidity at the primary wound site. This procedure is obviously limited by the availability of functional tissue that can be transplanted without additionally harming the patient. Major clinical uses of autografting have been associated with skin grafting in massively burned patients (36), the use of the sural nerve to bridge a severe peripheral nerve injury, primarily in the case of hand injuries (37–41), and the use of autologous vein graft to bypass a restricted artery (42).

Permanent Prosthetic Device. The implantation of a permanent prosthetic device to replace the functionality of lost or damaged tissues offers a number of advantages and disadvantages. Typical examples of prosthetic devices are artificial hips and knees (43), cardiac pacemakers (44), heart valves (45), stents (46), cochlear implants (47), and contact lenses (48). Prosthetic devices are typically fabricated from biologically inert materials such as metals, ceramics, and synthetic polymers. Hence, these devices do not provoke the immune response problems inherent to transplanted tissues and organs and can also be manufactured on a mass scale. Even though these devices are fabricated from bioinert materials, interactions with the biological environment surrounding the prosthesis lead to a number of unfavorable physical and biological manifestations. Specific examples of negative biomaterial–tissue interactions are the formation of a thick, fibrous scar tissue capsule around a silicone breast implant (49,50), stress-shielding due to the implantation of a relatively stiff (compared to the host bone) hip prosthesis that eventually leads to a loss of bone mass (51), platelet aggregation to implanted surfaces, also known as biofouling (52–55), and the accumulation of polyethylene particles in the lymph system as a result of wear of an orthopedic implant (56,57). The spontaneous remodeling process of the surrounding tissues can also be significantly altered, negatively or positively, by the presence of the prosthetic device (58). The often-serious side effects that appear as a result of interactions between nonbiological materials and the surrounding tissues illustrate the difficulty of replacing bioactive tissues with bioinert materials with drastically different material and mechanical properties.

Stem Cells. Stem cells present an exciting possibility for replacement of lost or damaged organs and tissues. The pluripotential nature of stem cells offers the possibility for the synthesis of tissues from the least differentiated cells in the body (59,60). Current efforts in stem cell research have focused on identifying protocols to harvest stem cells, expand them in culture, and reimplant them at a site of injury, as well as to develop technologies to introduce genes into stem cells so that when reintroduced to the patient, they will synthesize the required proteins *in vivo*. Currently, mesenchymal stem cells (61), epithelial stem cells (62), and neural stem cells (63) have been grown *in vitro* and studied. While few advances in the use of stem cells for replacement of damaged tissues have been made to this point, stem cell technologies present a new area of scientific study for future exploration with a great deal of promise.

In Vitro Synthesis. *In vitro* synthesis requires the growth of a functional replacement tissue using an *in vitro* culture environment to replace a lost or damaged organ or tissue. Traditional *in vitro* synthesis techniques have utilized both cell culture systems and culture systems based on interactions between cells and an extracellular matrix analogue. *In vitro* synthesis allows for total control over the culture environment, specifically the inclusion or exclusion of soluble regulators (i.e., growth factors, cytokines), insoluble regulators (i.e., extracellular matrix proteins), and a variety of cell culture media and conditions; the complexity of biological systems, specifically their cytokine and growth factor needs, and the necessity for developing an efficient method for providing nutrients and cell–cell signals to the site of the developing tissue have to date precluded the formation of complex tissues *in vitro* (1,64).

Early successes with *in vitro* synthesis were encountered using cultured epithelial cells to produce a physiological epidermis (65). In these studies, keratinocyte sheets were grown *in vitro* from skin explants; these keratinocyte sheets were implanted into skin wounds, and induced the formation of a fully mature, stratified epidermis, the uppermost tissue layer that makes up skin (66,67). Later study found that keratinocyte sheets could be grown *in vitro* starting from disaggregated epidermal cells and could then be implanted into a skin wound, inducing the development of a mature, fully stratified epidermis as well (68). This technique has been used to prepare autologous sheets of keratinocytes to treat skin wounds in severely burned patients (69–77). Known as a cultured epithelial autograft (CEA), this technology will be discussed in greater depth later in this article as one of the five major techniques that lead to at least partial regeneration following severe skin injuries. *In vitro* synthesis of more complex tissues and organs began with work aimed at developing an epithelial–mesenchymal bilayer in order to produce a material that could be implanted to replace damaged skin. The fabrication of a “living skin equivalent” (LSE) involved culture of fibroblasts within a collagen gel, followed by introduction of keratinocytes in order to produce an immature skin equivalent; this bilayer was implanted into a skin wound and was observed to lead to the formation of a mature, stratified epidermis as well as an immature neodermis (78–82). This technology will also be covered in greater detail later in this article.

In vitro synthesis of increasingly complex tissues has necessitated the development of technologies for culturing cells in three-dimensional (3D) scaffolds known as ECM analogs and modifying the surface chemistry of these scaffolds to control cell–substrate interactions (20,48,83–86). Synthetic polymer meshes have been used as an ECM analog for culturing keratinocytes and fibroblasts as a potential skin replacement that has been used clinically in the treatment of severe burns and ulcers (87–94). This technology (Living Dermal Replacement, LDR) will be described in greater detail later in the article as one of the major techniques utilized to treat severe skin wounds. *In vitro* synthesis using ECM analogs as a culture environment has also been studied using hepatocytes in an attempt to synthesize a functioning liver (95–97), and chondrocytes in an attempt to synthesize a cartilaginous

network (98–101). Continued exploration of *in vitro* techniques to synthesize tissue replacements has been hampered by the complexity of biological systems and an inability to provide the proper nutrient cocktails (i.e., cytokines, growth factors), the structures necessary to deliver these nutrients (i.e., arteries, veins, and capillary systems), and the correct mechanical environments (ECM analog structures) for complex tissue and organ growth outside of the organism.

Induced Organ Synthesis *In Vivo* (Regeneration). Induced organ synthesis *in vivo* relies on the healing processes that are inherently active in a wound site to regenerate lost or damaged tissue. In this method, an analog of the ECM is implanted in the wound site and combined with the biological processes in the wound induce synthesis of a physiologic replacement tissue. Induced organ synthesis was first identified following the development of fabrication techniques that allowed for synthesis of ECM analogs with a well-defined macromolecular structure, specifically controlling the ECM specific surface, chemical composition, and degradation rate (102–106).

The first application of induced organ synthesis was in the fabrication of an ECM analog able to induce skin regeneration, the dermal regeneration template (DRT). The DRT showed very high biological activity when implanted into a full-thickness skin wound and was capable of inducing regeneration of the underlying dermal layer of skin as well as the epidermal and basement membrane layers (3,103,107–113). The speed of this regeneration process was significantly increased when the DRT was further seeded with autologous keratinocytes prior to implantation (3,111,112,114–116). This technology will be covered in greater detail later in the article as one of the major tissue engineering techniques utilized to treat severe skin injuries.

Induction of organ regeneration, first observed in the study of skin regeneration, has also been observed in other tissues and organs with the use of other specialized ECM analogs. Regeneration of peripheral nerves has been achieved using a tubular device that incorporates an ECM analog as a filling, known as the nerve regeneration template (NRT) (1,117–120). The long-term morphological structure and electrophysiological function of nerves regenerated using the NRT has been observed to be at the level of an autografted nerve, the current gold-standard for peripheral nerve injury treatment (121,122). In addition, the DRT was implanted into a conjunctiva stromal wound model, where induced regeneration of the conjunctival tissue structure was also observed (123).

In vitro synthesis and induced organ regeneration (*in vivo* synthesis) currently constitute an area of study termed tissue engineering (20). For the remainder of this article, we will focus on studies on the structure and function of ECM analogs for use in tissue engineering and on an overview of the tissue engineering approaches that have been utilized to treat skin wounds.

Basic Parameters of the Living Environment During *In Vivo* Synthesis

The process of *in vivo* regeneration of lost or damaged tissue can be modeled as if the entire process was taking

place within a “bioreactor” that is surrounded by a reservoir with constant properties, representing the entire organism with its complex homeostatic mechanisms. The “bioreactor” itself has a defined anatomical and physicochemical environment (environment of the wound site); parameters include the temperature, the pH, the structure of a template within the “bioreactor”, and the flow rate and composition of the exudate. The flow of exudate entering the “bioreactor” starts almost immediately following the creation of the wound site, while the structure in the wound bed is provided by implantation of an analog of the ECM. The ECM analog, and any cells that may be seeded within, constitutes the exogenous stimulus provided to the wound bed. During the remainder of the healing process, the reservoir (surrounding organism) is considered to maintain the “bioreactor” environment: temperature, pH, as well as cytokines, growth factors, and cells present in the exudate. While these factors are considered standardized in a wound bed, it is the ECM analog that provides the variable structure and ECM components (proteins) that are critical for inducing regeneration. If no active ECM analog is present or if its structural features are changed, the reaction sequence within the “bioreactor” is observed to follow almost completely normal repair processes that result in wound contraction and scar formation. In contrast, when the appropriate ECM analog is introduced into the “bioreactor” normal repair mechanisms are replaced by induced regeneration, also referred to as *in vivo* or *in situ* regeneration (1).

Experimental study of *in vivo* regeneration is complicated by a lack of reproducibility between different reaction sites (anatomical sites); unless the wound site is standardized, it will be impossible to accurately study differences between ECM analogs when implanted into wound models and results obtained in independent laboratories will not be statistically significant. Billingham and Medawar (124,125) introduced the concept of an anatomically constant wound for the study of massive skin injuries. For skin injury models, the entire thickness of the skin (epidermis, basement membrane, and dermis) was consistently excised down to the layer of the underlying muscle and fascia. Except for edge effects, this model standardized the wound environment from one animal to another, making it possible to obtain statistically significant results from a study with several animals. For the remainder of this article, all animal results that are discussed will be from this wound model. In clinical cases, the nature of the wounds is different; such issues will be discussed in greater detail later in this article.

Mammals possess a small, quite limited, ability to spontaneously regenerate damage inflicted to most of their tissues and organ systems. The epidermis and basement membrane in the skin regenerates spontaneously, provided that there remains an intact, underlying dermal layer (1). In another example, a small, cylindrical defect (<1 cm in diameter) in mammalian long bones is spontaneously refilled with normal bone (126), and a gap <5 mm in a transected rat sciatic nerve can be spontaneously regenerated to a moderate extent (127–131). In these cases, regeneration is observed without the implantation of any active ECM analog. More specifically in the case of skin

injuries, the epidermis and basement membrane have been observed to spontaneously regenerate to form fully mature tissue, but the adult mammal is unable to spontaneously regenerate the dermis (1,124,125,132). In studies of induced regeneration using grafting techniques, it is important to first understand the fundamentals of behavior of the tissues involved, specifically the type of tissue that can be spontaneously regenerated, in order to understand the effectiveness of the graft. Regeneration of tissue that has resulted from the exogenous graft will be referred to as induced regeneration. In the case of studies of skin regeneration, the presence of new dermis following grafting of a full thickness skin wound with an appropriate device will be considered induced regeneration.

CHARACTERISTICS OF EXTRACELLULAR MATRIX ANALOGS THAT DEFINE BIOACTIVITY

Fundamental Design Principles for Tissue Regeneration Scaffolds

Porous scaffolds are utilized in the study of tissue regeneration; the term active extracellular matrix analogs (bioactive analogs) refers to scaffolds that induce regeneration of normally nonregenerative tissues following severe injury. Bioactivity is observed in only a limited number of scaffold variants, and is measured by the scaffold’s ability to prevent irreversible repair behavior while inducing regeneration. With a bioactive scaffold, the cells, cytokines, and biological exudate in the wound site interact with the scaffold such that the mechanisms and kinetics normally associated with spontaneous wound closure by wound contraction and scar synthesis (repair) are modified leading to the induced regeneration response. Bioactivity of tissue regeneration scaffolds has been observed to depend on the structural characteristics of the scaffolds, notably the chemical composition, the template pore structure, and the template residence time (1). These characteristics, and any governing models to describe cellular behavior in regeneration scaffolds, will be discussed in the following sections.

Template Residence Time

The residence time of an implanted scaffold is a critical variable that helps to define the bioactivity of the scaffold. For physiologic tissue to be synthesized in the wound bed, the scaffold must degrade in such a way that it does not interfere with the production of physiologic tissue. Empirical evidence supports a requirement for the implanted active ECM analog to be capable of isomorphous tissue replacement, that is, degradation of the active ECM analog at a rate of the same order as the rate of synthesis of new tissue (1).

These considerations are consistent with a model that defines a scaffold residence time with both an upper and a lower bound. Using the isomorphous tissue replacement model, the appropriate time period for scaffold residence is approximately equal to the time period required to synthesize a mature tissue at the specific site by regeneration. A reasonable approximation of the time for regeneration is the time period observed for the conventional healing

process of a wound that involves wound contraction and scar formation at the anatomic site of interest. In the case of a full-thickness skin wound, the healing time is ~ 25 days (1,3,133). As the intact scaffold cannot diffuse away from the wound bed, the simplest method for achieving isomorphous tissue replacement requires the macromolecular scaffold structure to be degraded by enzymes in the wound bed into low molecular weight fragments that are able to diffuse away. In a model of template degradation characteristics, the lifetime of the scaffold in the wound bed can be defined by the time constant for degradation (t_d) and can be compared to the time constant for a normal healing process of a wound at the anatomic site of interest (t_h). For isomorphous tissue replacement:

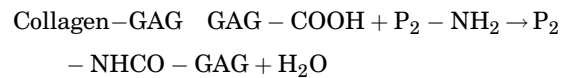
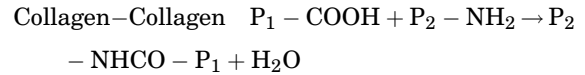
$$\frac{t_d}{t_h} = O(1)$$

The isomorphous tissue replacement hypothesis has been supported by observations made with different implanted devices that were studied in the context of *in vivo* synthetic processes. When the ratio of t_d/t_h was much >1 , the scaffold remained in the wound bed virtually as a nondegradable implant, and dense fibrotic tissue similar to scar was observed to be synthesized underneath the scaffold (1,133). As the ratio became considerably <1 , the initially insoluble scaffold became rapidly degraded and did not induce a regenerative healing process in the wound bed; instead, wound bed healing was marked by scar synthesis similar to the healing process observed in an ungrafted wound (109).

Satisfying the principle of isomorphous tissue replacement requires adjustment of the structure of the ECM analog such that the biodegradation time constant of the scaffold (t_d) closely matches the healing time constant (t_h). In the case of collagen-based ECM analogs, degradation of the scaffold in the wound bed is accomplished primarily by collagenases present in the wound site. Reduction of the biodegradation rate of collagen scaffolds has been achieved both by introducing glycosaminoglycans (GAGs) into the collagen mixture and by cross-linking collagen fibers to one another: scaffold resistance to degradation is increased with an increase in the density of cross-links (103,105,108,117,134). For collagen-GAG scaffolds, precipitating the two polymers together under an acidic pH mixes the glycosaminoglycan and collagen components, and the scaffold is fabricated from the coprecipitate. Cross-linking of the collagen fibers can be accomplished by a variety of different techniques, both physical and chemical.

Dehydrothermal (DHT) cross-linking is a physical technique that has been often utilized for cross-linking collagen and collagen-copolymer scaffolds, and allows for the creation of a variety of cross-linking densities. Dehydrothermal cross-linking involves exposure of the scaffold to a high temperature (90–120 °C) under vacuum (i.e., 95 kPa), leading to removal of water from the scaffold. When the water content in the ECM analog is removed below $\sim 1\%$, drastic dehydration of the scaffold leads to the formation of inter-chain amide bonds through condensation (135). This cross-linking reaction is a condensation reaction involving carboxylic groups from glutamyl/aspartyl residues on collagen

polypeptide chain P_1 and the amino groups of lysyl residues on an adjacent P_2 chain to yield covalently bonded collagen fibers. In addition, graft copolymers of collagen and GAG are formed by dehydration, leading to cross-link formation by condensation of amine groups on collagen chains (denoted P_1 and P_2 below) with carboxylic groups of glucuronic acid residues on GAG chains (104).



The density of cross-links formed through DHT cross-linking depends on the temperature as well as the length of exposure, with higher temperatures and longer exposure lengths producing a higher density of cross-links (104,105). Ultraviolet (UV) treatment is a second physical cross-linking technique that can create cross-links between collagen fibers (136–138). Additionally, cross-linking can be induced chemically by introduction of glutaraldehyde (GT) (104) or with 1-ethyl 3-(3-dimethylaminopropyl)carbodiimide (EDAC) (139). These chemical cross-linking techniques are considerably more powerful than the previously noted physical methods and lead to a much higher cross-link density and a much longer time constant of degradation (t_d). Chemically cross-linked scaffolds must be extensively washed to remove all trace of the cytotoxic cross-linking chemicals prior to use in the wound bed; in addition to concerns regarding washing away the excess chemical, some chemical cross-linkers function by having all or a portion of the chemical compound becoming part of the cross-link. In this case, slow degradation of the scaffold could result in the release of potentially cytotoxic agents into the wound site (1,3).

By using both physical and chemical methods, the cross-linking density of a particular collagen or collagen-GAG scaffold can be effectively adjusted to create a wide range of enzymatic degradation rates. These cross-linking tools can also be applied to a multitude of other scaffolds, fabricated from both natural and synthetic materials. As different wound sites and the same wound sites in different species have been observed to exhibit very different time constants for healing (t_h), it is necessary to adjust the degradation rate of the ECM analog to the characteristics of the specific wound bed site and species in order to satisfy the principle of isomorphous tissue replacement and to induce regeneration.

Critical Cell Path Length

Migration of cells into the active ECM analog (regeneration template) is critical for the synthesis of new tissue. The use of porous templates allows for more rapid incorporation of cells into the template. While the effect of the structural characteristics of the template will be discussed in the following sections, there is another important characteristic to consider: an adequate source of metabolites (i.e., oxygen, nutrients) available to the cells. There are two mechanisms available for transport of metabolites to and

waste products from the migrating cells: diffusion to and from the surrounding wound bed or transport along capillaries that have sprouted into the scaffold as a result of angiogenesis. While angiogenesis becomes the limiting factor for long-term cell survival and growth, significant angiogenesis is not observed for the first few days after implantation of the template. Therefore, early cell survival inside the scaffold is completely controlled by the diffusional mode of transport of metabolites from the wound bed.

A simplified model of cellular metabolic requirements and nutrient diffusion characteristics can be utilized to describe the critical cell path length (L_c) for cellular migration into a scaffold; the model assumes a totally diffusion-based mode of metabolite transport. Beyond the critical cell path length, diffusional transport will not provide for cellular survival; this model is important for predicting the initial “take” of a graft, as diffusional transport, along with prevention of graft rejection or infection, will be responsible for initial cell survival and graft “take”. Such a model simplifies the cell’s metabolic requirements by defining a “critical” nutrient that is metabolized by the cell at a rate of R mol/cm³/s. This nutrient is transported to the scaffold from the wound bed, where there is a constant concentration of nutrient, C_0 , that remains essentially constant due to resupply from the organism that acts as an infinite source. The transport processes are modeled by the diffusivity of the nutrient through the scaffold, D , with units of square centimeters per second (cm²/s) and a length of diffusion, L , with units of centimeters. The *cell lifeline number*, S , is a dimensionless number that is derived from these variables by dimensional analysis to express the relative importance of the consumption of an essential nutrient by the cell and the diffusion of the nutrient from the wound bed into the scaffold:

$$S = \frac{RL^2}{DC_0}$$

The magnitude of the cell lifeline number describes three distinct cases of cellular migration into the scaffold in the absence of angiogenesis. For $S \gg 1$, the rate of consumption of the critical nutrient exceeds the supply provided through diffusion processes, resulting in the death of cells that have migrated that distance L into the scaffold. For $S \ll 1$, the supply of the critical nutrient by diffusion exceeds the rate of consumption, resulting in cell proliferation and migration further into the scaffold. The steady-state condition, $S = O(1)$, describes the state where the value of L becomes the critical cell path length, L_c . L_c describes the longest distance from the wound bed edge that a cell can migrate while depending solely on diffusional processes as the source of metabolites before angiogenesis begins. In the case of many common nutrients with a low molecular weight, L_c can be predicted using this model on the order of 100 μm (109). This estimation of L_c indicates that there needs to be close contact between the wound bed and the scaffold immediately after grafting in order for the initial migration of cells into the wound site to take place prior to angiogenesis and suggests that very stiff ECM analog grafts will show inferior results. A stiffer scaffold will be unable to conform to the wound bed, thus reducing the

effective distance that cells can migrate into the scaffold and be metabolically supported prior to angiogenesis, eventually reducing overall graft “take”. Unfavorable values of the scaffold-tissue surface tension can also prevent close contact.

Chemical Composition

The chemical composition of an ECM analog plays a critical role in defining the bioactivity of the device. In order to design a suitable device for use in wound healing that will induce regeneration rather than repair mechanisms, it is important for the device to keep the wound edges apart and to drastically modify the healing processes in the wound bed to yield physiologic tissue rather than scar. At the outset, the chemical composition defines the ligands that are displayed on the scaffold surface. Cellular activities such as binding, migration, and contraction are all mediated by interactions between the integrins expressed by a specific cell type and the ligands available on the scaffold surface. The design of a scaffold to be used for tissue engineering purposes must accordingly be fabricated in such a manner and using such specific materials as to allow for appropriate binding between cell and scaffold. The specific cell–matrix binding that appears to be required in cases of induced regeneration is that which blocks contractile cells from implementing their program of wound contraction (1).

A number of different materials have been used for the production of scaffolds for tissue engineering. Several synthetic nondegradable polymers, such as poly(dimethyl siloxane), have been utilized; these polymers, which parenthetically violate the principle of isomorphous tissue replacement, have not exhibited the ability to induce regeneration. Degradable synthetic polymers, such as poly(lactic acid), can be fabricated to satisfy isomorphous tissue replacement and the surfaces can be modified to properly induce cellular binding, but have not been shown to prevent contraction and scar formation. A chemical composition that has been used successfully to induce regeneration has been a graft copolymer of type I collagen and a sulfated GAG (1). These natural polymers are capable of facilitating cell binding in part due to their expression of natural ligand-binding sites.

Template Pore Structure

Although the chemical composition is a critical component in defining the biological activity of the scaffold, it is not the sole characteristic. The biological activity of a particular ECM analog also depends significantly on the architecture of the three-dimensional (3D) network. Having migrated into the scaffold, the cell interacts with the structure that defines the porous scaffold, making use of its cell surface receptors (integrins) to bind to specific ligands on the scaffold surface. The first critical components of the scaffold pore structure to consider is the open- or closed-cell nature of the scaffold and its relative density: A tissue engineering scaffold must possess an open-cell pore structure with a relative density below a critical value. Three-dimensional porous structures can possess an open- or a closed-cell pore structure; in a closed-cell structure, each

individual pore is separated from adjacent pores by membrane like faces while open-cell pore structures exhibit interconnectivity between adjacent pores. Interconnectivity is critical for cells to be able to migrate through the 3D structure. The most important structural feature of porous scaffolds is the relative density (R_d): the density of the scaffold divided by the density of the solid from which it is made. The porosity of a scaffold, or pore volume fraction, is defined as $(1 - R_d)$. The relative density defines the amount of solid material available for cells to bind to; when the pores are closed or when the relative density is too large, cells are not able to migrate through the scaffold, a significant impediment for using such a scaffold for a tissue engineering application. Such structural aspects also suggest that in designing scaffolds for tissue engineering applications, there is a critical number of cells required for scaffolds to appear bioactive. There must be a large enough area available for cells to bind to in order to support a large enough population of cells within the scaffold; the existence of a critical density of cells has been hypothesized as a result of a number of experiments studying cell-scaffold interactions (1).

The development of a highly detailed model, describing the number of receptors utilized per bound cell and the nature of the binding and receptor sites, is required to describe even a simple interaction between a cell and a generic scaffold surface. However, a more generic explanation can be used to indicate both the complexity of the cell-scaffold interaction and the significant influence the scaffold pore structure has on scaffold bioactivity. In particular, we will examine the affect of another critical factor on scaffold bioactivity: mean scaffold pore size.

The structure of a porous scaffold is defined by the pore volume fraction, mean pore size, and pore orientation in the scaffold. All of these characteristics have been shown to significantly affect the bioactivity of the scaffold. The pore volume fraction and the mean pore size together define the specific surface area of the scaffold, the total surface area of pore walls available for cellular binding. Increasing the mean pore size while keeping the pore volume fraction constant decreases the specific surface area of a scaffold. Decreasing the pore volume fraction and keeping the mean pore size constant increases the specific surface area (140). It has been estimated that a 30-fold increase in pore diameter leads to a 27-fold decrease in specific surface (2). A change in the specific surface area of the scaffold significantly changes the area available for cells to bind to. More specifically, the surface density of bound cells (Φ) in a 3D dimensional porous scaffold is a function of both the density of bound cells in the scaffold (ρ) and the specific surface of the scaffold (σ):

$$\Phi = \frac{\rho}{\sigma}$$

This calculation suggests the significance of the specific surface of the scaffold in defining the scaffold bioactivity. If the specific surface is too small due to large pores, an insufficient number of cells will be able to bind to the scaffold and the cells that remain free will contribute to the spontaneous repair mechanism. There is a minimum pore size as well, defined by the characteristic dimension of

the cell, $\sim 10\text{--}50\ \mu\text{m}$ for most cell types. When the scaffold pore size is smaller than this critical dimension, cells will be unable to migrate through the porous structure, and will be unable to infiltrate and bind to the template. These upper and lower bounds of the scaffold pore diameter, mediated by cell size and specific surface requirements, have been determined experimentally for each cell type for tissues where regenerative templates have been used (3,141); however, future work is necessary to develop a better understanding of cell adhesion and its relation of scaffold structure.

The shape of the pores that make up the porous scaffold must also be considered; slight changes in the mean shape of the pores can result in significant variations in the mechanical properties of the scaffold (140). In addition, changes in mean pore shape may also play a role in defining the areas of the scaffold available or unavailable for binding and in defining available directions for cell migration. The template pore structure plays a very significant role in defining the overall bioactivity of the scaffold and the open or closed-cell nature, the mean pore size, relative density, and pore shape and orientation all are critical components to consider.

TISSUE ENGINEERING OF THE SKIN

Structure and Function of Skin

Mammalian skin is a stratified tissue made up of three distinct layers of tissue: the epidermis, the basement membrane, and the dermis. Each tissue displays unique structural and functional properties as well as distinctive responses to injury or damage. These tissues, an epithelial layer (epidermis in skin), a basement membrane layer, and a stromal layer (dermis in skin) make up the previously described tissue triad for the skin (1,8,142). Using the tissue triad to model a generic tissue, the epithelia covers all body surfaces, tubes, and cavities, and is separated from the underlying stroma by a continuous basement membrane layer. As the basement membrane is totally acellular and is not penetrated by the vascular system, the survival of the epithelia depends on diffusion of metabolites, nutrients, oxygen, and waste products across the basement membrane to and from the stroma. The stromal layer contains the vascular system and other supporting tissues (connective tissues) that serve to nourish and anchor the basement membrane and the epithelia. In addition, while both the stromal and the basement membrane layers contain an extracellular matrix structure, the epithelial layer does not. In the following sections, the structure and function as well as the response to injury of each of the layers of the tissue triad that constitutes adult skin will be described in detail.

Morphology and Function of the Epidermis

The epidermis is the exterior layer of tissue that makes up the skin. The epidermal layers act as a physical barrier to protect the organism against microorganisms; prevents organismic dehydration; and protects the organism from mechanical, thermal, chemical, and UV insults. It is a

stratified tissue, consisting of five distinct tissue layers (strata) that form a tissue ~0.1 mm thick. These layers represent a cell-maturation gradient along which cells move, from the interior to the most exterior layers; cells become increasingly mature and keratinized during this migration process. The most interior layer is the basal cell layer, known as the *stratum malpighii* or *stratum germinativum*. The next layer is the prickle cell layer, also known as the *stratum spinosum*, describing the prickly morphology of the cells. The granular layer, also known as the *stratum granulosum*, is the next layer consisting of keratohyalin granules (intracellular granules) that contribute to the process of keratinization. The *stratum lucidum* is the fourth epidermal layer, found only in the very thick skin associated with the fingertips, palms, and soles of the feet. The most exterior layer is the cornified or horny layer, the *stratum corneum*, made up of flattened cell remnants that are fused together forming a compact layer of keratin, the fibrous protein that makes up the external armor of the epidermis (142). The cell-maturation gradient observed in the epidermis starts with immature cells (keratinocytes) in the basal layer; these cells undergo mitosis and migrate through the cell layers toward the cornified layer over a period of 25–50 days. Along this path, the cells become increasingly keratinized, until they reach the cornified layer where the dead cells are desquamated. This stratified tissue is avascular, relying on the underlying dermis for a nutrient supply.

The five cell layers and the individual keratinocytes within each layer are bound together by desmosomes; keratin filaments, a meshwork of filaments inside the keratinocyte cytoplasm (also known as tonofilaments), anchor neighboring cells to one another. Additional mechanical stability is provided to the epidermis by its attachment to the underlying dermis; bonding at the epidermal-dermal junction is mediated by the basement membrane. Hemidesmosomes, located inside the cell membrane of basal cells, attach to the epidermis and to the basement membrane by means of tonofilaments via junctions on the subbasal plates (142). This construct forms a stratified, mechanically stable keratinized epithelium able to withstand the thermal, mechanical, chemical, and UV insults to which the body is continuously exposed.

Morphology of the Basement Membrane

The basement membrane, also known in the literature as the basal lamina, is found in many different tissues as an acellular, avascular layer between the avascular, cellular epidermis (no ECM) and the cellular, vascularized dermis (developed ECM). The basement membrane performs a number of significant duties; notably, it provides a secure and adhesive layer to facilitate a strong connection between the epithelia and stroma, serves as a boundary that can regulate cell and molecular movement, provides a scaffold to facilitate repair following injury, and facilitates differentiation and growth of the epithelial and stromal layers (9,143–146).

The basement membrane is an acellular, avascular stratified tissue made up of three distinct strata that are, in total, ~100 nm in thickness. The basement mem-

brane structure is often observed under light microscopy as having only a single layer, termed the *lamina densa*; electron microscopy of the basement membrane reveals the *lamina lucida* and the *fibroreticularis* (142). The first layer of the basement membrane, sitting adjacent to the basal cell layer of the epidermis, is termed the *lamina lucida*, an electron-lucent membrane ~20–40 nm in thickness that consists primarily of the glycoprotein laminin. The middle layer of the basement membrane, termed the *lamina densa*, consists predominantly of type IV collagen, is ~40–50 nm in thickness, is significantly more electron-dense than the *lamina lucida*, and is the region visible under light microscopy. The final layer of the basement membrane is located adjacent to the underlying dermis and is termed the reticular layer, also known as *fibroreticularis*. This electron-lucent layer is composed primarily of type VII collagen fibers and is responsible for fixing the basement membrane to the underlying dermis by anchoring fibrils attached to specific anchoring plaques that are embedded in the underlying dermis (147,148). While the primary components of the basement membrane are type IV and type VII collagen, the basement membrane also contains significant amounts of chondroitin sulfate, heparin sulfate, fibronectin, tenascin, nidogen, enactin, thrombospondin, and 1-microglobulin (149). When viewed from a more macroscopic scale, the topography of the basement membrane surface appears as an undulating line between the dermis and the epidermis, significantly increasing the surface area between these two structures. This undulating structural feature, termed rete ridges, will be discussed in greater depth in the next section describing the dermis.

Morphology and Function of the Dermis

The dermis is the final component of the skin tissue triad, lying below the basement membrane and above the underlying muscle and fascia. The dermis, considered anatomically to be a single layer, actually consists of two zones: the papillary dermis and the reticular dermis. The papillary dermis is the upper zone adjacent to the basement membrane and consists primarily of loosely packed collagen fibers. The papillary dermis forms the upward projections of the dermis into the epidermis that define the rete ridges; these projections are filled with capillary loops responsible for providing metabolites, nutrients, and oxygen to the epidermis. In addition, the papillary dermis contains fine axonal connections of unmyelinated sensory nerves that extend up to the basement membrane. The bulk of the dermis, termed the reticular zone, is found below the papillary dermis. The reticular zone is comprised of thicker and more closely packed collagen fibers as well as a significant content of elastin fibers that are interlaced with the collagen fibers to form an isotropic, collagen-elastin network. While collagen fibers are highly crystalline microfibrils that have a limited stretching ability and provide the strength to a tissue, elastin fibers are considerably thinner and amorphous (noncrystalline), providing the ductile strength (stretching without yielding) of a tissue (1). While the strength of the dermis is defined by the collagen content and the ability of the dermis to bend and deform without

permanent damage is defined by the elastin content, it is the combination of collagen and elastin fibers that is responsible for the robust nature of the dermis.

The dermis has two major functional roles: providing mechanical stabilization for and metabolic support to the epidermis. The combination of mechanical strength and deformability gives the dermis the ability to provide a stable base for the epidermis that is able to withstand the substantial shear, tensile, and compressive forces associated with ordinary activities that would cause an unsupported epidermis to fail. In addition, the undulatory nature of the dermo-epidermal junction allows for the intimate presence of an extensive dermal vascular system that provides metabolic support (providing nutrients and oxygen while removing waste products) to the avascular epidermis. The rete ridges also provide increased surface area for attachment of the epidermis to the basement membrane and the basement membrane to the dermis, increasing the strength of the dermo-epidermal connection and enhancing the surface area available for the capillary loops to provide metabolic support to the epidermis. In addition to the two zones that make up the dermal layer, hair follicles, sweat glands, and oil-secreting glands originate in the dermis and extend through the basement membrane and epidermis to the skin surface. The dermis also provides tactile sensation through the unmyelinated sensory nerves that extend through the dermis up to the dermo-basement membrane junction, and allows for thermoregulatory control (1,142).

Current Treatment of Massive Skin Loss

Traditional treatments of a severe skin wound have focused on developing a temporary technique or product that serves to close the wound, preventing infection and dehydration (important for large skin wounds) (1). Historically, attempts have been made to treat severe wounds and burns dating back almost 3500 years, and have included a wide variety of temporary devices such as membranes of organic and synthetic polymers, skin grafts from animals (heterografts or xenografts), skin grafts from human cadavers (homografts or allografts), and skin grafts from the patient (autografts) (133). Allografts are used as a temporary covering for excised (cleaned) wounds prior to autograft, where the allograft is removed and the permanent autograft is placed into the wound. Xenografts are typically taken from the pig due to the great affinity between human and pig skin. Like the allograft, xenograft skin is a temporary wound dressing used until autograft. Temporary dressings immediately reestablish the skin barrier, decrease inflammation and risk of infection, decrease fluid loss, and reduce patient mortality. More recently, the need to develop technologies suitable for treating severe skin injuries over large areas has increased, increasing the requirement for a material to rapidly close a severe skin wound. The fundamental reasons for such a change are that an increasing percentage of patients survive the acute phase of the injury due to improved medical care and that the widespread use of early escharectomy (complete debridement of the wound immediately after injury) requires immediate coverage of large wound areas. It is often not

possible to harvest enough autograft tissue to cover a large wound. In these cases, the graft is perforated and then stretched to cover much more space than the original tissue; this meshing process decreases the quality of regeneration, but increases the area that can be treated and will be discussed in greater detail later in this section. There are also a number of problems related to the use of temporary dressings such as allografts and xenografts. Especially in cases of severe burns over large areas of the body, the xenograft or allograft may need to remain in place for a significant period of time until autografting or other treatments are possible. In transplantation of donor tissue, histocompatibility becomes an issue; typically, xenografts and allografts are rapidly rejected, usually within a month of transplantation. This phenomenon illustrates the concept of host-versus-graft disease, where the patient's (host's) body mounts a host defense that ultimately destroys the implanted tissue (graft). Histocompatibility antigens expressed in the transplanted tissue are identified as foreign by the patient's immune system, leading to an inflammatory and immune response that destroys the grafted skin; such problems are also seen in transplantation of almost all other organs and tissue in the body (i.e., liver, kidney, heart, lungs, bone marrow). Patients with allografts or xenografts can be immunosuppressed using a variety of drugs to prevent host-versus-graft disease and prolong the viability of the transplanted tissue (150); however immunosuppression introduces a variety of complications such as a decreased ability to fight infection, a prime concern for people with severe skin injuries.

Polymeric membranes used as temporary dressings often lack biological activity due to the chemical composition and structure of the membrane; these membranes often have to be removed after only a few days due to lack of formation of physiological structures and incidence of infection (133). Often a temporary graft, such as a synthetic or organic polymeric membrane, xenograft, or allograft, is useful in early management of a wound while an autograft site is prepared. Permanent treatments for massive skin loss have traditionally been focused on the autograft technique.

Despite the presence of other grafting techniques, the autograft is the current clinical standard; it addresses both the urgent need to cover an exposed wound and results in an adequate long-term result. Under ideal conditions and in the case of small wounds, when full-thickness skin wounds are treated with an autograft, an almost fully functional skin has been observed to regenerate. However, the skin replaced via the autografts has been observed to lack hair follicles and other adnexa. Despite these missing components, this skin replacement remains functional for the remainder of the patient's lifetime. Two major problems complicate the use of autografts to treat full-thickness skin wounds: the creation of a second wound (donor site), and the requirement for large autografts in the case of massive skin injury. The removal of the autograft results in a secondary full-thickness skin wound that eventually becomes reepithelialized, but considerable scar formation and contraction are observed at the donor site. This factor coupled with the usual need for large amounts of autograft

tissue due to the typical size of severe skin defects in humans has resulted in the surgical meshing procedure, where a small amount of autograft tissue is harvested, then passed through a device to cut a pattern of slits in the autograft tissue; this tissue is then stretched, greatly increasing the area of coverage and thus decreasing the amount of harvested tissue needed to cover the wound. This technique is not without problems as scar synthesis is observed in the areas of the wound not covered by the stretched autograft mesh, resulting in a pattern of scar that greatly reduces the value of the autografting procedure. Due to clinical attempts to minimize the size of the autograft wound, meshing is used almost exclusively to treat skin wounds. This results in adequate coverage and closure of the wound, but the coverage is marked by considerable scar synthesis and contraction, reducing the aesthetic and functional value of the treatment.

It is this inability to utilize the autograft without considerable scar formation as well as the requirement for the creation of a secondary wound site that have provided the stimulus to investigate alternative dressing options that could potentially lead to regeneration of physiologic tissue rather than healing by contraction and scar formation. The resulting technologies will now be discussed in detail in the following sections.

TECHNOLOGIES FOR REPLACEMENT OF THE SKIN

A number of technologies have been developed in an attempt to induce regeneration of skin following injury, both in conjunction with or without the use of an autograft. These technologies have met with differing levels of success. There are five main technologies for treating massive skin wounds by grafting that will be discussed in this article: sheets of epidermis cultured *in vitro* (Cultured Epithelial Autograft, CEA), cell-seeded nylon scaffolds (Living Dermal Replacement, LDR), a 3D living bilayer first cultured *in vitro* with dermal and epidermal cells (Living Skin Equivalent, LSE), a collagen scaffold that was either seeded with keratinocytes or implanted as an acellular construct (DRT), and a naturally derived collagen matrix (NDCM, Alloderm). These distinct procedures will now be discussed in detail, describing the design and manufacture of each device as well as the attendant experimental and clinical results.

Cultured Epithelial Autograft

Cultured epithelial autografts have been studied in both experimental and clinical settings as a possible treatment for massive skin injuries. This technique uses an epidermal graft that is grown *in vitro* and then implanted into the wound site to cover the skin defect. While initially used to provide immediate coverage of the wound site to prevent excessive fluid loss and infection, the CEA has also been studied in models to assess its potential as a permanent skin replacement graft. In the literature, the CEA technology has also been referred to as a keratinocyte sheet, cultured epithelia sheets, or cultured autologous keratinocyte sheets. This technique to culture keratinocyte cells to form an epithelial sheet was utilized because keratino-

cytes make up ~90% of mammalian epidermal cells (1). There are three major sources for keratinocytes that have been used in cultivating keratinocytes sheets: keratinocytes from disassociated cells, keratinocytes from epidermal explants, and suspensions of pellets of disaggregated keratinocytes (65).

The CEA technology relies on culturing keratinocytes isolated from the patient to produce a graft of autologous tissue, removing any immune complications observed in the case of xenografts or allografts; additionally, since the epidermis can spontaneously regenerate, epidermal tissue harvested from the patient will regenerate without further scarring. Therefore, development of a successful CEA relies upon the development of *in vitro* methods for rapid, serial cultivation of human keratinocytes from a disaggregated cell suspension; these techniques allow for expansion of the (small) harvested cell population by >10,000-fold in 3–4 weeks, a rate necessary to culture the volume of cells required to rapidly produce a keratinocyte sheet large enough to cover a wound site in a clinically acceptable time period (76,77,151,152). In a clinical setting, keratinocytes are typically isolated from skin biopsies; the biopsy tissue is then treated enzymatically to allow removal of the dermal tissue and to dissociate the remaining epidermis. This sequential process prevents contamination of the keratinocyte cell line with dermal cells (mainly fibroblasts). The keratinocytes can then be cultured using a defined *in vitro* process (73,76,77). All of these techniques can be utilized to culture the requisite cell expansion.

Using these established cell culture techniques, an intact, coherent sheet of stratified epithelium can be produced *in vitro* that is on the order of four to six cell layers thick and is bound together by the desmosomes seen in the normal epidermis. Similar to normal epidermis, sparse keratin fibers are observed running parallel to the long axis of the flattened keratinocyte cells in this new epithelial layer. While keratinization is not always observed, the maturity of this newly formed neoepidermis is moderately high. The epithelial maturity can be acutely affected by the identity of the substrate on which the CEA sheet is grown. When the stratified neoepithelium is grown on collagen gels, hemidesmosomes are not synthesized; however, the use of a surface formed from reconstituted basement membrane led to synthesis of hemidesmosomes and a more mature epidermis (1,148,153).

Studies of CEA development have indicated that a partially mature epidermis can be synthesized *in vitro* starting from disaggregated keratinocytes. There is no requirement for the presence of any dermal component or for fibroblasts in the synthesis of a neoepidermis. However, there is a temporary requirement for a nondiffusible substrate onto which the cells are grown in order to develop stratified and keratinized cell layers. While contact with specific connective tissue surfaces can induce formation of the mature neoepidermis, it is still possible to develop a less mature epidermal layer with culture on plastic or glass surfaces (1).

In the clinical setting, keratinocytes are harvested from the patient via a biopsy; the cells are then dissociated, cultured, and expanded *in vitro* for ~3 weeks to form a neoepidermis. At the end of this period, the mature,

keratinizing epidermal layer that forms *in vitro* is then implanted directly into the wound site. One major drawback to the use of the CEA is that due to its extreme friability, handling and grafting the CEA into a wound bed require extreme care. Additionally, the grafted site must be kept immobile so that the CEA can remain in place and not break apart. After implantation, the epidermal cells continue to multiply and spread, covering the entire wound. Clinically, the success of the CEA treatment depends significantly on the condition of the skin wound. The adhesion (take) of the CEA was very different depending on whether the CEA was grafted onto a full-thickness or a partial-thickness skin wound. In the case of partial-thickness wounds, the *take* of the graft has been very good and the CEA has been used to cover significant areas as large as half of the total body surface, making the CEA a life-saving, although temporary graft (133). *Take* was considerably inferior in the case of full-thickness skin wounds, where there was no underlying dermis to support the neoepidermis; in particular, one persistent problem was the formation of blisters under large areas of the graft (avulsion). Regardless of whether the graft was placed upon a partial- or full-thickness skin wound, the resulting CEA graft exhibited mechanical fragility due to a lack of three specific structural features present in normal, adult epidermal and basement membrane that serve to tether these layers onto the underlying dermis: the 7-S domain of type IV collagen, anchoring fibrils, and rete ridges (154). These structures are required for the formation of a physiological dermo-epidermal junction, and CEA grafts have failed to induce formation of these structures, a collagen fiber architecture, or the elastin fiber network that are all observed in the normal, adult dermis (133). Without these structures, the CEA cannot be used as a permanent skin replacement; instead, the CEA is often used as a temporary coverage as part of a larger treatment regimen. The CEA also exhibits high vulnerability to cytotoxins and bacterial proteases, as the CEA does not behave as a full-thickness graft in preventing infection. During the initial period after grafting, the CEA is extremely sensitive to the effects of bacterial or fungal infections of the wound bed: a full-thickness graft such as the autograft can tolerate infections that result in a near or complete loss of the CEA.

Living Dermal Replacement

The LDR was developed to be a more permanent treatment for severe skin injuries. While the CEA relied upon developing a stratified epidermal layer *in vitro* that could be used to permanently treat injuries to the epidermis and basement membrane and to temporarily treat full-thickness skin injuries, the LDR was developed in an attempt to utilize a structure that could permanently treat full-thickness skin injuries. The LDR technology used an acellular scaffold cultured with both fibroblasts and keratinocytes *in vitro* prior to implantation. In addition to introducing a cell population to the wound site, the scaffold structure was included to provide an immediate 3D architecture for both structural and cellular support that could be synthesized and implanted rapidly into the wound site. The acellular LDR scaffold consists of a copolymer of glycolic acid (90

wt%) and lactic acid (10 wt%), termed polyglactin-910 surgical mesh (PGL). The PGL fibers, $\sim 100\ \mu\text{m}$ in thickness, were knitted into a mesh that exhibited a pore structure with a characteristic dimension of $280 \times 400\ \mu\text{m}$. This mesh structure presented a large-weave structure to the cells, relative to the characteristic cell dimension of $\sim 10\ \mu\text{m}$, allowing rapid cell incorporation into the mesh as well as diffusion of an ample supply of nutrients to support cell activity. The polyglactin mesh was cultured with fibroblasts until all of the pores in the mesh were covered with cells; the confluent cells were observed to have begun to synthesize several important ECM components *in vitro*. Immediately prior to grafting, the upper surface of the polyglactin scaffold that was confluent with fibroblasts was seeded with keratinocytes in an attempt to form a bilayer graft that would mimic the structure of skin. Once the keratinocytes reached confluence on the surface of the PGL mesh, the entire structure was grafted into the wound site (88,92).

The LDR was studied as a stand-alone graft to be used temporarily prior to eventual, permanent closure by an autograft, primarily in a mouse model where the device was grafted into full-thickness skin wounds (1,133). A thin, fragile epidermal layer developed initially by 10 days postgrafting and it became cornified as early as 20 days postgrafting. In addition, by 20 days the nylon scaffold degraded completely with minimal inflammatory response. While the interface between the graft and the wound bed stained positive for laminin, consistent with the synthesis of a lamina lucida layer, no other component of the basement membrane was synthesized. In addition, rete ridges were not synthesized and a thick fibrotic tissue layer characterized by a large fibroblast population and vascular in-growth was observed below the newly synthesized epidermal layer. Additionally, the cellular component of the LDR was critical in achieving these results. When fibroblasts were not seeded into the scaffold, the mesh rapidly separated from the wound bed and fibrovascular in-growth did not occur; the presence of keratinocytes in the graft was required to prevent contraction of the wound site (88,92). While the LDR showed the ability to induce regeneration of a neoepidermis, the LDR did not exhibit the ability to induce regeneration of a complete basement membrane or a dermal layer.

Living Skin Equivalent

Preparation of the LSE graft utilized a novel approach for producing a full thickness graft with 3D architecture. While the LDR consisted of an acellular, synthetic scaffold that was cultured *in vitro* with dermal cell then seeded with epidermal cells immediately prior to implantation, the LSE approach utilized dermal cells to create a cellular, organic structure *in vitro* that could be then seeded with epidermal cells and implanted into the wound site as a cellular, bilayer neotissue (78–80,82). The skin equivalent was formed by populating a collagen lattice with heterologous fibroblasts that *in vitro* contracted the lattice and synthesized additional extracellular matrix proteins that were incorporated into the base lattice. After this contraction period, the upper surface of the neodermal

layer was then seeded with a suspension of epidermal cells, primarily keratinocytes. Once seeded, the epidermal cells attached to the collagen scaffold, proliferated, and differentiated to form a multilayered, epidermal structure within 1–2 weeks. At this point, the collagen scaffold populated by fibroblasts with a neomature epidermis upper structure was then grafted into the wound site. This technology requires a significant culture period, necessitating temporary treatment of the wound for patients during the *in vitro* culture period if this technology were to be used in the clinical setting.

Detailed studies have been made of the structure of the LSE following *in vitro* culture but prior to grafting. The keratinocytes that were seeded onto the contracted collagen scaffold formed a multilayered, partially keratinized epidermis *in vitro* that included tonofilaments, keratohyalin granules, and desmosomes (82,155). The intercorneocyte lipid lamellae that were synthesized in the *stratum corneum* of the neoepidermis did not have the same repeating pattern of narrow and broad electron lucent bands that are responsible for the epidermal barrier properties in the normal, adult epidermis. As a result of this structural abnormality, the LSE has an increased water permeability compared with normal skin, and hence a greater chance for wound site dehydration (156). Although short segments of the lamina densa were observed along the dermo–epidermal convergence, the LSE did not exhibit a complete basement membrane layer at the end of the *in vitro* culture period (157,158). Rete ridges and skin appendages were also consistently absent. Continued structural and biological changes were observed in the LSE following grafting, indicating that remodeling was taking place (1,133).

The LSE was tested experimentally in full-thickness skin wounds in a rat model; the bilayer graft exhibited remodeling in both the neodermal and neoepidermal layers. A functional, fully differentiated epidermis was observed as early as 7 days following grafting, and by 14 days a vascularized subepidermal layer with many of the structural characteristics of normal dermis, such as a “basketweave” collagen fiber pattern, was present. While a pattern of collagen fibers was observed in the subepidermal layer, the fibers were much thinner and much more tightly packed than those fibers observed in physiologic dermis, and no rete ridge structure was synthesized (155). When the LSE was implanted into a mouse model, similar remodeling effects were observed in both the epidermal and subepidermal layers. A mature epidermal layer and basement membrane were synthesized *in vivo*, with only anchoring fibrils missing from these tissue structures. The subepidermal layer again exhibited densely packed collagen fibers, but a physiologic dermis was not observed, and rete ridges again failed to form (159). Experimental results across these animal models and a number of experimental trials were consistent in indicating that the LSE was able to induce regeneration of a mature epidermis and basement membrane, but a mature dermal layer was not observed and the dermo–epidermal junction remained flat, without any sign of rete ridges (1).

The results observed in experimental trials of the LSE led to a series of clinical trials for the LSE graft that focused

on its potential use for treatment of severe burn patients. Full-thickness skin wounds that covered >15% of their body surface area were grafted with the LSE. Extensive lysing of the graft was observed as early as 2 days post-grafting, and after 2 weeks only one patient showed any significant amount of *take* (~40%). The investigators concluded that the LSE was not appropriate to use as a permanent treatment to replace the autograft for burn patients (160). Additional clinical studies were performed on patients who had skin tumors removed where these acute wounds were not full-thickness skin wounds. While acceptable *takes* and no evidence of graft toxicity or rejection were observed in these trials, wound contraction by 10–15% was also observed; wound contraction to this extent was significantly larger than what is observed after grafting with an autograft. A biopsy taken from the graft site showed evidence of scar formation, and the authors hypothesized that the LSE was replaced by host tissue through more traditional modes of injury response (i.e., contraction and scar synthesis) (161).

A final series of clinical trials involving the LSE used the graft to treat chronic wounds due to venous ulcers, a chronic skin defect that has been observed to be of variable depth. Patients were diagnosed as having chronic ulcers when their wounds remained open for at least 1 month prior to LSE implantation, although the median duration of the ulcers for these patients was ~1 year. No information on the initial depth of each of the ulcers was reported, so it is not clear whether these were full- or partial-thickness skin wounds. In these studies, the use of the LSE was compared to a standard clinical treatment for such ulcers: bandaging with a compression regimen. Following a 6 month study, it was concluded that the time to wound closure was significantly shorter for patients treated with the LSE compared to those treated with standard bandages. However, observations from both experimental (animal models) and clinical trials indicated that while the LSE displayed a significant ability to regenerate both mature epidermal and basement membrane layers, dermal regeneration was not observed and normal skin architecture (i.e., collagen fiber architecture, rete ridges) was lacking in both animal and human models (1,162,163). More recent animal experiments utilizing the LSE have indicated that once implanted, the graft is rapidly incorporated into the host tissue and began to undergo remodeling; basic dermal organization, the appearance of specific ECM constituents such as type I, III, V, and VI collagen as well as elastin, and the gradual disappearance of contractile myofibroblasts occurred by 1-year postimplantation. However, complete regeneration of the tissue triad (epidermis, basement membrane, dermis) was not observed. The LSE possesses the adequate environment to remain bioactive >1 year post-implantation and provide a platform for long-term *in situ* therapy studying wound healing (164). While the LSE is used clinically to treat chronic (nonclosing) venous leg and diabetic foot ulcers and significantly increases the rate of wound closure and the likelihood of wound closure compared to conventional treatments (debridement followed by the application of a synthetic dressings), the LSE has yet to be able to replace the wound with normal, physiological skin.

Dermal Regeneration Template

The DRT is the fourth major paradigm investigated to treat massive skin injury. Like the LDR and the LSE, the DRT utilized a 3D scaffold as the basis for an implant to be grafted into a severe skin wound. But unlike the cell-seeded, synthetic, acellular scaffold of the LDR and the cell-seeded, organic, cellular LSE where the focus was not on developing a specific scaffold structure, the DRT was an acellular scaffold fabricated from primary components of the extracellular matrix (i.e., collagen, proteoglycans). The focus was on fabricating a bioactive ECM analog that presented a 3D architecture that induced the endogenous cell population in the wound site to regenerate the lost tissue (1,3). The DRT device sequentially performs two separate tasks during the wound healing process: prevent wound dehydration and infection while modifying the healing response to induce regeneration. The first task addressed by the DRT is the management of the acute phase of the clinical healing process; the graft must protect against severe fluid loss and prevent massive infection as a result of the open skin wound. The second task for the DRT to modify is the chronic phase of the wound healing process; in the chronic phase, wound contraction and scar synthesis take place, leaving a mechanically, functionally, and aesthetically inferior tissue compared to normal skin (repair mechanism).

The DRT is a two-stage device with a top layer of poly(dimethyl siloxane), a silicone elastomer, bonded to an active ECM analog beneath. The active ECM analog was designed to induce synthesis of new, physiological dermal tissue while the silicone elastomer was designed to prevent flow of exudate outside the defect, acting as a barrier to control moisture permeability and to shield the wound site from bacteria (109). The top silicone layer was designed so it would be easily removed after an initial healing period (Stage 1 of the healing response) so that a keratinocyte sheet could be grafted on top of the newly synthesized dermal bed, producing a stratified epithelial tissue on top of the neodermis. This grafting procedure was developed based on observations that the epidermis regenerates spontaneously provided that there is an underlying dermal structure, while the dermis does not. The graft design is also compatible with a view of the active ECM analog (template) as a structure that prevents contraction and scar formation, while the silicone elastomer layer protects the wound site from dehydration and infection. In the clinical setting (107), the silicone layer was replaced by a thin autoepidermal graft harvested from the patient once initial regeneration of the dermis occurred (~10–15 days). Following the introduction of a keratinocyte sheet, a mature epidermal and basement membrane layer forms on top of the regenerated dermis.

The DRT structure was optimized in studies utilizing animal models where it was observed that skin regeneration did not occur unless the ECM analog effectively delayed wound contraction. Preliminary optimization of the structural features of the ECM analog was quantitatively based on comparing the delay in the onset of contraction. The structural characteristics of an active ECM analog were individually studied to determine the optimal

chemical composition, template residence time, and template pore structure needed to create a bioactive scaffold that prevented contraction and induced regeneration (1,3,134).

The ECM analogs that were used successfully to induce skin regeneration are graft copolymers of type I collagen and chondroitin-6-sulfate, a GAG, with a collagen:GAG weight ratio of 98:2 (3). The porous structure of the DRT was produced using a freeze-drying process. The collagen-GAG (CG) copolymer was produced in slurry form from microfibrillar collagen, an aqueous glycosaminoglycan solution, and acetic acid. The slurry was frozen to a final temperature of -40°C and then sublimated [pressure <100 mtorr (13.3 Pa), temperature = 0°C]; after freezing, an interpenetrating network of ice crystals that is surrounded by collagen and GAG fibers has formed. The sublimation process converts all of the ice crystals formed in the frozen slurry into vapor and generates empty pores, creating a collagen-GAG scaffold formed from interconnected sublimated pores (3,165). The scaffold structure (pore size, specific surface) can be adjusted by varying the temperature of freezing of the CG slurry prior to sublimation; a lower temperature of freezing increases the frequency of nucleation of ice crystals in the slurry and decreases the rate of material transport within the slurry. These two processes result in the formation of smaller ice crystals, and therefore smaller pores after sublimation, for lower temperatures of freezing (3,166,167). In addition to being able to adjust the composition and the pore structure, the template residence time can be adjusted by changing the cross-link density, allowing the fabrication of scaffolds with specific degradation rates (3,105). Both chemically (i.e., glutaraldehyde or carbodiimide based cross-linking) and physically induced (i.e., dehydrothermal crosslinking) covalent bonds between collagen fibers (cross-links) can be introduced; as the scaffold becomes more highly cross-linked, it is increasingly difficult to degrade through enzymatic digestion. Adjustment of the severity of cross-linking (increasing the chemical exposure time or increasing the heat and/or exposure time to a dehydrating environment increases the severity of cross-linking) allows for adjustment of the scaffold cross-link density and therefore degradation rate (3,120). Previous research has determined that crosslinked collagen scaffolds degrade at a rate that monotonically decreases with increasing crosslink density (105). Systematic use of the contraction inhibition criterion was used to select the average pore diameter and biodegradation rate of the DRT that results in an optimal regeneration result. The speed of contraction of a wound site was measured using the wound half-life criterion, the time it took for the wound to decrease in area by 50% from its original size.

The kinetics of contraction of full-thickness skin wounds in the guinea pig were used to separate CG copolymer grafts into three classes: Class 0, I, or II (Fig. 2). The guinea pig model was employed during early studies because of the vigorous contraction observed in skin wounds; this rapid and significant contraction was used to identify templates that were active in preventing contraction and inducing regeneration (108). While the skin of most mammals is securely tethered to the underlying fascia and skin wounds

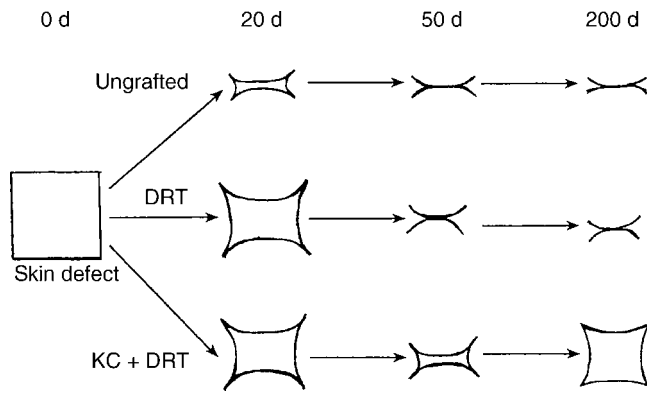


Figure 2. Schematic representation of wound healing kinetics in the adult guinea pig. A full-thickness skin wound was either ungrafted, grafted with the DRT, or grafted with a keratinocyte-seeded DRT, leading to three classes of wound healing: class 0, class I, or class II, respectively. *Top:* The ungrafted defect closed rapidly by contraction and formation of a characteristic “stellate” scar (class 0 healing). The ungrafted defect exhibited a half-life of 8 ± 1 days, and the wound contracted to $<10\%$ of the original area within 20 days. *Middle:* Grafting with a cell-free DRT led to a significant delay in wound contraction, and synthesis of a small mass of dermis and epidermis was observed. In the guinea pig wound model, however, the wound eventually closed by wound contraction (class I healing), but the resulting scar was more rounded than the “stellate scar”. The unseeded DRT exhibited a half-life of 27 ± 2 days, but the wound contracted to $<10\%$ of the original area within 40 days. *Bottom:* Grafting with a keratinocyte-seeded DRT led to delay in wound contraction, followed by the complete arrest of contraction and expansion of the defect parameter due to synthesis of new, physiologic skin (class II healing). The keratinocyte-seeded DRT exhibited a half-life of 22 ± 2 days, and as a result of synthesis of new skin the wound edges reached a steady-state condition by 200 days after grafting that was $72 \pm 5\%$ of the original wound area (1).

close by contraction, but mainly scar synthesis, the skin of the guinea pig is not nearly as well tethered and skin wounds close almost entirely by contraction with very little scar synthesis and no regeneration (Fig. 3) (1,3,133). Class 0 healing, attained by grafting either no scaffold or a biologically inactive scaffold in the guinea pig model is characterized by a wound half-life of ~ 1 week (8 ± 1 day); the wound was observed to contract to $<10\%$ of the original area within 20 days. Class I healing, attained by grafting a bioactive, cell-free scaffold into the full-thickness wound, is characterized by a significantly longer wound half-life (27 ± 2 days); however, the wound is still observed to contract to $<10\%$ of its original area after ~ 40 days. While this implant in the guinea pig still shows significant contraction of the wound site, in other animal models that more closely mimic human skin contraction kinetics, class I devices with the guinea pig prevented significant wound contraction (1). The final mode of guinea pig wound healing (class II) was observed when the wound was grafted with a bioactive scaffold seeded with cells (keratinocytes); this healing was characterized by an extended wound half life (22 ± 2 days) compared to class 0 healing. After ~ 200 days, the wound was observed to have stabilized at $\sim 72 \pm 5\%$ of the original wound area. Due to the more secure attachment of the skin to the underlying fascia with humans,

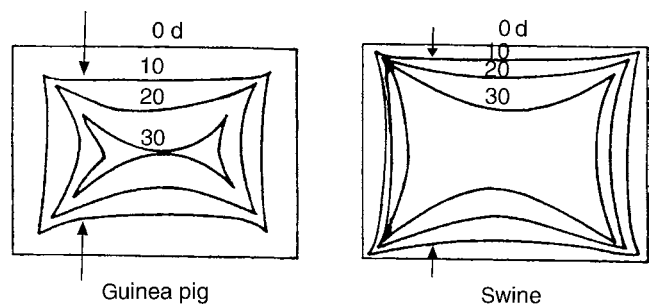


Figure 3. Kinetics of wound closure of full-thickness skin wounds (mediated primarily by contraction) in guinea pig and swine animal models. Each time point represents the change in defect area relative to the original wound dimensions on the indicated day after wounding. The guinea pig exhibits much more robust wound contraction compared to the swine model, arriving at a significantly smaller asymptotic wound area (1).

both class I and class II healing devices are useful clinically. The scaffold that led to these two healing modes was identified as having maximal biological activity, termed the DRT.

A homologous series of ECM analogs, varying in scaffold pore diameter from ~ 10 to $1000 \mu\text{m}$, was used in the above studies to determine the optimal pore diameter necessary to prevent class 0 contraction (Fig. 4). Maximum delay of wound contraction half-life, up to 27 ± 3 days, was observed when the average pore diameter was in

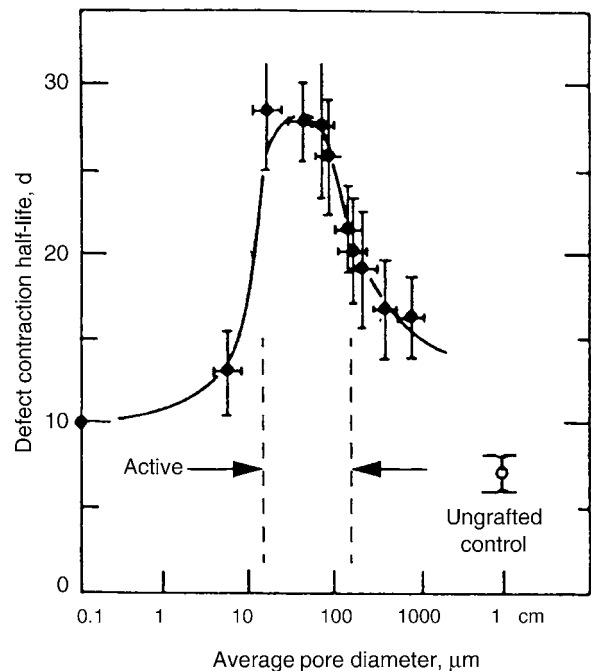


Figure 4. Identification of the optimal pore diameter for a bioactive extracellular matrix analog designed to induce skin regeneration. The range of maximal contraction-delaying activity for collagen-GAG scaffolds was observed when the average pore diameter of the scaffold was between 20 and $120 \mu\text{m}$. The limits to the area of maximal activity are indicated by broken lines and correspond to the range where contraction was most delayed (13).

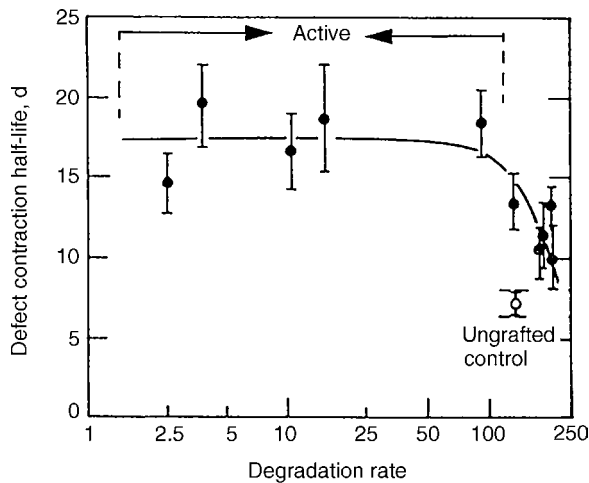


Figure 5. Identification of the optimal degradation rate for a bioactive extracellular matrix analog designed to induce skin regeneration. Contraction delaying activity of the collagen–GAG scaffold was observed when the device degradation rate was maintained between the two levels indicated in the figure. The degradation rate is expressed in units of standardized collagenase solution (*in vitro* assay). When the degradation rate was <2.5 enzyme units or >100 units, scar synthesis and contraction were observed. The limits to the area of maximal activity are indicated by broken lines, and correspond to the range where contraction was delayed the most (13).

the range between a lower limit of $20 \pm 4 \mu\text{m}$ and an upper limit of $125 \pm 35 \mu\text{m}$ (3). The lower limit in pore diameter is on the order of two cell diameters, suggesting that the scaffold pore size must be large enough to maintain space for dermal fibroblasts, the major cellular constituent of the dermis, to migrate from the wound bed into the DRT. The upper limit has been hypothesized to be as a result of an apparent requirement for attachment of a minimal number of cells onto the surface of the scaffold (167). Another homologous series of ECM analogs, where the scaffold degradation rate was varied, was also studied to match the degradation rate of the scaffold with synthesis of new physiologic tissue, as outlined in the section discussing the isomorphous tissue replacement rule (Fig. 5). A significant delay in wound contraction was observed when the degradation rate was $<115 \pm 25$ enzyme units; the enzyme units used here are arbitrary units based on an *in vitro* assay using bacterial collagenase to degrade the scaffolds and then correlating these results with the ability of each device to prevent wound contraction in the guinea pig model (3,103,105). The upper limit in degradation rate is consistent with a lower limit for the time of biodegradation of the scaffold, indicating that the scaffold needs to be present for a specific amount of time in the wound site when the contractile response is active in order to prevent contraction ($\sim 2\text{--}3$ weeks in humans) (1).

The knowledge obtained from the series of experiments used to determine the optimal chemical composition, degradation kinetics, and pore structure that each prevented wound contraction were required in order to interpret the unusual biological activity of this specific ECM analog, termed the DRT. These characteristics are appar-

ently required to block contraction for the entire period that contraction was active in the healing wound bed. Host fibroblasts were observed to migrate into the DRT from the wound edges at a speed of $\sim 0.2 \text{ mm/d}$, making the endogenous cells able to cross the entire 0.5-mm thickness of the DRT within a few days, provided the appropriate contact with the wound bed was available. With a calculated critical cell path on the order of $100 \mu\text{m}$, a distance that is filled largely with wound exudate, growth factors and serum nutrients, it was expected that fibroblasts from the host (surrounding) tissue were able to migrate readily into the DRT, multiply, and differentiate, leading to the synthesis of a physiologic dermis (1,3).

In addition to inducing the synthesis of a physiological dermis using the silicone-covered DRT, it was necessary to induce formation of a mature epithelial layer if the DRT was to be used as a permanent treatment. In the case of small wounds, keratinocytes can migrate from the wound edges across the top of the regenerating dermis. Once the keratinocytes migrate over the entire surface of the new dermis, the keratinocytes multiply and differentiate, creating a mature, stratified epidermis and basement membrane (3,133). This process was termed sequential regeneration because regeneration of the dermis had to occur prior to keratinocyte migration from the wound edges across the top of the regenerating dermis to form the neoepidermis. For larger wounds, the cells can only migrate into the edges of the wound due to their average migratory speeds of $0.2\text{--}0.5 \text{ mm/d}$, and are unable to cover the surface of the DRT rapidly enough to create a functional epidermal layer over the entire wound surface. In this case, it is necessary to supply an exogenous source of keratinocytes. Two distinct techniques have been utilized to overcome this shortcoming. The first utilizes uncultured autologous epidermal cells that are harvested using a skin biopsy from the patient and are then seeded into the DRT prior to implantation. This procedure results in the formation of a confluent epidermis after ~ 2 weeks provided that a large enough amount of cells were seeded originally ($>5 \times 10^4$ epithelial cells/ cm^2 area of DRT); this modified process has been termed simultaneous regeneration (133). A second procedural option for introducing keratinocytes to the wound can be performed when the silicone elastomer is removed (after $\sim 2\text{--}3$ weeks). When the silicone is removed, a thin autoepidermal graft is applied to the surface of the regenerating dermis (107). Approximately equivalent final results have been observed when utilizing either of these techniques for creating a mature epidermal and basement membrane layer over the regenerating dermis (1,3,133).

The quality of regeneration using the DRT was determined through immunohistochemical analysis of the regenerated tissue and comparing these results to those seen in normal skin (Fig. 6). The regenerated skin was observed to have the three tissue layers present in normal skin (tissue triad): an underlying dermis, the stratified epidermis, and a basement membrane layer between the two. For all of the following images, observations, and analyses discussed in this paragraph, the DRT treatment utilized was the DRT seeded with autologous keratinocytes prior to implantation into a full-thickness skin wound in a

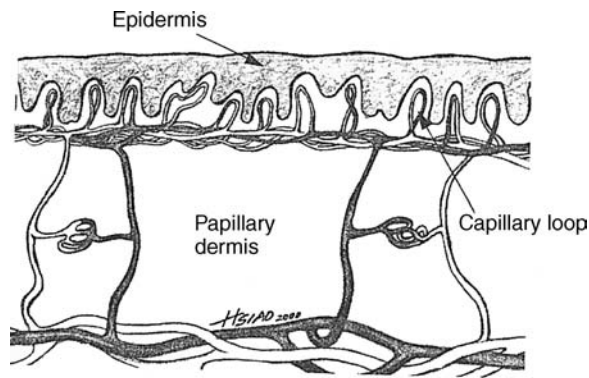


Figure 6. Diagram of normal skin showing the characteristic rete ridges at the dermo-epidermal junction and the vascular network (capillary loops) that populates the subepidermal region (1).

porcine model. As early as day 12 postgrafting, anchoring fibers were observed in the regenerated basement membrane (Fig. 7). By day 35 postgrafting, a rete ridge structure had formed, complete with vascular loops (Fig. 9), and a confluent hemidesmosomal staining pattern was observed at the dermo-epidermal junction (Fig. 8). All of these immunohistochemical results indicated that the regenerating skin was taking on all of the structural properties observed in the epidermis, dermis, and basement membrane of normal skin, except for skin appendages such as sweat glands and hair follicles (115).

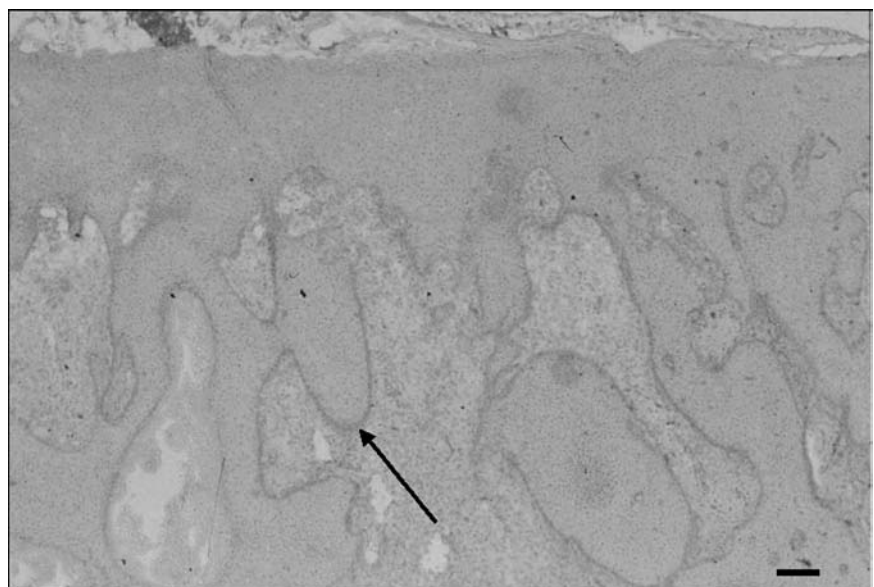
Clinical studies of the DRT initially emphasized treatment of massively burned patients while later studies have focused on patients who elected to have plastic and reconstructive surgery. In both patient populations, the primary defect is severe, traumatic skin loss down to the underlying muscle and fascia. In studies with burn patients, burns were excised down to muscle fascia prior to grafting with DRT. Wound closure was accelerated greatly by removing the silicone cover after ~2–4 weeks and grafting the newly

synthesized dermis with an ultrathin autoepidermal graft, nominally 100 μm in thickness, that was free from dermal components (107,168). In a detailed histological study of the resulting new organ the endogenous cell population degraded the DRT structure; remodeling then occurred and the newly synthesized collagen fibers became coarser and a distinction between papillary and reticular layers of the dermis appeared. Scar synthesis was not observed either at a gross or at a histological level at any time during the course of healing. In contrast to studies with the swine in which rete ridges had clearly formed (115), rete ridges were not reported in this study with humans and skin appendages were absent from the human as they were from studies with swine and rodents (169). A related immunological study showed a very small rise in immunological activity in patients' sera for the components of DRT: bovine skin collagen and chondroitin 6-sulfate. The overall conclusion from the clinical studies was that DRT presents few, if any, immunological problems to patients (170). In other clinical studies of DRT, the focus was on follow-up of massively burned pediatric patients over a 6- or 10-year period (171–173). It was reported that the new integument did not restrict joint function, suggesting the absence of crippling wound contraction and that the new skin had the ability to grow and mature long after grafting even when children were treated with the DRT (171,173). The DRT has also been used to treat patients with purpura fulminans (174), to release skin contractures (175), and to resurface scarred areas resulting from full-thickness burns (176,177); in all cases, regeneration of a functional epidermis, basement membrane, and dermis was observed.

Naturally Derived Collagen Matrices

Separate from the development of a series of scaffold materials for use to treat severe skin injuries, a separate technology has been developed to utilize naturally derived collagen matrices (NDCM) to treat skin injuries. Instead of relying on technologies to fabricate a 3D scaffold structure

Figure 7. As early as 12 days after grafting a full-thickness skin wound with a keratinocyte-seeded DRT, anchoring fibrils were observed in the regenerating basement membrane (arrow). The basal surface epithelium and the periphery of the epithelial cords are labeled with type VII collagen immunostaining, identifying the anchorage structures at the dermo-epidermal interface. Bar: 150 μm . Reprinted with permission from Reference (115).



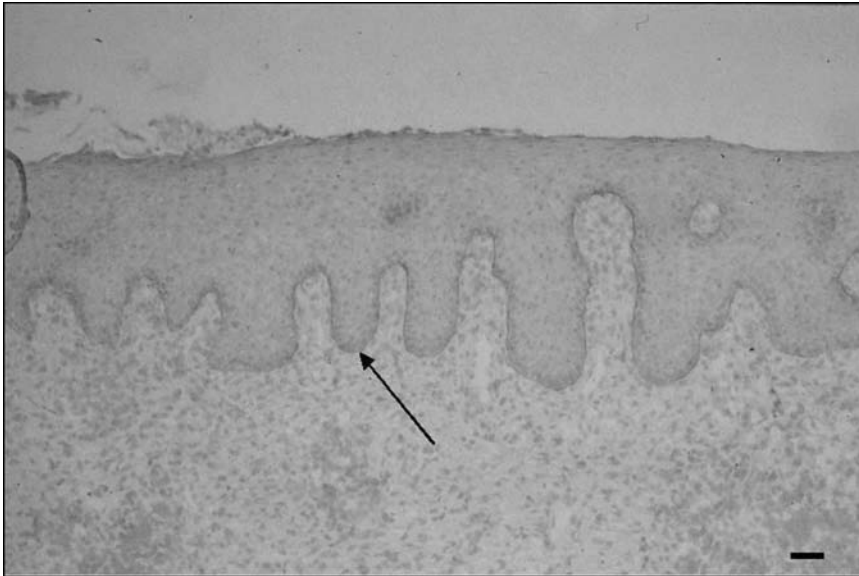


Figure 8. As early as 35 days after grafting a full-thickness skin wound with a keratinocyte-seeded DRT, a confluent hemidesmosomal staining pattern is observed at the dermo-epidermal junction (arrow) by immunostaining for the $\alpha_6\beta_4$ integrin. The pattern observed in the regenerating skin is identical to that observed in physiologic skin. Bar: 100 μm . Reprinted with permission from Reference (115).

from either synthetic or natural materials (i.e., CEA, LSE, LDR, or DRT), this technology uses decellularized dermal tissue as a scaffold structure. The most analogous treatment to the NDCM that has already been discussed in this article is the allograft and xenograft; the NDCM is a decellularized version of an allograft. The NDCM is designed to serve as a scaffold to support normal tissue remodeling following severe injury, thereby inducing regeneration.

The NDCM is produced through a three-step process. The epidermal tissue is completely removed from full-thickness autograft tissue, leaving both the dermal tissue and the basement membrane. The dermal cells are then removed using detergent washes. The decellularized tissue is then freeze-dried to preserve the NDCM structure and to maintain the bioactivity of the dermal matrix. The main advantage of the NDCM over the homograft and xenograft are that owing to decellularization, the antigenicity of the scaffold is significantly reduced. Reducing the antigenicity

reduces the immunological response of the patient to the graft and reduces the chance of implant rejection. To treat a severe skin wound, the NDCM can be rehydrated in saline solution then be implanted directly into a wound site in the same manner as a tissue autograft. The NDCM has been used primarily to treat full-thickness burns and burns to areas of the skin where contraction and scar formation would inhibit functionality (i.e., feet and hands). AlloDerm, a product of the LifeCell Corp., is a common NDCM available for experimental and clinical trials.

The NDCM is implanted into full-thickness skin wounds and is often covered by a thin autograft of the patient's own epidermis to speed the healing process. This treatment typically results in a high percentage of graft take; additionally, the thin autograft of epidermal tissue significantly decreases the time for complete reepithelialization of the graft and reduces the number of complications due to infection. Patients showed normal range of motion, grip strength, motor control, and functionality following

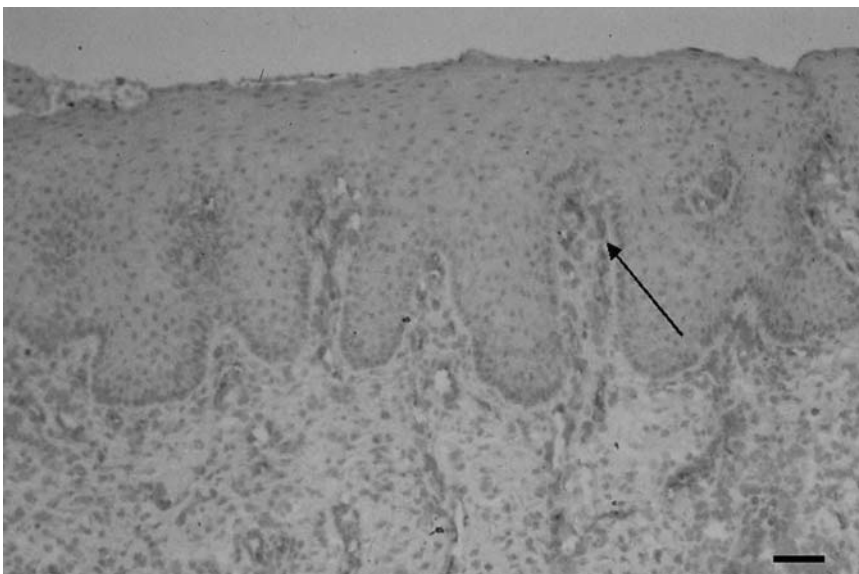


Figure 9. A full-thickness skin wound grafted with a keratinocyte-seeded DRT was observed to regenerate many of the structure observed in normal skin. Immunostaining for Factor VIII 35 days after grafting revealed that capillary loops had formed in the rete ridges of the regenerated dermis (arrow) similar to those observed in physiologic skin. Bar: 75 μm . Reproduced with permission from Reference (115).

treatment (178,179). The NDCM technology demonstrates the use of a naturally derived material in tissue engineering applications. A major restriction on the use of an NDCM such as Alloderm is the same as that faced by allografting: a limited supply of donor tissue available for grafting.

DISCUSSION AND CONCLUSIONS

Future improvements in treating injuries to various tissues and organs, including skin, via tissue engineering protocols will eventually depend critically on theoretical developments that explain the mechanism of tissue and organ replacement, thereby suggesting novel experimental studies and new paradigms for inducing regeneration. The goal of this article was to discuss the process of wound healing following injury, specifically the response of the skin to severe injuries and to discuss a number of paradigms that have been investigated in an attempt to alter the organismic (contraction-mediated) response to severe injuries. Evidence has been introduced to support the conclusion that contraction is the main engine for wound closure both in skin and peripheral nerve wounds, that scar formation is a byproduct of contraction, and that there is an antagonistic relationship between contraction and regeneration. Regeneration of skin and peripheral nerve requires selective blocking of contraction, but not blocking of other aspects of the healing process. The structural requirements for bioactive regeneration templates were also discussed in the light of this theory, and a number of distinct paradigms using very different scaffold structures were studied to determine the experimental and clinical outcomes of their use. This article concludes with a detailed discussion of the mechanism of regeneration after implantation of the DRT into severe wound sites.

The empirical evidence described here is consistent with the conclusion that synthesis of an organ *in vivo* (induced regeneration) requires grafting of an appropriate scaffold that has or has not been seeded with epithelial cells from the desired organ. In this approach, there is no requirement for addition of exogenous reactants such as mesenchymal cells (i.e., fibroblasts) or for addition of cytokines. These empirical findings appear to have direct relevance for the future selection of tissue engineering protocol based on both *in vitro* and *in vivo* environments.

ACKNOWLEDGMENTS

The authors acknowledge partial support from the NIH/NIBIB, grant DE13053, as well as a grant from the Cambridge University-MIT Institute.

BIBLIOGRAPHY

1. Yannas IV. Tissue and Organ Regeneration in Adults. New York: Springer-Verlag; 2001.
2. Yannas IV. In vitro synthesis of tissues and organs. In: Lanza RP, Langer R, Vacanti J, editors. Principles of Tissue Engineering. 2nd ed. San Diego: Academic Press; 2000.

3. Yannas IV, Lee E, Orgill DP, Skrabut EM, Murphy GF. Synthesis and characterization of a model extracellular matrix that induces partial regeneration of adult mammalian skin. Proc Natl Acad Sci USA 1989;86:933–937.
4. Mast BA, Diegelmann RF, Krummel TM, Cohen IK. Scarless wound healing in the mammalian fetus. Surg Gynecol Obstet 1992;174:441–451.
5. Mast BA, Nelson JM, Krummel TM. Tissue repair in the mammalian fetus. In: Cohen IK, Diegelmann RF, Lindblad WJ, editors. Wound Healing. Philadelphia: W.B. Saunders; 1992.
6. Wallace H. Vertebrate Limb Regeneration. New York: Wiley; 1981.
7. Yannas IV, Colt J, Wai YC. Wound contraction and scar synthesis during development of the amphibian *Rana catesbeiana*. Wound Rep Reg 1996;4:31–41.
8. Martinez-Hernandez A. Repair, regeneration, and fibrosis. In: Rubin E, Farber JL, editors. Pathology. Philadelphia: J.B. Lippincott; 1988.
9. Uitto J, Mauviel A, McGrath J. The dermal-epidermal basement membrane zone in cutaneous wound healing. In: Clark RAF, editor. The Molecular and Cellular Biology of Wound Repair. New York: Plenum Press; 1996.
10. Fu SY, Gordon T. The cellular and molecular basis of peripheral nerve regeneration. Mol Neurobiol 1997;14:67–116.
11. Haber RM, Hanna W, Ramsay CA, Boxall LB. Cicatricial junctional epidermolysis bullosa. J Am Acad Dermatol 1985;12:836–844.
12. Ikeda K, Oda Y, Tomita K, Nomura S, Nakanishi I. Isolated Schwann cells can synthesize the basement membrane *in vitro*. J Electron Microsc (Tokyo) 1989;38:230–234.
13. Stenn KS, Malhotra R. Epithelialization. In: Cohen IK, Diegelmann RF, Lindblad WJ, editors. Wound Healing. Philadelphia: W. B. Saunders; 1992.
14. Ferdman AG, Yannas IV. Scattering of light from histologic section: A new method for the analysis of connective tissue. J Invest Dermatol 1993;100:710–716.
15. Lehv M, Fitzgerald PJ. Pancreatic acinar cell regeneration IV: Regeneration after surgical resection. Am J Pathol 1968;53:513–535.
16. Oliver J. Correlations of structure and function and mechanisms of recovery in acute tubular necrosis. Am J Med 1953;15:535–557.
17. Stemerman MB, Spaet TH, Pitlick F, Cintron J, Lejniaks I, Tiell ML. Intimal healing. The patterns of reendothelialization and intimal thickening. Am J Pathol 1977;87:125–142.
18. Vracko R. Significance of basal lamina for regeneration of injured lung. Virchows Arch (Pathol Anat) 1972;355:264–274.
19. Yannas IV. Regeneration of skin and nerves by use of collagen templates. In: Nimni N, editor. Collagen: Biotechnology. Boca Raton: CRC Press; 1988.
20. Lanza RP, Langer R, Chick WL, editors. Principles of Tissue Engineering. San Diego: Academic Press; 1997.
21. Brown H. Wound healing research through the ages. In: Cohen IK, Diegelmann RF, Lindblad WJ, editors. Wound Healing. Philadelphia: W.B. Saunders; 1992.
22. Bach FH, Robson SC, Winker H, Ferran C, Stuhlmeiser KM, Wrighton CJ, Hancock WW. Barriers to xenotransplantation. Nature Med 1995;1:869–873.
23. Cooper DKC, Kemp E, Platt JL, White DJG, editors. Xenotransplantation. New York: Springer-Verlag; 1997.
24. Lanza RP, Chick WL. Endocrinology: Pancreas. Lanza RP, Langer R, Chick WL, editors. Principles of Tissue Engineering. San Diego: Academic Press; 1997.

25. Medawar PB. The behavior and fate of skin autografts and skin homografts in rabbits. *J Anat* 1944;78:176–199.
26. Medawar PB. The storage of living skin. *Proc R Soc Med* 1954;47:62–64.
27. Murray JE, Merrill JP, Harrison JH. Renal homotransplantations in identical twins. *Surg For* 1955;6:432–436.
28. Wickelgren I. Muscling transplants in mice. *Science* 1996;273:33.
29. Lau HT, Fontana A, Stoeckert CJ. Prevention of islet allograft rejection with engineered myoblasts expressing FasL in mice. *Science* 1996;273:109–112.
30. Avgoustiniatos ES, Colton CK. Design considerations for immunoisolation. In: Lanza RP, Langer R, Chick WL, editors. *Principles of Tissue Engineering*. San Diego: Academic Press; 1997.
31. Lim F, Sun AM. Microencapsulated islets as bioartificial endocrine pancreas. *Science* 1980;210:908–910.
32. Lanza RP, Cooper DKC, Chick WL. Xenotransplantation. *Sci Am* 1997;July: 54–59.
33. Kaiser J. IOM backs cautious experimentation. *Science* 1996;273:305–306.
34. Patience C, Takeuchi Y, Weiss RA. Infection of human cells by an endogenous retrovirus of pigs. *Nature Med* 1997;3:282–286.
35. Sikorski R, Peters R. Xenotransplanters turn xenovirologists. *Science* 1997;276:1893.
36. Burke JF, Bondoc CC, Quinby WC. Primary burn excision and immediate grafting: A method of shortening illness. *J Trauma* 1974;14:389–395.
37. Millesi H. Erfahrungen mit der Mikrochirurgie peripherer Nerven. *Chir Plast Reconstr* 1967;3:47–55.
38. Sunderland S. *Nerves and Nerve Injuries*. New York: Churchill Livingstone; 1978.
39. Terzis JK. *Microreconstruction of Nerve Injuries*. Philadelphia: W.B. Saunders; 1987.
40. Millesi H, Meissl G, Berger G. The interfascicular nerve grafting of the median and ulnar nerves. *J Bone Joint Surg* 1972;54-A:727–750.
41. Millesi H, Meissl G, Berger G. Further experience with interfascicular grafting of the median, ulnar, radial nerves. *J Bone Joint Surg* 1976;58-A:209–216.
42. Grondin CM, Campeau I, Thornton JC, Engle JC, Cross FS, Schreiber H. Coronary artery bypass grafting with saphenous vein. *Circ* 1989;79(I): 24–29.
43. Kohn DH, Ducheyne P. Materials for bone and joint replacement. In: Williams DF, editor. *Materials Science and Technology*. New York: VCH Publishers; 1992.
44. Neuman MR. Therapeutic and prosthetic devices. In: Webster JG, editor. *Medical Instrumentation*. 4th ed. New York: Wiley; 1998.
45. Schoen FJ, Hobson CE. Anatomic analysis of removed prosthetic heart valves: Causes of failure of 33 mechanical valves and 58 bioprostheses, 1980 to 1983. *Hum Pathol* 1985;16(6): 549–559.
46. Caldarone CA, McCrindle BW, Van Arsdell GS, Coles JG, Webb G, Freedom RM, Williams WG. Independent factors associated with longevity of prosthetic pulmonary valves and valved conduits. *J Thorac Cardiovasc Surg* 2000;120(6): 1022–1030.
47. Facer GW, Peterson A, Brey RH, Marion M, Cevette M, Balko K, Green JD, Rose D, Pool A. The mayo clinic experience with the cochlear implant. *Ear Nose Throat J* 1994;73(3): 149–152.
48. Peppas NA, Langer R. New challenges in biomaterials. *Science* 1994;263:1715–1720.
49. Ginsbach G, Busch LC, Kuhnel W. The nature of the collagenous capsules around breast implants. *Plast Reconstr Surg* 1979;64:456–464.
50. Rudolph R, Van de Berg J, Ehrlich P. Wound contraction and scar contracture. In: Cohen IK, Diegelmann RF, Lindblad WJ, editors. *Wound Healing*. Philadelphia: W.B. Saunders; 1992.
51. Spector M, Heyligers I, Roberson JR. Porous polymers for biological fixation. *Clin Orth* 1993;235:207–219.
52. Chandy T, Das GS, Wilson RF, Rao GH. Surface-immobilized biomolecules on albumin modified porcine pericardium for preventing thrombosis and calcification. *Int J Artif Organs* 1999;22:547–558.
53. Park JY, Davies JE. Red blood cell and platelet interactions with titanium implant surfaces. *Clin Oral Implants Res* 2000;11:530–539.
54. Snyder TA, Watach MJ, Litwak KN, Wagner WR. Platelet activation, aggregation, life span in calves implanted with axial flow ventricular assist devices. *Ann Thorac Surg* 2002;73:1933–1938.
55. Suggs LJ, West JL, Mikos AG. Platelet adhesion on a bioresorbable poly(propylene fumarate-co-ethylene glycol) copolymer. *Biomaterials* 1999;20:683–690.
56. Basle MF, Bertrand G, Guyetant S, Chappard D, Leonard M. Migration of metal and polyethylene particles from articular prostheses may generate lymphadenopathy with histiocytosis. *J Biomed Mater Res* 1996;30(2): 157–163.
57. Urban RM, Jacobs JJ, Tomlinson MJ, Gavrilovic J, Black J, Peoc'h M. Dissemination of wear particles to the liver, spleen, and abdominal lymph nodes of patients with hip and knee replacement. *J Bone Joint Surg Am* 2000;82(4): 457–476.
58. Willert HG. Reactions of the articular capsule to wear products of artificial joint prostheses. *J Biomed Mater Res* 1977;11(2): 157–164.
59. Prockop DJ. Marrow stromal cells as stem cells for nonhematopoietic tissues. *Science* 1997;276:71–74.
60. Solter D, Gearhart J. Putting stem cells to work. *Science* 1999;283:1468–1470.
61. Pittenger MF, Mackay AM, Berk SC, Jaiswal RK, Douglas R, Mosca JD, Morman MA, Simonetti DW, Craig S, Marshak DR. Multilineage potential of adult human mesenchymal stem cells. *Science* 1999;284:143–147.
62. Slack JMW. Stem cells in epithelial tissues. *Science* 2000; 287:1431–1433.
63. Gage FH. Mammalian neural stem cells. *Science* 2000; 287:1433–1438.
64. Yannas IV. Synthesis of organs: In vitro or in vivo? *Proc Natl Acad Sci USA* 2000;97:9354–9356.
65. Compton CC. Keratinocyte grafting models. In: Lane EB, Leigh I, Watt F, editors. *The Keratinocyte Handbook*. London: Cambridge University Press; 1994.
66. Karasek MA. In vitro culture of human skin epithelial cells. *J Invest Dermatol* 1966;47:533–540.
67. Karasek MA. Growth and differentiation of transplanted epithelial cell cultures. *J Invest Dermatol* 1968;51: 247–252.
68. Worst PKM, Valentine EA, Fusenig NE. Formation of epidermis after reimplantation of pure primary epidermal cell cultures from perinatal mouse skin. *J Natl Cancer Inst* 1974;53:1061–1064.
69. Compton CC, Gill JM, Bradford DA, Regauer S, Gallico GG, O'Connor NE. Skin regenerated from cultured epithelial autografts on full-thickness burn wounds from 6 days to 5 years after grafting. *Lab Invest* 1989;60:600–612.
70. Eldad A, Burt A, Clarke JA, Gusterson B. Cultured epithelium as a skin substitute. *Burns* 1987;13:173–180.

71. Gallico GG, O'Connor NE, Compton CC, Kehinde O, Green H. Permanent coverage of large burn wounds with autologous cultured human epithelium. *New Engl J Med* 1984;311:448-451.
72. Green H, Kehinde O, Thomas J. Growth of cultured human epidermal cells into multiple epithelia suitable for grafting. *Proc Natl Acad Sci USA* 1979;76:5665-5668.
73. Green H, Rheinwald JG. Process for serially culturing keratinocytes, US patent 4,016,036, 1977.
74. Munster AM. Use of cultured epithelial autograft in ten patients. *J Burn Care Rehab* 1992;13:124-126.
75. Munster AM. Cultured skin for massive burns. *Ann Surg* 1996;224:372-377.
76. Rheinwald JG, Green H. Formation of a keratinizing epithelium in culture by a cloned cell line derived from a tetroma. *Cell* 1975;6:317-330.
77. Rheinwald JG, Green H. Serial cultivation of strains of human epidermal keratinocytes: The formation of keratinizing colonies from single cells. *Cell* 1975;6:331-343.
78. Bell E, Ehrlich HP, Buttle DJ, Nakatsuji T. Living skin formed in vitro and accepted as skin-equivalent tissue of full thickness. *Science* 1981;211:1052-1054.
79. Bell E, Ehrlich HP, Sher S, Merrill C, Sarber R, Hull B, Nakatsuji T, Church D, Buttle DJ. Development and use of a living skin equivalent. *Plast Reconstr Surg* 1981;67:386-392.
80. Bell E, Ivarsson B, Merrill C. Production of a tissue-like structure by contraction of collagen lattices by human fibroblasts of different proliferative potential in vitro. *Proc Natl Acad Sci USA* 1979;76:1274-1278.
81. Bell E, Sher S, Hull B. The living skin equivalent as a structural and immunological model in skin grafting. *Scan Electr Micr* 1984;4:1957-1962.
82. Bell E, Sher S, Hull B, Merrill C, Rosen S, Chamson A, Asselineau D, Dubertret L, Coulomb B, Lapiere C. The reconstitution of living skin. *J Invest Dermatol* 1983;81:2s-10s.
83. Drumheller PD, Hubbell JA. Surface immobilization of adhesion ligands for investigations of cell-substrate interactions. In: Bronzio JD, editor. *The Biomedical Engineering Handbook*. Boca Raton: CRC Press; 1997.
84. Griffith Cima L. Polymeric biomaterials. *Acta Mater* 1994;48:263-277.
85. Langer R, Vacanti JP. Tissue engineering. *Science* 1993;260:920-928.
86. Lanza RP, Langer R, Vacanti J, editors. *Principles of Tissue Engineering*. 2nd ed. San Diego: Academic Press; 2000.
87. Cooper ML, Hansbrough JF. Use of a composite skin graft composed of cultured human keratinocytes and fibroblasts and a collagen-GAG matrix to cover full-thickness wounds on athymic mice. *Surgery* 1991;109:198-207.
88. Cooper ML, Hansbrough JF, Spielvogel RL, Cohen R, Bartel RL, Naughton G. In vivo optimization of a living dermal substitute employing cultured human fibroblasts on a biodegradable polyglycolic acid or polyglactin mesh. *Biomaterials* 1991;12:243-248.
89. Dore C, Noordenbos J, Hansbrough JF. Management of partial thickness burns with Dermagraft-TC. *J Burn Care Rehabil* 1998;19:S172.
90. Hansbrough JF, Cooper ML, Cohen R, Spielvogel R, Greenleaf G, Bartel RL, Naughton G. Evaluation of biodegradable matrix containing cultured human fibroblasts as a dermal replacement beneath meshed skin grafts on athymic mice. *Surgery* 1992;111:438-446.
91. Hansbrough JF, Dore C, Hansbrough WB. Clinical trials of a living dermal tissue replacement placed beneath meshed, split-thickness skin grafts on excised burn wounds. *J Burn Care Rehab* 1992;13:519-529.
92. Hansbrough JF, Morgan JL, Greenleaf GE, Bartel R. Composite grafts of human keratinocytes grown on a polyglactin mesh-cultured fibroblast dermal substitute function as a bilayer skin replacement in full-thickness wounds on athymic mice. *J Burn Care Rehab* 1993;14:485-494.
93. Naughton G, Mansbridge J, Gentzkow G. A metabolically active human dermal replacement for the treatment of diabetic foot ulcers. *Artif Org* 1997;21:1-7.
94. Purdue GF, Hunt JL, Still JM, Law EJ, Herndon DN, Goldfarb IW, Schiller WR, Hansbrough JF, Hickerson WL, Himel HN. A multicenter clinical trial of a biosynthetic skin replacement, Dermagraft-TC, compared with cryopreserved human cadaver skin for temporary coverage of excised burn wounds. *J Burn Care Rehab* 1997;18:52-57.
95. Park A, Wu B, Griffith LG. Integration of surface modification and 3D fabrication techniques to prepare patterned poly(L-lactide) substrates allowing regionally selective cell adhesion. *J Biomater Sci Polym Ed* 1998;9(2): 89-110.
96. Powers MJ, Domansky K, Kaazempur-Mofrad MR, Kalezi A, Capitano A, Upadhyaya A, Kurzawski P, Wack KE, Stolz DB, Kamm R, Griffith LG. A microfabricated array bioreactor for perfused 3D liver culture. *Biotechnol Bioeng* 2002;78(3): 257-269.
97. Xu J, Clark RAF. Integrin regulation in wound repair. In: Garg HG, Longaker MT, editors. *Scarless Wound Healing*. New York: Marcel Dekker; 2000.
98. Freed LE, Vunjak-Novakovic G. Cultivation of cell-polymer tissue constructs in simulated microgravity. *Biotechnol Bioeng* 1995;46:306-313.
99. Freed LE, Vunjak-Novakovic G, Briton RJ, Eagles DB, Lesnoy DC, Barlow SK, Langer R. Biodegradable polymer scaffolds for tissue engineering. *Biotechnology* 1994;12:689-693.
100. Lee CR, Breinan HA, Nehrer S, Spector M. Articular cartilage chondrocytes in type I and type II collagen-GAG matrices exhibit contractile behavior in vitro. *Tissue Eng* 2000;6(5): 555-565.
101. Nehrer S, Breinan HA, Ramappa A, Shortkroff S, Young G, Minas T, Sledge CB, Yannas IV, Spector M. Canine chondrocytes seeded in type I and type II collagen implants investigated in vitro. *J Biomed Mater Res* 1997;38(2): 95-104.
102. Dagalakis N, Flink J, Stasikelis P, Burke JF, Yannas IV. Design of an artificial skin. Part III. Control of pore structure. *J Biomed Mater Res* 1980;14:511-528.
103. Yannas IV, Burke JF, Huang C, Gordon PL. Suppression of in vivo degradability and of immunogenicity of collagen by reaction with glycosaminoglycans. *Polym Prepr Am Chem Soc* 1975;16(2): 209-214.
104. Yannas IV, Burke JF, Gordon PL, Huang C, Rubinstein RH. Design of an artificial skin II: Control of chemical composition. *J Biomed Mater Res* 1980;14:107-131.
105. Yannas IV, Burke JF, Huang C, Gordon PL. Correlation of in vivo collagen degradation rate with in vitro measurements. *J Biomed Mater Res* 1975;6:623-625.
106. Yannas IV, Burke JF, Umbreit M, Stasikelis P. Progress in design of an artificial skin. *Fed Proc Am Soc Exp Biol* 1979;38:988.
107. Burke JF, Yannas IV, Quincy WC, Bondoc CC, Jung WK. Successful use of a physiologically acceptable artificial skin in the treatment of extensive burn injury. *Ann Surg* 1981;194:413-428.
108. Yannas IV. Use of artificial skin in wound management. In: Dineen P editor. *The Surgical Wound*. Philadelphia: Lea & Febiger; 1981. p 171-190.
109. Yannas IV, Burke JF. Design of an artificial skin I. Basic design principles. *J Biomed Mater Res* 1980;14:65-81.

110. Yannas IV, Burke JF, Huang C, Gordon PL. Multilayer membrane useful as synthetic skin. US patent 4,060,081. 1977.
111. Yannas IV, Burke JF, Warpehoski M, Stasikelis P, Skrabut EM, Orgill DP, Giard DJ. Prompt, long-term functional replacement of skin. *Trans Am Soc Artif Intern Organs* 1981;27:19–22.
112. Yannas IV, Burke JF, Orgill DP, Skrabut EM. Wound tissue can utilize a polymeric template to synthesize a functional extension of skin. *Science* 1982;215:174–176.
113. Yannas IV, Orgill DP, Skrabut EM, Burke JF. Skin regeneration with a bioreplaceable polymeric template. In: Gebelein CC, editor. *Polymeric Materials and Artificial Organs*. Washington, D.C.: American Chemical Society; 1984. p 191–197.
114. Butler CE, Yannas IV, Compton CC, Correia CA, Orgill DP. Comparison of cultured and uncultured keratinocytes seeded into a collagen-GAG matrix for skin replacements. *Br J Plast Surg* 1999;52:127–132.
115. Compton CC, Butler CE, Yannas IV, Warland G, Orgill DP. Organized skin structure is regenerated in vivo from collagen-GAG matrices seeded with autologous keratinocytes. *J Invest Dermatol* 1998;110:908–916.
116. Murphy GF, Orgill DP, Yannas IV. Partial dermal regeneration is induced by biodegradable collagen-glycosaminoglycan grafts. *Lab Invest* 1990;62:305–313.
117. Chang AS, Yannas IV. Peripheral nerve regeneration. In: Smith B, Adelman G, editors. *Encyclopedia of Neuroscience*. Boston: Birkhauser; 1992. p 125–126.
118. Chang AS, Yannas IV, Perutz S, Loree H, Sethi RR, Krarup C, Norregaard TV, Zervas NT, Silver J. Electrophysiological study of recovery of peripheral nerves regenerated by a collagen-glycosaminoglycan copolymer matrix. In: Gebelein CC, Dunn RL, editors. *Progress in Biomedical Polymers*. New York: Plenum Press; 1990. p 107–119.
119. Yannas IV, Orgill DP, Silver J, Norregaard TV, Zervas NT, Schoene WC. Regeneration of sciatic nerve across 15 mm gap by use of a polymeric template. In: Gebelein CC, editor. *Advances in Biomedical Polymers*. New York: Plenum Publishing Corporation; 1987. p 1–9.
120. Harley BA, Spilker MH, Wu JW, Asano K, Hsu H-P, Spector M, Yannas IV. Optimal degradation rate for collagen chambers used for regeneration of peripheral nerves over long gaps. *Cells Tissues Organs* 2004;176:153–165.
121. Chamberlain LJ, Yannas IV, Hsu H-P, Strichartz G, Spector M. Collagen-GAG substrate enhances the quality of nerve regeneration through collagen tubes up to level of autograft. *Exper Neurol* 1998;154:315–329.
122. Chamberlain LJ, Yannas IV, Hsu H-P, Strichartz GR, Spector M. Near terminus axonal structure and function following rat sciatic nerve regeneration through a collagen-GAG matrix in a 10-mm gap. *J Neurosci Res* 2000;60:666–677.
123. Hsu WC, Spilker MH, Yannas IV, Rubin PA. Inhibition of conjunctival scarring and contraction by a porous collagen-glycosaminoglycan implant. *Invest. Ophthalmol Vis Sci* 2000;41:2404–2411.
124. Billingham RE, Medawar PB. The technique of free skin grafting in mammals. *J Exp Biol* 1951;28:385–394.
125. Billingham RE, Medawar PB. Contracture and intussusceptive growth in the healing of expensive wounds in mammalian skin. *J Anat* 1955;89:114–123.
126. Shapiro F. Cortical bone repair. *J Bone Joint Surg* 1988;70-A:1067–1081.
127. Lundborg G. Nerve regeneration and repair. *Acta Orthop Scand* 1987;58:145–169.
128. Lundborg G, Dahlin LB, Danielsen N, Gelberman RH, Longo FM, Powell HC, Varon S. Nerve regeneration in silicone chambers: Influence of gap length and of distal stump components. *Exp Neurol* 1982;76:361–375.
129. Lundborg G, Dahlin LB, Danielsen N, Johannesson A, Hansson HA, Longo F, Varon S. Nerve regeneration across an extended gap: A neurobiological view of nerve repair and the possible involvement of neuronotrophic factors. *J Hand Surg* 1982;7:580–587.
130. Lundborg G, Gelberman RH, Longo FM, Powell HC, Varon S. In vivo regeneration of cut nerves encased in silicone tubes: Growth across a six-millimeter gap. *J Neuropathol Exp Neurol* 1982;41:412–422.
131. Lundborg G, Longo FM, Varon S. Nerve regeneration model and trophic factors in vivo. *Brain Res* 1982;232:157–161.
132. Peacock EE. Wound healing and wound care. In: Schwartz SI, Shires GT, Spencer FC, Storer EH, editors. *Principles of Surgery*. New York: McGraw-Hill; 1984.
133. Yannas IV. Artificial skin and dermal equivalents. In: Bronzio JD, editor. *The Biomedical Engineering Handbook*. Boca Raton: CRC Press; 2000. p 138–1–138-15.
134. Yannas IV. Regeneration Templates. In: Bronzio JD, editor. *The Biomedical Engineering Handbook*. Boca Raton: CRC Press; 2000. p 113–1–113-18.
135. Yannas IV, Tobolsky AV. Crosslinking of gelatine by dehydration. *Nature (London)* 1967;215:509–510.
136. Lee JE, Park JC, Hwang YS, Kim JK, Kim JG, Sub H. Characterization of UV-irradiated dense/porous collagen membranes: morphology, enzymatic degradation, and mechanical properties. *Yonsei Med J* 2001;42(2): 172–179.
137. Weadock KS, Miller EJ, Bellincampi LD, Zawadsky JP, Dunn MG. Physical crosslinking of collagen fibers: Comparison of ultraviolet irradiation and dehydrothermal treatment. *J Biomed Mater Res* 1995;29(11): 1373–1379.
138. Weadock KS, Miller EJ, Keuffel EL, Dunn MG. Effect of physical crosslinking methods on collagen-fiber durability in proteolytic solutions. *J Biomed Mater Res* 1996;32(2): 221–226.
139. Lee CR, Grodzinsky AJ, Spector M. The effects of crosslinking of collagen-glycosaminoglycan scaffolds on compressive stiffness, chondrocyte-mediated contraction, proliferation, biosynthesis. *Biomaterials* 2001;22:3145–3154.
140. Gibson LJ, Ashby MF. *Cellular solids: Structure and properties*. Cambridge, U.K.: Cambridge University Press; 1997.
141. Yannas IV. Studies on the biological activity of the dermal regeneration template. *Wound Rep Reg* 1998;6:518–524.
142. Burkitt HG, Young B, Heath JW. *Wheater's Functional Histology*. 3rd ed. Edinburgh: Churchill Livingstone; 1993.
143. Farquhar MG. The glomerular basement membrane: A selective macromolecular filter. In: Hay ED, editor. *Cell Biology of Extracellular Matrix*. New York: 1981. Plenum Press; p 335–378.
144. Vracko R. Basal lamina scaffold: Anatomy and significance for maintenance of orderly tissue structure. *Am J Pathol* 1974;77:313–346.
145. Woodley DT, Briggaman RA. Re-formation of the epidermal-dermal junction during wound healing. In: Clark RAF, Henson PM, editors. *The Molecular and Cellular Biology of Wound Repair*. New York: Plenum Press; 1988.
146. Hay ED. Collagen and embryonic development. In: Hay ED, editor. *Cell Biology of the Extracellular Matrix*. New York: Plenum Press; 1981.
147. Briggaman RA, Wheeler CE. The epidermal-dermal junction. *J Invest Dermatol* 1975;65:71–84.
148. Carver N, Navsaria HA, Fryer P, Green CJ, Leigh IM. Restoration of basement membrane structure in pigs following keratinocyte autografting. *J Plast Surg* 1993;46:384–392.

149. Rigal C, Pieraggi M-T, Vincent C, Prost C, Bouissou H, Serre G. Healing of full-thickness cutaneous wounds in the pig I: Immunohistochemical study of epidermodermal junction regeneration. *J Invest Dermatol* 1991;96:777-785.
150. Gallico GG. Biological skin substitutes. *Clin Plast Surg* 1990;17:519-526.
151. Eisinger M, Lee JS, Hefton JM, Darzynkiewicz Z, Chiao JW, DeHarven E. Human epidermal cell cultures: Growth and differentiation in the absence of dermal components or medium supplements. *Proc Natl Acad Sci USA* 1979;76:5340-5344.
152. Freeman AE, Igel HJ, Herrman BJ, Kleinfeld KL. Growth and characterization of human skin epithelial cultures. *In Vitro* 1976;12:352-362.
153. Lillie JH, MacCallum DK, Jepsen A. Growth of stratified squamous epithelium on reconstituted extracellular matrices: Long-term culture. *J Invest Dermatol* 1988;90:100-109.
154. Woodley DT, Peterson HD, Herzog SR, Stricklin GP, Burgeson RE, Briggaman RA, Cronce DJ, O'Keefe EJ. Burn wounds resurfaced by cultured epidermal autografts show abnormal reconstitution of anchoring fibrils. *JAMA* 1988;259:2566-2571.
155. Hull BE, Sher SE, Rosen S, Church D, Bell E. Structural integration of skin equivalents grafted to Lewis and Sprague-Dawley rats. *J Invest Dermatol* 1983;81:429-436.
156. Cumpstone MB, Kennedy AH, Harmon CS, Potts RO. The water permeability of primary mouse keratinocyte cultures grown at the air-liquid interface. *J Invest Dermatol* 1989;92:598-600.
157. Nolte CJM, Oleson MA, Bilbo PR, Parenteau NL. Development of a stratum corneum and barrier function in an organotypic skin culture. *Arch Dermatol Res* 1993;285:466-474.
158. Nolte CJM, Oleson MA, Hansbrough JF, Morgan J, Greenleaf G, Wilkins L. Ultrastructural features of composite skin cultures grafted onto athymic mice. *J Anat* 1994;185:325-333.
159. Bosca AR, Tinois E, Faure M, Kanitakis J, Roche P, Thivolet J. Epithelial differentiation of human skin equivalents after grafting onto nude mice. *J Invest Dermatol* 1988;91:136-141.
160. Wassermann D, Slotterer M, Toulon A, Cazalet C, Marien M, Cherruau B, Jaffray P. Preliminary clinical studies of a biological skin equivalent in burned patients. *Burns* 1988;14:326-330.
161. Eaglstein WH, Iriondo M, Laszlo K. A composite skin substitute (Graft-skin) for surgical wounds. *Dermatol Surg* 1995;21:839-843.
162. Falanga V, Margolis D, Alvarez O, Auletta M, Maggiamco F, Altman M, Jensen J, Saboinski M, Hardin-Young J. Rapid healing of venous ulcers and lack of clinical rejection with an allogeneic cultured human skin equivalent. *Arch Dermatol* 1998;134:293-300.
163. Sabolinski ML, Alvarez O, Auletta M, Mulder G, Parenteau NL. Cultured skin as a 'smart material' for healing wounds experience in venous ulcers. *Biomaterials* 1996;17:311-320.
164. Guerret S, Govignon E, Hartmann DJ, Ronfard V. Long-term remodeling of a bilayered living human skin equivalent (Apligraf) grafted onto nude mice: Immunolocalization of human cells and characterization of extracellular matrix. *Wound Rep Reg* 2003;11:35-45.
165. O'Brien FJ, Harley BA, Yannas IV, Gibson LJ. Influence of freezing rate on pore structure in freeze-dried collagen-GAG scaffolds. *Biomaterials* 2004;25:1077-1086.
166. Kurz W, Fisher DJ. *Fundamentals of Solidification*. Switzerland: Transtech Publications; 1989.
167. O'Brien FJ, Harley BA, Yannas IV, Gibson LJ. The effect of pore size on cell adhesion in collagen-GAG scaffolds. *Biomaterials* In press, 2004.
168. Heimbach D, Luteran A, Burke J, Cram A, Herndon D, Hunt J, Jordan M, McManus W, Solem L, Warden G. Artificial dermis for major burns. *Ann Surg* 1988;208:313-320.
169. Stern R, McPherson M, Longaker MT. Histologic study of artificial skin used in the treatment of full-thickness thermal injury. *J Burn Care Rehab* 1990;11:7-13.
170. Michaeli D, McPherson M. Immunologic study of artificial skin used in the treatment of thermal injuries. *J Burn Care Rehab* 1990;11:21-26.
171. Burke JF. Observations on the development and clinical use of artificial skin: An attempt to employ regeneration rather than scar formation in wound healing. *Jpn J Surg* 1987;17:431-438.
172. Tompkins RG, Hilton JF, Burke JF, Scoenfeld DA, Hegarty MT, Bondoc CC, Quinby WC, Jr., Behringer GE, Ackroyd FW. Increased survival after massive thermal injuries in adults: Preliminary report using artificial skin. *Child Care Med* 1989;17:734-740.
173. Sheridan RL, Hegarty M, Tompkins RG, Burke JF. Artificial skin in massive burns: Results to ten years. *Eur J Plast Surg* 1994;17:91-93.
174. Besner GE, Klamar JE. Integra Artificial Skin as a useful adjunct in the treatment of purpura fulminans. *J Burn Care Rehab* 1998;19:324-329.
175. Spence RJ. The use of Integra for contracture release. *Burn Care Rehabil*. 1998;19:S173.
176. Lorenz C, Petracic A, Hohl H-P, Wessel L, Waag K-L. Early wound closure and early reconstruction: Experience with a dermal substitute in a child with 60 percent surface area burn. *Burns* 1997;23:505-508.
177. Pandya AN, Woodward B, Parkhouse N. The use of cultured autologous keratinocytes with Integra in the resurfacing of acute burns. *Plast Reconstr Surg* 1998;102:825-828.
178. Lattari V, Jones LM, Varcelotti JR, Latenser BA, Sherman HF, Barrette RR. The use of permanent dermal allograft in full-thickness burns of the hand and foot: a report of three cases. *J Burn Care Rehabil* 1997;18:147-155.
179. Tsai CC, Lin SD, Lai CS, Lin TM. The use of composite acellular allodermis-ultrathin autograft on joint area in major burn patients—one year follow-up. *Kaohsiung J Med Sci* 1999;15:651-658.

See also ENGINEERED TISSUE; SKIN, BIOMECHANICS OF; SKIN SUBSTITUTE FOR BURNS, BIOACTIVE.

SKIN, BIOMECHANICS OF

YAN CHEN
BRIAN L. DAVIS
Department of Biomedical
Engineering/ND20 Lerner
Research Institute
The Cleveland Clinic
Foundation
Cleveland, Ohio

INTRODUCTION

Skin, the body's largest organ, accounts for about 16% of an average adult's body weight and occupies a surface area of

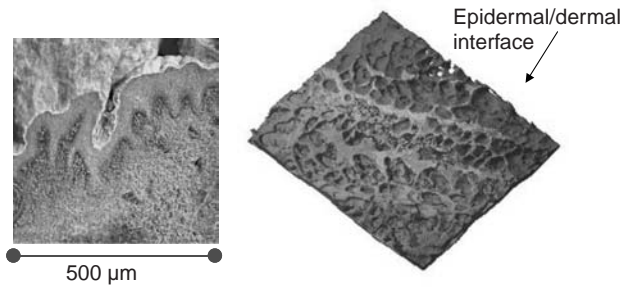


Figure 1. Layers of skin (left) and view of rete ridges (right) (1).

approximately 2 m^2 . It has multiple essential functions: It provides the inner organs and vasculature with a protective barrier against injury or invasion by microorganisms; it prevents water loss; its sensory nerves can sense contact, pain, pressure, heat, and cold; and it plays an important role in maintaining normal body temperature via the function of sweat glands.

Structurally, the skin consists of two principal layers—epidermis and dermis—and various sublayers (Fig. 1). The epidermis is the thinner outer layer, composed of stratified squamous epithelium organized in four or five layers. The main cells in the epidermis are keratinocytes, which are generated by continuous divisions of stem cells that form the basal layer of the epithelium. As keratinocytes migrate to the surface of the epidermis, they grow and differentiate, synthesizing large amounts of a cytoskeletal protein called keratin. This protein builds into 10 nm filaments that gradually come to occupy 80% of the cell volume.

The epidermis is 0.07–0.12 mm in thickness over most of the body surface. It has no blood vessels and thus relies on the dermis for nutrients. The dermis is a supportive layer beneath the epidermis, composed of fibroblasts, fibrocytes, collagen, elastic fibers, glycosaminoglycan (GAG) matrix, blood vessels, and nerves. It is about 0.5–2.5 mm thick. The dermis has two distinct layers: papillary dermis (about 20% the overall thickness) and reticular dermis. The papillary layer is a loose connective tissue, containing large blood vessels, interlacing elastin fibers and bundles of collagen fibers. In contrast, the reticular dermis is a dense, irregularly arranged connective tissue containing interlacing bundles of type I collagen fibers and coarse elastic fibers.

The interface between the dermis and the epidermis varies from region to region. In thick skin, such as that found on the plantar aspect of the foot, the epidermis–dermis interface has the configuration of rete ridges (downgrowths of epidermis and dermis). The upper dermis exhibits a pattern of primary ridges separated by deep primary grooves (Fig. 1). Rows of conical dermal papillae project upward into the conforming concavities between the ridges in the deep surface of the epidermis. Tethering fibers connecting dermis and epidermis to the base membrane keep the dermis and epidermis from separating. In thin skin (e.g., facial skin), the dermo-epidermal junction is much simpler. The dermal papillae in thin skin are shorter, wider, and fewer and not arranged in the pattern of ridges and grooves as observed in thick skin.

Beneath the dermis is a subcutaneous layer (also called the hypodermis) consisting of areolar and adipose tissues. This layer is sometimes considered a third layer of skin. Collagen fibers from the dermis extend down into the superficial fascia and anchor the skin to the subcutaneous layer, firmly attaching it to underlying tissues and organs. The superficial fascia provides the skin's loose flexible connection with the other internal soft tissues, whereas the upper layers (epidermis and dermis) protect skin from injuries.

MECHANICAL PROPERTIES OF SKIN

Microstructure of Skin

The mechanical properties of skin have long interested dermatologists and bioengineers (2–11). Studies of the mechanical properties of skin could provide objective information related to skin and the changes it undergoes with age and disease. For instance, knowledge of the skin's mechanical properties is of potential use in assessing conditions of connective tissue disease, skin aging, wound healing, and scarring.

Like any other material, the mechanical properties of skin are determined by those of its structural constituents. Major structural components of skin are elastin fibers, collagen fibers, and the ground substance. The individual mechanical and structural properties of collagen fibers, elastin fibers, and their interaction with the ground substances in skin form the basis of mechanical properties of skin.

Collagen. Collagen, the body's most abundant and common protein, represents approximately 72% of the dry weight of dermal tissue. There are now more than 19 known collagen types, each of which is a genetically distinct product. The collagen of adult dermis is mainly type I (85–90%), with 8–11% of type III and 2–4% of type V collagen. In all connective tissues, collagen exists as fibers made up of crimped fibrils. Depending on the tissue, a collagen fibril varies from 50 nm to 300 nm in diameter. Collagen fibers in skin are randomly oriented in layers or lamellae, which give skin large extensibility and resiliency under stress. This arrangement contrasts with that found in the tendon and ligament, where the parallel alignment of collagen fibers gives these tissues higher values of tensile strength requirement.

Collagen fibrils are stabilized and strengthened by the formation of covalent cross-links. These cross-links are formed by lysine and hydroxylysine residues. The intermolecular cross-links are formed by the joining of two hydroxylysine residues and one lysine residue. The cross-links are formed between residues near the amino terminus of one collagen molecule and the carboxyl terminus of another.

There is evidence that in the presence of hyperglycemia (such as in diabetes), some proteins, such as collagen, undergo nonenzymatic glycation (12–15). This process modifies the structure of collagen and has a direct effect on the mechanical properties of collagen, resulting in increased mechanical stiffness and decreased flexibility

of these collagen-containing tissues (16–18). Increased collagen cross-linking due to a buildup of advanced glycation endproducts (AGEs) is also believed to be a major contributor to many complications of diabetes (18,19). AGE content has been shown to be four times higher in the collagen of diabetic versus nondiabetic patients (20). Besides being more rigid, highly crosslinked tissues are more resistant to enzymatic digestion by collagenase (18,21).

The strength of collagen fiber is similar to that of skin, which suggests that collagen fibers are the dominant load-bearing material at high strains. However, the extension of collagen at failure is found to be only 10% that of skin. Studies of the geometry of collagen fibers show that the apparently randomly coiled collagen fibers do not carry any part of the applied stress during the initial elastic deformation of skin. As fibers become oriented and straightened out in the stress direction, they start to carry stress, and the high stiffness of the collagen prevents any further large strain.

Elastin. Elastin accounts for 4% of the dry weight of dermal tissue. Elastin polypeptide molecules are cross-linked together to form rubberlike, elastic fibers. Each elastin molecule uncoils into a more extended conformation when the fiber is stretched and will recoil spontaneously as soon as the stretching force is relaxed. Elastin fibers are highly extensible, and their extension is reversible even at high strains. They exhibit linear elasticity with low stiffness up to about 200% elongation followed by a short region where the stiffness increases rapidly until failure. The loading and unloading paths of elastin fibers do not show significant hysteresis. Elastin fibers are characterized as low-modulus elastic material. Analyzing skin disks from cadavers mechanically and histologically, Dick (22) concluded that the loss of elastin fibers in older persons resulted in a loss of skin resistance to deformation at low stress. King and Lawton (23) investigated the behavior of elastin-rich tissues in terms of elastomer theory and found good agreement with experimental data. When other structural components such as collagen and ground substance are present, however, the elastomer theory is no longer applicable. Daly (24) found that the elastic recovery of skin is completely lost when elastin is removed. Thus, he concluded that the elastic behavior of skin during the initial low strain region is due to the small amount of elastin in dermis.

Ground Substance. Ground substance is the intercellular, non-fibril material in connective tissue. It is composed of tissue (extracellular) fluid, amorphous component proteoglycans, and GAGs. Ground substance accounts for 20% of the dry weight of dermal tissue.

Ground substance contributes to the time-dependent properties of skin. Minns et al. (25) performed stress-strain and relaxation studies on human tendon, aorta, and bovine ligamentum nuchae after removing the ground substance with an enzyme or chelating agents. They noted a decrease in stress level, stiffness, stress relaxation, hysteresis, and other time-dependent effects in all three tissues.

Much is known about the structure of collagen, elastin, and proteoglycans at the molecular level. However, less is known about the higher level of structure. For example, how the collagen fibers and elastin fibers are connected to each other in the network of skin and how this might be affected by disease, such as diabetes, are not known.

Skin Mechanics

Most engineering materials are elastic material, which can be described, for small strains, by Hooke's law of linear elasticity: Stress σ is linearly proportional to strain ϵ . Elastic materials show time-independent material behavior, that is, they deform instantly when external loading is applied and return back to their original state immediately when the applied loads are removed.

Unlike engineering materials, the responses of most biological tissues to mechanical loading are complex. Biological tissues have self-adapting and self-repairing characteristics. That is, they can alter their mechanical properties to adapt to changing mechanical demands and can repair themselves. The mechanical properties of biological tissues tend to change with age and some connective tissue diseases. Most biological tissues are composite materials with nonhomogeneous and anisotropic properties, and almost all biological tissues are viscoelastic in nature.

Viscoelasticity. The response of viscoelastic materials not only depends on the strain level applied but also on the time and rate at which the external loading are applied. Skin is such a composite material, made up of collagen fibers, elastin fibers, keratin, and ground substance. Collagen fibers and elastin fibers explain the solid behavior of skin; ground substance contributes to the fluid behavior of skin. The mechanical properties of skin are determined not only by the individual component properties, but also by the geometry configuration and interaction of these components.

As a viscoelastic material, skin has the characteristic of creep, stress relaxation, hysteresis, and precondition.

- Creep: When skin is suddenly loaded with a constant tension, its lengthening velocity decreases against time until equilibrium.
- Stress relaxation: When skin is suddenly extended and maintained at its new length, the stress gradually decreases over time.
- Under cyclic loading, the stress-strain curve shows two distinct paths corresponding to the loading and unloading trajectories. This phenomenon is called hysteresis. With repeated loading cycles, the load-deformation curves shift to the right in a load-elongation diagram and the hysteretic effects decrease. By repeated cycling, eventually a steady state is reached at which no further internal changes in tissue structure will occur unless the cycling routine is changed. At this point, tissue is defined as preconditioned.

Nonlinear Elasticity. The elasticity of skin is strongly nonlinear (Fig. 2). In the initial low-strain region, elongation of skin occurs without appreciable force (a). In the

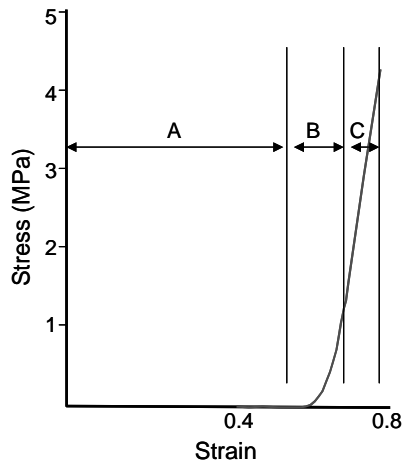


Figure 2. Typical stress-strain curve of skin under incremental loading (4).

mid-strain region, the curve increases in a roughly exponential manner (b) until, at the high-strain region, it becomes almost a straight line (c) immediately before rupture occurs. The accepted explanation of the curve is that the initial deformation (Fig. 2a) is due to deformation of the delicate elastin network; the second part of the curve (b) is due to a gradual straightening of the randomly oriented collagen fibers; the third part of the curve (c) is because most collagen fibers are elongated in the direction of the stress. If the initial part of this curve is magnified, a straight-line approximation will allow the calculation of elastic properties such as Young's modulus. From a functional part of view, the first parts of the curve are more important because they correspond to the physiological range in which the skin normally functions.

Anisotropy. Like most composite materials, skin is anisotropic. This mechanical behavior of human skin can be demonstrated by stretching skin in different directions or by a biaxial loading test (26). Results from tests such as these show directional variance related to lines of tension or cleavage (known as Langer's lines) within the skin that are characteristic for each part of the body. They are named after Austro-Hungarian anatomist Karl Langer (1819–1887), the first to systematically investigate the configuration of the tension on cuts in the human body. Microscopic examination of sections of skin along and across Langer's lines shows a preferential orientation of the collagen fibers of skin. The orientation seems to be at an angle slightly less than 45° from the direction of the lines. Ridge and Wright (27) performed a series of experiments with an extensometer in which they stretched pieces of skin with and against the normal skin tension lines. They could show a clear difference in the directional mechanical properties in the skin, which corresponded to Langer's lines.

Measurement of Skin Mechanical Properties

Information about the mechanical properties of skin falls under three broad categories: strength values (e.g., breaking strain), time-dependent values (e.g., creep, relaxation),

and non-time-dependent values (e.g., elasticity). The measurement of mechanical properties of skin can be performed either *in vitro* or *in vivo*. Results from *in vitro* tests are estimates of properties that are assumed to have an important influence on function *in vivo* and can provide information about the basic mechanical properties of strength, elasticity, or density inherent in the architecture of skin. However, *in vitro* studies are limited in that the characteristics of isolated skin are modified. Dissected skin samples no longer have the same physiological properties as *in vivo* skin, such as those of internal tension, hydration, and vascularization. *In vivo* tests may provide information on the function and the kinetics of change in mechanical properties.

In Vitro Studies. In testing the strength characteristics of biological material, Yamada (28) suggested that the material should be in a mechanically stabilized state in which constant values are obtained. Such a state can be achieved by putting the biological material in a physiological saline solution and storing it in a refrigerator overnight or longer. Yamada noted that the duration of the mechanically stabilized state for skin is 3 days.

There have been many investigations of the mechanical properties of both animal and human skin. Most have used uniaxial test procedures. Many of these early experiments were flawed due to unsophisticated equipment and poor control of such variables as temperature and humidity. Ridge and Wright (1964) developed a skin extensometer, which stretched 1 cm × 0.4 cm strips of skin at a constant rate of 0.2 in. per min. They characterized the resulting stress-strain curves with the equation:

$$e = c + kL^b$$

where e is extension; L is load; and c , b , and k are constants. The authors felt that constant b reflected a specific property of the collagen fibers and constant k represented the condition of the fiber network, which was related to the length and area of the fibers.

Daly (3) further refined uniaxial testing. He first appreciated that for consistent results the specimens must be tested at a constant temperature, humidity, and pH and with a constant strain rate. In addition to measuring stress-strain curves and Poisson's ratio, he performed extensive experiments documenting stress relaxation and creep in human cadaver skin.

An accurate biomechanical description of skin requires constitutive equations that characterize soft tissues in three dimensions. For practical analysis, skin can be considered an incompressible solid material. It is possible to determine three-dimensional mechanical properties from two-dimensional tests for an incompressible solid. Lanir and Fung (29) developed a two-dimensional experimental system for biaxial testing. Each specimen of skin is hooked along its edges by many small staples. Each hook connects by means of a small thread to a force-distributing platform, where tension can be individually adjusted. The investigators used this device to measure the biomechanics of rabbit skin. In their study, the deformed skin always returned to its predeformed state so long as no surface dimension had been allowed to decrease below its initial

unstressed value. This study is obviously different from uniaxial tests, in which the skin essentially always diminishes in lateral directions and usually does not return to its prestress state. Biaxial tests showed that:

1. The stress-strain relationship was extremely nonlinear.
2. Hysteresis was present at all strain levels.
3. Stress-strain relationships were minimally dependent on the strain rate.
4. The relaxation curve did not terminate at the origin and returned to it only after a long period of relaxation.

Reihsner et al. (9) studied the two-dimensional biomechanical behavior of human skin samples from different anatomical sites. Using the *in vivo* geometry of the specimen as reference, a set of incremental strains was applied to the skin. After stress relaxation was completed, the final values of stresses were recorded and related to the incremental strains. Six independent elastic constants were determined. The average deviation of the orientation of maximum principal stress from the direction of the Langer cleavage lines was in the range of angles of -10° to $+10^\circ$. The effect of a directional dependency was most pronounced in skin samples from patella and abdomen. Across a range of ages, Reihsner et al. observed no uniform trend in the principal stresses except that the oldest subject showed the highest principal stresses. The strain necessary to restore the *in vivo* configuration decreased with age in both principal axes of strain. Reihsner et al. suggested that the regional differences in anisotropic behavior could be explained by the different interdigitation between the epidermis and the dermis, the polydispersity of fiber diameter distribution, and the differences in fiber bundle orientation.

Daly (4) conducted tensile tests on specimens oriented at right angles to each other and found that anisotropy is related only to the magnitude of the initial large extension region and the initial low stiffness and the final high stiffness is not orientation dependent. He concluded that the geometry of elastin is not entirely random, as the elastin network determines the initial configuration of the collagen.

In Vivo Studies. *In vivo* tests can be classified as static or dynamic. In static tests, a single modulating stimulus is applied and some resulting change is measured. In dynamic tests, a cyclical stimulus is applied and the initial adapting and final steady-state reactions are monitored. Dynamic tests can provide more information than static tests. However, there is a potentially unlimited combination of frequencies, magnitude, and attachment area. The effects of these parameters on the skin may not be fully understood and may result in difficulties in interpretation of data.

Several techniques have been used to study the mechanical properties of human skin *in vivo*:

1. Tonometric measurements that evaluate the ability of the skin to withstand vertical forces of extension (8).

2. Traction, which applies a linear displacement in the horizontal plane of the skin.
3. Indentation.
4. Torsion, which applies a torque to the skin. The response of skin to shear force can be measured with this device. A device using this technique, the "dermal torque meter," has been commercialized.
5. Suction, which involves placing a suction cup or cylinder on the skin surface and applying a negative pressure to raise a dome of skin. The pressure and the height of the dome are used to calculate elasticity parameters. Two devices based on suction, the Cutometer SEM 474 (Courage and Khazaka, Cologne, Germany) and the Dermaflex A (Cortex Technology, Copenhagen, Denmark), have been commercialized.
6. Elastic wave propagation systems involve measurements of shear wave velocities and the rate at which they are dissipated during their passage in the skin. One study using this technique suggests a probable reduction in skin water content with age (10).
7. Mechanical impedance method. Skin is made to vibrate by a probe over a rectangular frequency spectrum of 20 to 500 Hz after applying a standard preload. One device using this method (DPE, Cotas Computer Technology, Denmark) has been commercialized.

When studying the mechanical properties of skin *in vivo* by such methods, consideration should be given to the contribution of different tissue layers (i.e., dermis and subcutaneous fat). Diridollou et al. (11) reported a device called the "echorheometer," comprising a suction system with an ultrasound scanner. Using this device, the behavior of the dermis and subcutaneous fat under suction was investigated. They observed that there is a certain amount of infiltration of fluid into the tissue under suction. Upon returning to atmospheric pressure, excessive fluid remained in the tissue, which might explain the hysteresis of skin. The authors found that resistance of skin to the applied vertical stress is essentially due to the dermis rather than the subcutaneous fat. The relative contribution of each is not easily evaluated. With high suction pressures applied to large surface areas or low suction pressures to smaller surface areas, the effects on skin cannot be isolated from those of the subcutaneous fat.

Biomechanics of Diabetic Skin Ulceration

The mechanics of skin breakdown under the foot incorporate several aspects described in this article: biaxial skin properties and applied stresses, differing stiffness values within each skin layer, effects of subcutaneous fat, collagen cross-linking and nonlinear, and viscoelastic properties of connective tissue. Although it is recognized that neuropathy and ischemia are primary predisposing factors in the formation of diabetic foot ulcers, an initiating factor, such as physical or mechanical stress, is required for an ulcer to develop (30). For patients with neuropathy, this

mechanical stress need not be excessive, only repetitive, for ulcers to develop.

Brand (31) conducted a series of experiments on rat footpads in which lightly anesthetized rats had one foot placed in a machine that applied repeated intermittent stress to the footpad to simulate a human walking 7 miles. The footpads became swollen after 2 to 3 days, blistered by 1 week, and ulcerated and necrotic after 10 days. In neuropathic animals, the number of cycles that could be tolerated was reduced. Histologic studies on the rats showed inflammation of the footpads after a few days, with small foci of necrosis in areas with large numbers of inflammatory cells. Brand postulated that a similar mechanism occurs in the formation of neuropathic foot ulcers, with a sterile inflammatory autolysis occurring in the subcutaneous tissue due to the repetitive moderate stresses imposed by walking.

The stresses that the foot experiences during walking are vastly different from those encountered by any other skin surface. At the foot-ground interface, a complex combination of shear and pressure exists (32) that could explain why some patients get foot ulcers and others do not. Mathematical modeling of (1) the effects of stresses at the skin/fat interface and (2) the influence of more cross-linked collagen (33) showed that small increases in shear stresses and/or increased skin stiffness had a profound effect on the stresses at the interface between dermis and underlying soft tissue (Fig. 3).

The effects of increased dermal interface stresses have also been assessed in an experimental study (1). In total, 21 diabetic (68.0 ± 2.4 years) and 17 nondiabetic (74.6 ± 3.8 years) amputated foot specimens were tested. The procedure involved peeling strips of skin off the plantar surface of each foot specimen. Each strip was approximately 8 mm wide (in the antero-posterior direction) and 20 mm long (in the medio-lateral direction). While tension was applied via a Materials Testing System (MTS), force and displacement of the loading head were recorded at a sampling rate of 60 Hz. Data across all 38 specimens are shown in Fig. 4.

Within the nondiabetic group, analysis of variance showed that the required forces for removing skin at different sites were significantly different ($p < 0.001$), with the skin-fat interface being significantly stronger

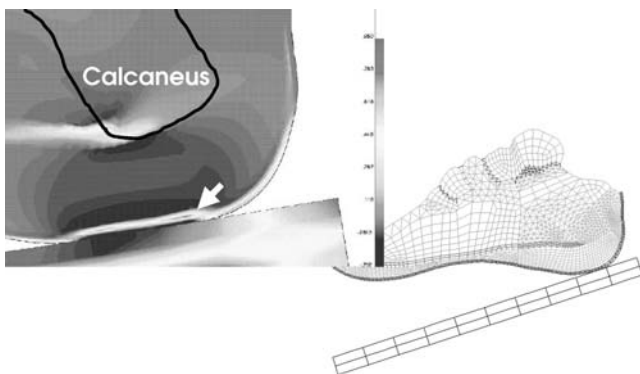


Figure 3. Through finite element modeling of the hindfoot, the predicted effect of diabetes-induced skin stiffening is to increase the stresses at the skin-fat interface by 100% (33).

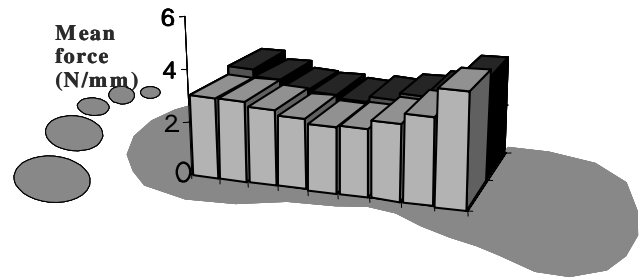


Figure 4. Combined diabetic and control data showing the forces required to remove skin strips off the plantar surface of amputated foot specimens. These forces are significantly higher ($p < 0.05$) in the forefoot and hindfoot regions compared with the midfoot area (1).

in both the metatarsal head and heel regions than in the midfoot area. These regions also withstand the largest weight-bearing loads during normal locomotion. Neuropathic diabetic patients (who typically have higher plantar pressures) also showed higher strengths at the skin-fat interface than age-matched nondiabetic subjects (1). Every strip (e.g., first strip on medial side, or third strip on lateral side, etc.) for the diabetic group required larger forces to pull it off the underlying soft tissue ($p < 0.05$). Across all skin strips, diabetic specimens required a 20% greater removal force.

SUMMARY

Skin has multiple essential functions ranging from protection against injury or invasion by microorganisms, to a barrier to water loss, to maintaining normal body temperature. It consists primarily of two principal layers—epidermis and dermis—and various sublayers. From a mechanical point of view, these layers have stiffness values that can differ by three orders of magnitude! The area of skin that withstands the greatest daily history of loading—that under the foot—has an interface between the dermis and the epidermis characterized by rete ridges. The precise mechanical function of these ridges is not yet clear. What is known is that this interface is profoundly affected by the magnitudes of shear and pressure loading that plantar soft tissue is required to withstand. Areas that have higher loading (e.g., heel or metatarsal head region) have a dermis that is more tightly bonded to the subcutaneous soft tissue than other regions. In this respect, skin seems to follow the “form follows function” law (or “Wolff’s Law”) that is usually used to describe bone’s adaptation to applied loading. When these loads become extreme (e.g., greater than 1 MPa applied pressure), even the short weight-bearing intervals during gait are sufficient to lead to skin breakdown. The precise site for the initiation of failure (e.g., which layer of skin fails first) has yet to be determined. As skin breakdown under the foot is a clinical problem with significant risks for the patient, it is likely that research into this problem will continue until skin failure mechanisms are better understood. These may then serve as the basis for better prevention or treatment strategies.

BIBLIOGRAPHY

1. Chen Y. The influence of diabetes on mechanical properties of skin. Unpublished doctoral dissertation. Cleveland, OH: Cleveland State University; April 2003.
2. Ridge MD, Wright V. The description of skin stiffness. *J Biorheol* 1964;10:139–155.
3. Daly CH. The biomechanical characteristics of human skin. Ph.D. Thesis, University of Strathclyde, Glasgow, Scotland, 1966.
4. Daly CH. Biomechanical properties of dermis. *J Invest Dermatol* 1982;79(Suppl 1):17s–20s.
5. Danielson DA. Human skin as an elastic membrane. *J Biomech* 1973;6:539–546.
6. Lanir Y, Fung YC. Two-dimensional mechanical properties of rabbit skin. II. Experimental Results. *J Biomech* 1974;7:171–182.
7. Lanir Y. Structural theory for the homogeneous biaxial stress-strain relationships in flat collagenous tissues. *J Biomech* 1979;12:423.
8. Warren R, Gartstein V, Kligman AM, Montagna W, Allendorf RA, Ridder GM. Age, sunlight, and facial skin: A histologic and quantitative study. *J Am Acad Dermatol* 1991;25:751–760.
9. Reihnsner R, Balogh B, Menzel EJ. Two-dimensional elastic properties of human skin in terms of an incremental model at the in vivo configuration. *Med Eng Phys* 1995;17:304–313.
10. Potts RO, Buras EM Jr, Chrisman DA Jr. Changes with age in the moisture content of human skin. *J Invest Dermatol* 1984;82:97–100.
11. Diridollou S, Berson M, Vabre V, Black D, Karlsson B, Auriol F, Gregoire JM, Yvon C, Vaillant L, Gall Y, Patat F. An in vivo method for measuring the mechanical properties of the skin using ultrasound. *Ultrasound Med Biol* 1998;24:215–224.
12. Brownlee M, Vlassara H, Cerami A. Nonenzymatic glycosylation and the pathogenesis of diabetic complications. *Ann Intern Med* 1984;101(4):527–537.
13. Buckingham B, Reiser KM. Relationship between the content of lysyl oxidase-dependent cross-links in skin collagen, nonenzymatic glycosylation, and long-term complications in Type I diabetes mellitus. *J Clin Invest* 1990;86:1046–1054.
14. Delbridge L, Ellis CS, Lequesne LP. Non-enzymatic glycosylation of keratin from the diabetic foot. *Br J Surg* 1983;70:305.
15. Monnier VM, Kohn RR, Cerami A. Accelerated age-related browning of human collagen in diabetes mellitus. *Proc Natl Acad Sci USA* 1984;81:583–587.
16. Cerami A, Vlassara H, Brownlee M. Role of advanced glycosylation products in complications of diabetes. *Diabetes Care* 1988;11(Suppl 1):73–79.
17. Hamlin CR, Kohn RR, Luschin JJ. Apparent accelerated aging of human collagen in diabetes mellitus. *Diabetes* 1975;24:902–904.
18. Reiser KM. Nonenzymatic glycation of collagen in aging and diabetes. *Proc Soc Exp Biol Med* 1991;196:17–29.
19. Kennedy L, Baynes JW. Non-enzymatic glycosylation and the chronic complications of diabetes: An overview. *Diabetologia* 1984;26:93–98.
20. Makita Z, Radoff S, Rayfield EJ, Yang Z, Skolnik E, Delaney V, Friedman EA, Cerami A, Vlassara H. Advanced glycosylation end products in patients with diabetic nephropathy. *N Engl J Med* 1991;325(12):836–842.
21. Elkeles RS, Wolfe JHN. ABC of vascular diseases. The diabetic foot. *BMJ* 1991;303:1053–1055.
22. Dick JC. The tension and resistance to stretching of human skin and other membranes. *J Physiol* 1951;112:102–113.
23. King AL, Lawton RW. Elasticity of body tissues. In: Glasser O, editor. *Medical Physics*, Vol. 3. Chicago: Yearbook Publishers; 1960. 234–247.
24. Daly CH. The role of elastin in the mechanical behavior of human skin. *Proc. 8th Intl. Conf. Med. Biol. Engr.*. Chicago, IL: 1969: 18–27.
25. Minns RJ, Soden PD, Jackson DS. The role of the fibrous components and the ground substance in the mechanical properties of biologic tissues: A preliminary investigation. *J Biomech* 1973;6:153–165.
26. Tong P, Fung YC. The stress-strain relationships for the skin. *J Biomech* 1976;9:649–657.
27. Ridge MD, Wright V. The directional effects of skin. A bioengineering study of skin with particular reference to Langer's lines. *J Invest Dermatol* 1966;64:341–346.
28. Yamada H. *Strength of Biological Materials*. Baltimore, MD: Williams & Wilkins; 1970.
29. Lanir Y, Fung YC. Two-dimensional mechanical properties of rabbit skin. I. Experimental System. *J Biomech* 1974;7:29–34.
30. Laing P. The development and complications of diabetic foot ulcers. *Am J Surg* 1998;176(Suppl 2A):11S–19S.
31. Brand PW. Pathomechanics of diabetic (neurotrophic) ulcer and its conservative management. In: Bergan JJ, Yao JST, editors. *Gangrene and Severe Ischaemia of the Lower Extremities*. New York: Grune & Stratton; 1978. p 185–189.
32. Davis BL. Foot ulceration: Hypotheses concerning shear and vertical forces acting on adjacent regions of skin. *Med Hypoth* 1993;40:44–47.
33. Thompson DL. Finite element modeling of the diabetic foot. Unpublished Masters Dissertation, Ohio State University, 1998.

See also SKIN SUBSTITUTE FOR BURNS, BIOACTIVE; SKIN TISSUE ENGINEERING FOR REGENERATION; TACTILE STIMULATION; ULTRAVIOLET RADIATION IN MEDICINE.

SLEEP LABORATORY

RICHARD K. BOGAN
SHAWN D. YOUNGSTEDT
University of South Carolina
Columbia, South Carolina

INTRODUCTION

Sleep is a fundamental, homeostatic process necessary for human existence and quality of life. Sleep disorders account for impairment of alertness (1); cognitive function (2), especially short-term memory and divided-attention tasks (3); mood (4); work productivity (5); and driving ability (and hence, an increase in automobile accidents). Obstructive sleep apnea has been clearly linked to cardiovascular disease.

There are > 80 sleep disorders, which generally increase in prevalence with age. Indeed, > 50% of adults over age 60 years suffer from some sleep-related complaint (6), including daytime fatigue, difficulty initiating or maintaining sleep, snoring, obstructive apnea, insomnia, restless legs syndrome, narcolepsy, or circadian rhythm disorders. Most disorders have attendant morbidity issues. For example, patients with sleep apnea have an increased risk of cardiovascular disease, stroke, and diabetes. Patients with protracted insomnia have

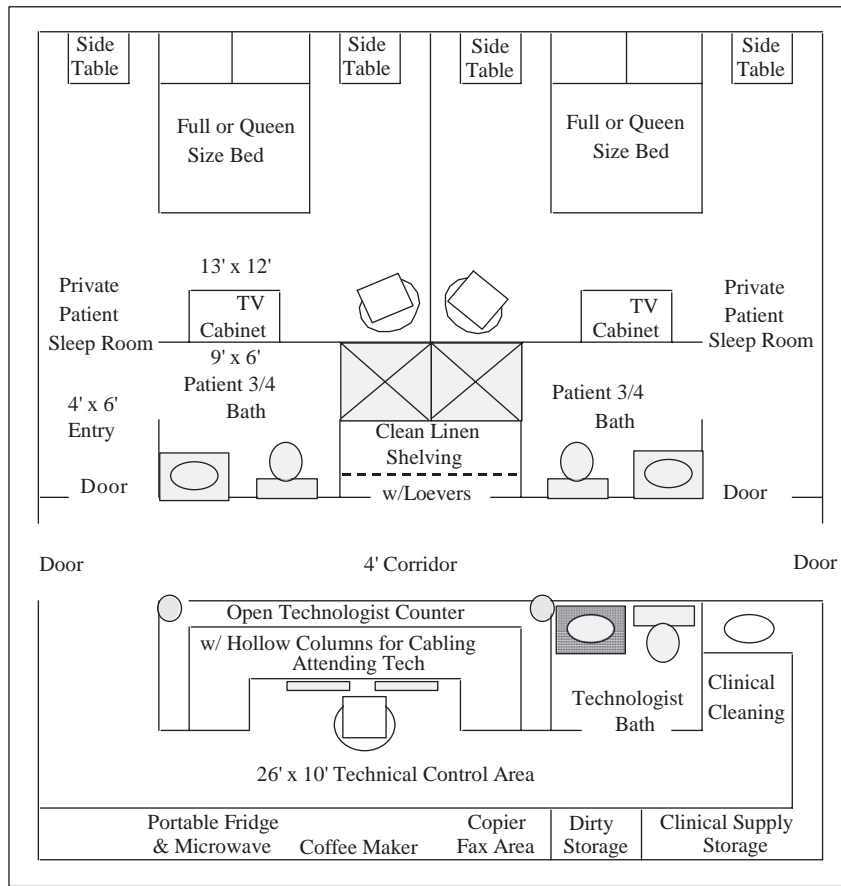


Figure 1. Sample sleep lab layout (26 ft × 32 ft 832 ft² for a two-bed sleep lab).

an increased risk of depression, anxiety, and substance abuse. Many sleep disturbances are secondary to other medical problems, such as nocturia, pain associated with arthritis, and cardiopulmonary disorders.

POLYSOMNOGRAPHIC RECORDING

The traditional evaluation of normal sleep and sleep disorders incorporates the gold standard, nocturnal polysomnography, performed in a sleep clinic or laboratory (Fig. 1). Growth of laboratories has occurred in the last 30 years to 2515 estimated labs, generally located in hospitals. Most current laboratories require a night tech, a scoring tech, a clinician, or a scientist to acquire and process sleep data. Some laboratories are freestanding or exist in private clinics. Contract service entities also provide outsourcing of sleep diagnostics.

The sleep laboratory classically consists of diagnostic bedrooms, accompanied by a central technical area for collection of data, observation of patients, and data processing. Ancillary areas include a business office, exam rooms, a lounge area, a break area, and file storage, much like a clinical practice. Sleep diagnostics for the sleep bedroom are focused primarily on a monitoring EEG to determine sleep states, with expanded measurements to allow quantitative assessment of EEG (see Table 1). The sleep laboratory environment attempts to achieve a bedroom environment,

with the patient interface allowing minimally invasive techniques.

Electroencephalography

The EEG is the fundamental measurement of polysomnography and consists of application of electrodes according to the International 10–20 electrode placement system (7) (Fig. 2). The skin interface is established through cleansing and removal of dead skin layers, with electrode application, typically using an electrode cup with a conductive medium. There are many derivations of electrode placement, but for sleep, typical placements are at occipital lobe sites (O1 or

Table 1. Sleep Laboratory Measures (Possible)

Eye movements	Inductance plethysmography
Pupillometry	Intercostal EMG
Arms, legs and chin electromyography	Esophageal pressure
Nasal and oral airflow	Esophageal pH
End tidal CO ₂	Pulse oximetry
Nasal pressure-flow	Transcutaneous oximetry
Pneumotach	Transcutaneous CO ₂
Snoring microphone	Electrocardiogram
Chest wall movement	Blood pressure
Abdominal movement	Penile tumescence
	Body position

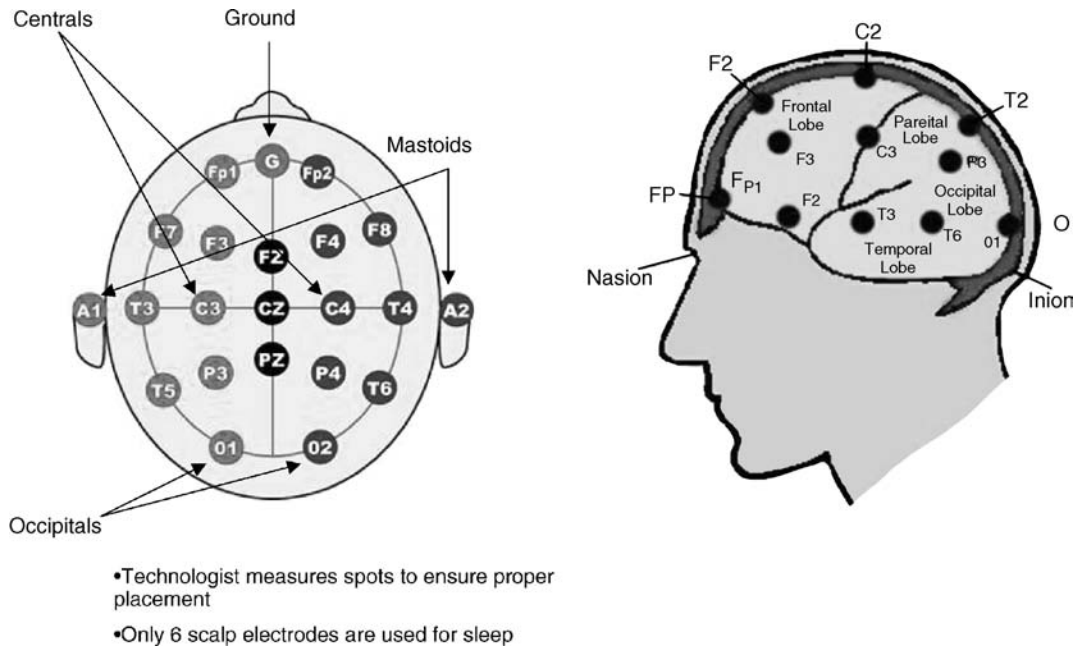


Figure 2. Electrode placement.

O2) with contralateral placement of a reference electrode on the mastoid process or earlobe. This gives the standard O1/A2 or O2/A1 electrode placement. Many labs also incorporate central electrode placement (usually C3/A2 or C4/A1). Occasionally, frontal electrodes are placed. The contralateral reference of electrodes typically allows a high amplitude of EEG signal, enabling adequate measurement of frequency and amplitude information. Normal sleep scoring rules have been based on central electrode placement. Occipital electrodes enhance alpha frequency (8–12 Hz) measurements and frontal electrodes enhance delta frequency (2–4 Hz) measurements.

Full montage EEG electrode placement is sometimes done to enhance the capture of nocturnal seizure activity. This replicates bipolar studies done for seizure monitoring and is especially useful for nocturnal movement disorder evaluation.

Electroculography

The electrooculogram (EOG) is a measurement of eye activity. Phasic bursts of rapid eye movement occur during rapid eye movement sleep (REM sleep). The EOG is important for recognition of REM sleep, as well as for detecting the transition from wakefulness to Stage 1 sleep, which is characterized by slow, rolling eye movements.

The EOG recordings are possible due to a small electro-potential difference from the cornea to the retina. The electrodes are positioned at the right outer canthus (ROC) and the left outer canthus (LOC), eferenced to auricular electrodes. Thus, ROC/A1 and LOC/A2 will register as out-of-phase pen deflections. This facilitates artifact recognition, as well as EEG activity recorded in the eye electrodes. EOG placements incorporate a somewhat superior placement of the electrode on one eye and inferior on the other eye, which allows recognition of vertical eye movements.

Electromyography

Three electromyographic (EMG) electrodes placed in the mentalis–submental area allow detection of EMG activity. During sleep, there is a gradual decline in EMG activity, and the dramatic reduction of muscle tone during REM, makes the EMG an important confirmation of REM sleep. Again, adequate removal of oils and dead skin cells with a conductive electrode placement enhances signal processing.

Electrodes or strain gauge sensors are also commonly placed on the tibialis muscles (Fig. 3). The EMG and movement signals from the lower extremities, and sometimes the arms, allow measures of isolated movements, as well as periodic limb movements that may be seen in restless legs syndrome, periodic limb movements of sleep, REM behavior disorder, seizures, parasomnias, and other disorders.



Figure 3. Electrodes applied to legs.



Figure 4. Technologist monitors through the night.

Other Measures and Filtering of Signals

Other biological measures have been incorporated into the polysomnographic measurements. These include body position, body movement, transient arousals, respiratory abnormalities, heart rate, oxygen saturation, esophageal pH, and esophageal pressure.

Video monitoring, as well as audio monitoring, are desirable features and occur in most laboratories. Video recordings with infrared or low light allow correlations of behavioral states with the physiologic measures.

The multiple channels of biological data are gathered in a bundle of electrical wires (the ponytail) at the back of the scalp. The electrical signals require signal processing, amplification, filtering, monitoring, and subsequent collation. The impact of signal output, sampling rates, filtering, and video recording are all factors in the design and in the monitoring of the patient (Fig. 4), as well as scoring information that is processed according to peer-reviewed criteria and event definitions.

The essential measurements in a sleep laboratory involve sleep staging (described below) and this requires calibration, with close attention to signal amplitude and filtering. Frequency measurements ranging from 2–50 Hz are common for measurement of EEG signals. The EMG signals have much higher frequency and sample filtering usually up to 75 Hz, with a notch filter at 60 Hz used to reduce alternating current noise. Signals <10 Hz are usually filtered for an EMG signal.

SCORING AND INTERPRETING PSG

Sleep Staging

Standard sleep stage scoring is guided by Rechtschaffen and Kale's scoring rules (8). The standard EEG recording is negative-up with amplitude measured from peak to valley of the waveform. Delta rhythm is 2–4 Hz; theta is 4–8 Hz; alpha rhythm is 8–12 Hz; beta rhythm is 13–30 Hz; and gamma is >30 Hz. Sleep is scored in 30 s epochs, with a chart speed equivalent to 10 mm·s⁻¹. The EEG, EOG, and EMG signals are necessary for sleep state assessment and are used to study clinical and research physiology, as well

as pathologic processes. The epoch is scored based on the majority population of EEG/EOG/EMG activity during the 30 s epoch.

Sleep is thus staged as non-REM sleep Stages 1, 2, 3, and 4 and REM sleep. The data are collated in both graphic, as well as tabulated, form to present information from lights-out to sleep onset, as well as quantifying total sleep time, time in bed, sleep efficiency (sleep time divided by time in bed), wake after sleep onset, and sleep stage distribution. A sleep histogram of the night of sleep summarizes the stage distribution pictorially.

Special rules are established for movement, arousals, and periodic limb movements in sleep, as well as consideration for specific disease processes. For example, narcolepsy may be associated with excessive motor activity during sleep, with an elevated EMG amplitude, particularly during phasic rapid eye movement sleep. Obstructive sleep apnea patients may have <15 s of sleep during a 30 s epoch due to arousals from obstructive events, forcing the scoring of the epoch as awake. Special rules to allow for these variances are in place.

Sleep-Disordered Breathing

Abnormal breathing during sleep is one of the most common disorders and accounts for as much as 70% of patients presenting to a sleep laboratory. Some laboratories specialize in sleep-disordered breathing alone. Snoring with increase in upper airway resistance due to functional relaxation of dilator muscles in the pharynx is a common disorder, ranging from 20% to 40% of the adult population and 13% of children. Anatomic changes, particularly tonsillar hypertrophy, add to the increase in upper airway resistance. This increase in the work of breathing may cause unstable breathing, with episodes of apnea or hypopnea, with associated changes in oxygenation, autonomic tone, and sleep state.

Obstructive apnea is defined as 10 s of cessation of airflow, despite a continued effort to breathe, in the adult population. Hypopnea is defined as a discernable reduction in flow or effort that produces a 4% desaturation. Central apnea is cessation of flow associated with cessation of effort for at least 10 s in the adult population. In the pediatric population, the duration for defining apneas is shorter. In some laboratories, hypopneas are also defined as a discernable reduction in flow or effort, which terminates in a 3% desaturation or an arousal. Mass loading of the diaphragm due to obesity and intrinsic central nervous system regulation of breathing are other variables to be considered clinically, as well as in research.

Current laboratories monitor flow, effort, and oxygen saturation as surrogate measures of minute ventilation and gas exchange. The intent is to assess the effort to breathe, the results of the output of ventilation, and, ideally, gas exchange, especially O₂ content or oxygen saturation. These measurements along with clinical correlation provide risk assessment and guide treatment plans.

Methods to detect airflow include pneumotachography, nasal airway pressure, thermistors, and thermocouples, as well as expired CO₂. Nasal airway pressures resemble the signals from a pneumotachograph, and therefore give a



Figure 5. Sleep setup with respiratory belts.

qualitative measure of increase in upper airway resistance. The pressure gradient intranasally compared to ambient atmospheric pressure is used to calculate the flow. Effort signals may be generated by flexible bands around the chest and the abdomen to detect movement. This is the most common technique (Fig. 5). Strain gauges, impedance pneumography, inductance plethysmography, and intercostal muscle electromyography have been used. Snoring sensors include microphones and piezoelectric assessment of vibration.

Oxygen content is most commonly measured by pulse oximetry. Decrease in oxygen saturation during sleep is most commonly due to changes in ventilation. Oxygen desaturation is an important criterion for scoring respiratory events, and thus critical for diagnostic studies. Intrinsic lung disease or cardiac disease may produce ventilation-perfusion mismatch and may also produce decreases in O_2 saturation. In a stable state, however, most decreases in oxygen saturation reflect changes in minute ventilation.

Pulse oximetry uses a two wavelength light transmitter, using spectrophotoelectrical techniques based on oxyhemoglobin absorption. Pulsatile tissues are necessary for pulse oximetry, and therefore these are usually applied to the finger, earlobe, or nasal sites. Reduction in pulsation, skin pigment, and dyshemoglobinemias may interfere with signal processing. The sampling rate, filtering, and proprietary algorithms for signal processing may affect the resolution of the signal.

Transcutaneous oxygen and carbon dioxide are other techniques available, but they are not commonly used, except in neonates. Transcutaneous CO_2 may be useful to assess chronic hypoventilation, but the resolution time for these measures limits their use in routine polysomnography.

Esophageal pressure measurements using an esophageal balloon provide a highly accurate measure of intrapleural pressure and, therefore, work of breathing. Some sleep centers use esophageal pressure measurements; however, this is not routine, as the invasive technique disturbs sleep. However, esophageal pressure measurements enhance the ability to measure upper airway resistance syndrome and central alveolar hypoventilation.

Cardiac data primarily consists of heart rate and rhythm. Limited EKG electrodes, equivalent to lead II (right arm–left arm) or precordial leads (V5 or V6) are used, depending on the electrical axis and lab protocol. Clinical studies focus on high, mean, and low heart rates, as well as semiquantitative data on arrhythmias as to frequency and type. These are correlated with respiratory events or sleep state.

In patients with obstructive sleep apnea, efficacy of treatment is measured by applying positive airway pressure (PAP) or bilevel positive airway pressure (BiPAP). Flow signals are generated from the hardware device, allowing quantitative measures of flow and upper airway resistance based on waveform characteristics. Sleep state, body position, oxygen saturation, electrocardiogram, flow, and effort data are recorded to assure proper titration. The goal is to reduce the respiratory disturbance index, ideally < 5 , and to minimize oxygen desaturations.

Daytime Sleepiness

Objective assessment of daytime sleepiness is another important laboratory diagnostic measure. The multiple sleep latency test involves a series of nap opportunities at 2 h intervals during the day. The standardized technique and normative data facilitate objective measures of excessive daytime sleepiness. This test is primarily used in the diagnosis of narcolepsy. The maintenance of wakefulness test is similar to the multiple sleep latency test, but requires the individual to attempt to remain awake during 20 or 40 min nap opportunities in a dark room.

Treatment Guidelines. Practice-based guidelines by the American Academy of Sleep Medicine and medical specialty societies have established guidelines for diagnosing and treating sleep disorders. These can be accessed on their website (www.aasmnet.org).

ADDITIONAL AMBULATORY MEASURES

Alternative or ancillary devices have been developed, primarily in the area of sleep-disordered breathing and brain state (sleep–wake estimate). Screening devices have been developed for obstructive sleep apnea primarily. Portable outpatient systems for unsupervised polysomnography are available. The unattended study is considered a level II evaluation. Level III is portable sleep apnea testing, monitoring airflow, effort, electrocardiogram, and oxygen saturation. Level IV is arterial oxygen saturation measurement alone.

Actigraphy incorporates an accelerometer used to measure wrist movement. Complex algorithms have been developed to estimate sleep/wake state, periodic limb movements, and circadian rhythm abnormalities. Actigraphy is particularly useful for long-term monitoring of patients in their homes, and for detection of daytime napping, which is usually not possible for EEG, as it is restricted to the night.

DIGITAL PGS AND FUTURE TRENDS

The EEG, EOG, and EMG signals were traditionally recorded in a sleep laboratory with limited channels on

analog systems, with pen deflections recorded on paper. Digital recording systems are more often used now, replicating a chart-paper speed of $10 \text{ mm} \cdot \text{s}^{-1}$. Digitized signals with current hardware and software systems allow enhanced collation of data, increased numbers of channels, enhanced filtering, and the potential for automated analysis. Archiving, data compression, and feature extraction are also possible.

Digitized signals of the ECG with high sampling rates, preferably at 200 Hz or higher, provide an opportunity for high resolution analysis. High resolution analysis might provide clearer insights into autonomic tone and vascular resistance and, the influence of sleep states, pathological conditions, and transient arousals. Further development of quantitative measures of ventilatory response, changes in gas exchange, as well as central nervous system and cardiovascular changes, could be helpful for diagnosing and treating sleep-disordered breathing.

Software processing collates data, presenting information in a tabular, as well as in a graphic, format. Clinicians and scientists review raw data, as well as human-supervised or scored changes. These changes can produce an audit trail referenced to the raw data. The fields of data can populate an electronic medical record, as well as a database, including demographics, comorbidities, medications, and fields of data from the polysomnogram, as well as the final diagnosis, treatment, and outcome measures.

Conceivably, advanced EEG analysis might reduce interscorer variability, thus better defining state and process. Enhanced resolution of EEG with improved consistency might allow better feature extraction, such as transient arousals and continuity measures that may improve research and clinical care. This could significantly improve the efficiency, cost and safety in drug development.

Opportunities for the future in sleep diagnostics are numerous. The duration of a polysomnogram, as well as long-term EEG monitoring, create challenges in the quality of the signal. Improvement of electrode-skin interface and conductivity are desirable. Application of wireless technologies to minimize the ponytail bundle of wires would be extremely useful. Audiovisual monitoring quality, advanced respiratory analysis, high resolution cardiac assessment, technical acquisition, collation, and data management are all current challenges.

BIBLIOGRAPHY

1. Schnieder C, Fulda S, Schulz H. Daytime variation in performances and tiredness/sleepiness ratings in patients with insomnia, narcolepsy, sleep apnea and normal controls. *J Sleep Res* 2004;13:373–383.
2. Ancoli-Israel S, Cooke JR. Prevalence and comorbidity of insomnia and effect on functioning in elderly populations. *J Am Geriatr Soc* 2005;53(Suppl. 7):S264–S271.
3. Thomas RJ, et al. Functional imaging of working memory in obstructive sleep-disordered breathing. *J Appl Physiol* 2005; 98:2226–2234.
4. Drake CL, Roehrs T, Roth T. Insomnia causes, consequences, and therapeutics: An overview. *Depress Anxiety* 2003;18:163–176.

5. Metlaine A, Leger D, Choudat D. Socioeconomic impact of insomnia in working populations. *Indus Health* 2005;43:11–19.
6. Foley DJ, et al. Sleep complaints among elderly persons: An epidemiologic study of three communities. *Sleep* 1995;18:425–432.
7. Jasper HH. The ten-twenty electrode system of the International Federation. *Electroencephalogr Clin Neurophysiol* 10:370–375.
8. Rechtschaffen A, Kales A. A Manual of Standardized Terminology, Techniques, and Scoring Systems for Sleep Stages of Human Subjects. Los Angeles: Brain Information/Brain Research Institute UCLA; 1968.

Further Reading

- Bassetti CL. Sleep and stroke. *Semin Neurol* 2005;25:19–32.
- Drummond SP, Gillin JC, Smith TL, DeModena A. The sleep of abstinent and pure primary alcoholic patients: Natural course and relationship to relapse. *Alcohol Clin Exp Res* 1998;22:1796–1802.
- George CF. Sleep. 5: Driving and automobile crashes in patients with obstructive sleep apnoea/hypopnoea syndrome. *Thorax* 2004;59:804–807.
- Phillips B. Sleep disordered breathing and cardiovascular disease. *Sleep Med Rev* 2005;9:131–140.
- Spira AP, Friedman L, Flint A, Sheikh JI. Interaction of sleep disturbances and anxiety in later life: Perspectives and recommendations for future research. *J Geriatr Psychiatry Neurol* 2005;18:109–115.
- Vgontzas AN, Bixler EO, Chrousos GP. Sleep apnea is a manifestation of the metabolic syndrome. *Sleep Med Rev* 2005;9:211–224.

See also CONTINUOUS POSITIVE AIRWAY PRESSURE; ELECTROENCEPHALOGRAPHY; SLEEP STUDIES, COMPUTER ANALYSIS OF; VENTILATORY MONITORING.

SLEEP STUDIES, COMPUTER ANALYSIS OF

DAVID WM. SHUCARD
BRETT A. PARMENTER
State University of New York
Buffalo, New York

ALI A. EL SOLH
Erie County Medical Center
Buffalo, New York

INTRODUCTION

Human beings spend approximately one-third of their lives asleep. Prior to the advent of physiological records, technology, and digital computers, our understanding of this ubiquitous phenomenon was based on observations of people or animals sleeping. Yet, this observational approach placed severe limitations on our understanding of the functions of sleep and the neurological processes that underlie this apparently quiescent period. It was because of several major scientific developments, all of which occurred during the twentieth century, that the concept of sleep as an active process (rather than a passive one) became appreciated, making the advent of modern sleep medicine possible.

Sleep was generally regarded as a passive process, “. . . the intermediate state between wakefulness and death. . .” (1). In 1939, University of Chicago physiologist Nathaniel Kleitman published a monograph “Sleep and Wakefulness” based on observations of human subjects who were sleep deprived. He determined that sleep may be viewed as a “let down of waking activity” and that “there may be different kinds of wakefulness” (2). In the 1950s, Kleitman and his students, Aserinsky and Dement, went on to describe rapid eye movement sleep (REM) in humans and the relationship between REM and dreaming (3,4). To follow up on the earlier work of, Dement and Kleitman (5) undertook a particularly ambitious project, considering the available technology. They recorded the electroencephalogram (EEG) and electrooculogram (EOG) continuously from 33 subjects for a total of 126 nights. Recordings were all done on paper and manually scored. This work revealed that sleep had an architecture and that the EEG showed cyclical variation throughout the night. REM periods occupied ~20–25% of sleep. Dement and Kleitman characterized sleep stages as they are known today and determined that sleep consisted of two distinct states (REM and non-REM), neither of which could be considered as a time of brain inactivity.

Importantly, the foundation for this work was the recording of the human EEG in 1928 by Berger who showed that there were differences in the rhythm (or EEG frequency) when subjects were awake or asleep. Berger’s work allowed, for the first time, the EEG to be monitored continuously during sleep without disturbing the individual. Hobson (6) suggested that Berger’s work was the turning point of sleep research. Much of what is known about the EEG characteristics of the different stages of sleep was described in the 1930s. Further, it was determined that sleep was an active process that involves the synchronization of a number of neurological systems.

Perhaps one of the most important events in the history of sleep medicine was the discovery of sleep apnea by two separate groups in France and Germany in 1965. Narcolepsy and complaints of insomnia by patients further spurred the development of sleep centers that monitored patients’ physiology as they slept. According to Dement (7), the “consolidation and formalization of the practice of sleep disorders medicine was largely completed” (p. 13) at the end of the 1970s. This labor intensive all night monitoring of patients was done without the benefit of digital computers, which did not become generally available for over a decade.

The purpose of this article is to discuss the use of computers in sleep disorders medicine and how this technology has enhanced the diagnostic and scientific capabilities of the sleep laboratory. The focus of the remainder of this article is on the computer analysis of the data obtained during polysomnography.

INTEGRATION OF COMPUTERS INTO POLYSOMNOGRAPHY

Over the years, and in particular with the development of the digital desktop computer, tools for studying sleep have

advanced rapidly. Polysomnography (PSG), a term coined in 1974 by J Holland at Stanford University, allows for the assessment of sleep by simultaneously measuring multiple physiologic parameters. These parameters generally include EEG, EOG, electromyogram (EMG), electrocardiogram (ECG), respiration (both effort and air flow or pressure), blood oxygenation, and body–limb movement (8).

The results of an 8 h PSG can consist of >1000 pages (depending on the paper speed, or epoch length (e.g., 10, 15, 30 s) of data from multiple recording channels, making manual and visual interpretation a difficult and overwhelming task (9). Typically, each page (e.g., 30 s) of the recording was examined and sleep stage was determined based on the Rechtschaffen and Kales scoring system (10). Depending on the reason for the study, respiratory events such as nasal–oral air flow and respiratory effort, as well as blood oxygen saturation, were frequently recorded. Paper speeds were used so that each page of recording represented a 15–30 s epoch over the full night. So, for example, for a 7 h study and a paper speed set at 30 s epochs, there would be 840 pages of data to review. Scoring was done during the day after the study was completed and, as noted, each page was examined not only for sleep stage, but for changes in other measures including respiration, oxygen saturation, and limb movements. These events, along with stages, were manually tabulated and summarized. For clinical studies, each record was reviewed by the polysomnographer or sleep specialist and a report was composed that summarized the major diagnostic aspects of the study. For research purposes, variables of interest were entered into a data base for later analysis.

The development of digital computers (in particular the PC), their general availability, and the evolution of software for these systems specifically designed for the collection, storage, and manipulation of sleep data have played a significant role in the growth of the field of sleep medicine, and insights into sleep, in general. The vast amount of data collected from a full night sleep study can now be rapidly consolidated, allowing for comprehensive visualization of the measures obtained. Summary statistics pertaining to the sleep data recorded during the night are also obtained and easily accessed with the commercial computer data acquisition and analysis systems available for sleep studies. For example, once data are collected in an all-night study, the study can be examined over a variety of epoch window sizes from 10 s up to the thousands of seconds required to represent the entire study as one epoch. The ability to change the epoch window width allows for evaluation of different waveforms (e.g., EEG, respiration) using an optimal time scale. Further examples of sleep analysis software capabilities are presented below. By using such sleep systems, many laboratories that focus on the diagnosis of sleep disorders are now able to study as many as 6–10 patients in a single night.

COMPUTER ANALYSIS OF POLYSOMNOGRAPHIC DATA

Stages of Sleep

Sleep can be dichotomized into REM sleep and nonrapid eye movement (NREM) sleep (11). The NREM sleep can be

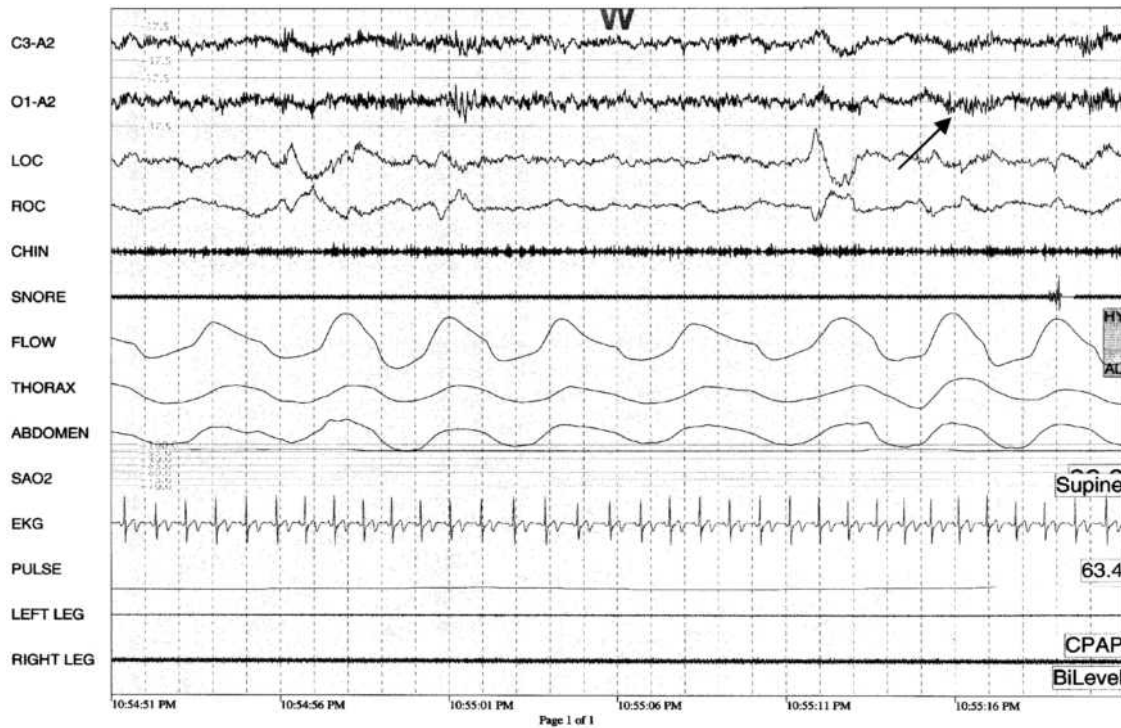


Figure 1. Illustration of Stage 1 sleep in Patient C. Epoch width and recording channels are the same as those in Fig. 1. Note the slowed eye movements (arrow).

further divided into four stages, according to rules established by Rechtschaffen and Kales (R&K) (10). The following outlines these stages and summarizes the major criteria that characterize them. *Stage 1 Sleep, illustrated in Fig. 1*, consists of a relatively low voltage EEG, with activity prominently from 2 to 7 cycles per second (Hz). Waves at this frequency are referred to as theta waves (12). Rapid eye movements are not present and eye movements are generally slow and rolling. This stage typically occurs during the transition from wakefulness (seen in Fig. 2) to the other sleep stages, and is relatively short. *Stage 2 Sleep, illustrated in Fig. 3*, is marked by K complexes and sleep spindles on a background of relatively low voltage EEG. The K complexes are negative, sharp, high voltage waves that are usually followed by a positive component. The complex duration, according to the R&K rules, should not exceed 0.5 s. The K complexes can occur spontaneously or in response to sudden stimuli and they are highest in amplitude over the vertex scalp regions (10,11). Sleep spindles are defined as “bursts of waves having a frequency of 12 to 15 [Hz]” (11, p. 16). In addition to the K complexes and sleep spindles, Stage 2 sleep lacks waves of high amplitude and slow activity seen in later stages (10). *Stage 3 Sleep, illustrated in Fig. 4*, is defined by delta waves, or slow waves of 2 Hz or less and amplitudes $>75 \mu\text{V}$ (11,12). The EEG activity in Stage 3 should account for 20–50% of the time interval (epoch) studied (10). Sleep spindles may or may not be present. *Stage 4 Sleep, illustrated in Fig. 5*, is defined when $>50\%$ of the epoch consists of high amplitude waves that are 2 Hz or slower.

Stage REM, illustrated in Fig. 6, typically occurs after the first 70–100 min of sleep. Rapid eye movements characterize this stage of sleep. The EEG during REM sleep is similar to a waking or Stage 1 EEG, with low amplitude mixed frequency waves. If eye movements and EEG activity were the only available data by which to characterize this sleep stage, one would suspect that the person was likely awake. It is because of the similarity to wakefulness that REM sleep has often been called paradoxical sleep. Other significant features of REM are as follows: saw tooth waves that may or may not be present in conjunction with bursts of REM activity, the lack of sleep spindles, and mental–submental EMG that is almost always lowest during REM than that seen for other stages of sleep. After REM sleep, there is generally a cycling through Stages 2, 3, and back to 4. These cycles continue throughout the night (11).

Computer Identification of Sleep Stages

Manual scoring of the various sleep stages is a tedious and time-consuming task, with interrater reliability between 67 and 91% (13). As such, both research and clinical laboratories have explored the use of computers to objectively classify these stages. Initial attempts at automating sleep analysis concentrated on programming computers to identify sleep stages according to the R&K rules. However, newer methods have focused on pattern recognition of distinct sleep stages (13,14).

One method, the polysomnogram assay (PSGA), was developed by Bartolo et al. (9), and condenses the large

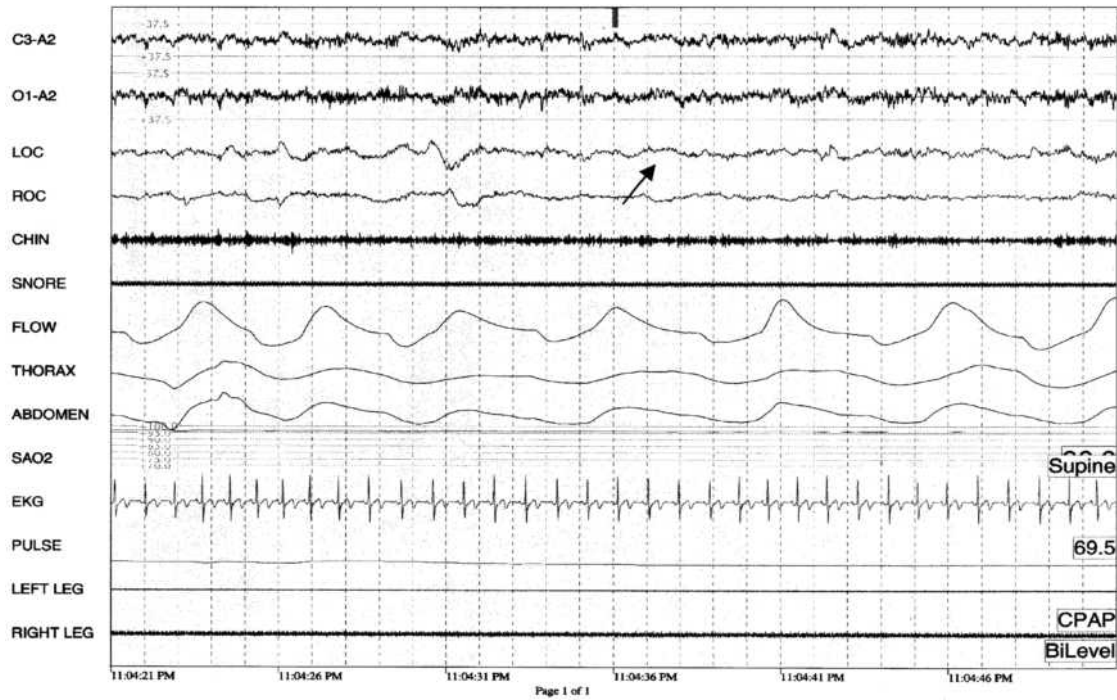


Figure 2. Example of a 30 s epoch of wakefulness from patient C being evaluated for a sleep disorder. The labels on the left indicate the measures at each channel of recording. The first two channels are EEG from the left central and left occipital scalp sites. LOC and ROC = left and right eye movements followed by chin EMG, nasal/oral air flow, thorax and abdominal respiratory effort, oxygen saturation, electrocardiogram, average pulse, left and right leg movement. Labels on the right indicate the patient's body position, average oxygen saturation (not shown), average pulse rate, level of continuous positive or bilevel air pressure (CPAP/BIPAP), if administered.

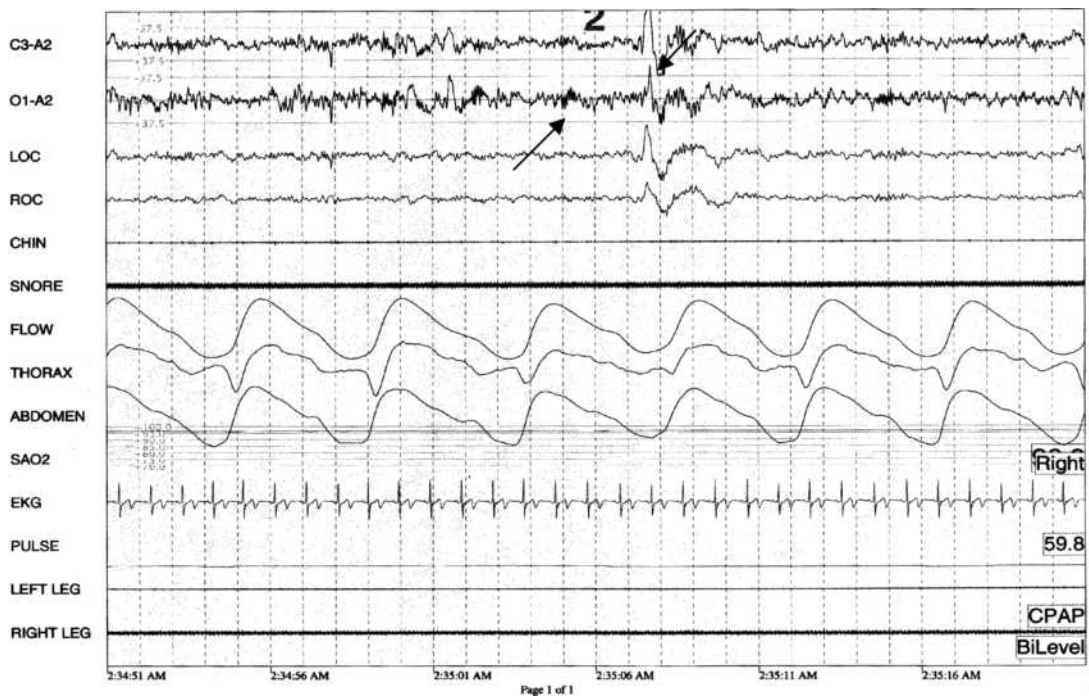


Figure 3. Illustration of Stage 2 sleep in patient C. Epoch width and recording channels are the same as those in Fig. 2. Note the K complexes (arrow Channel 1) and sleep spindles (arrow Channel 2).

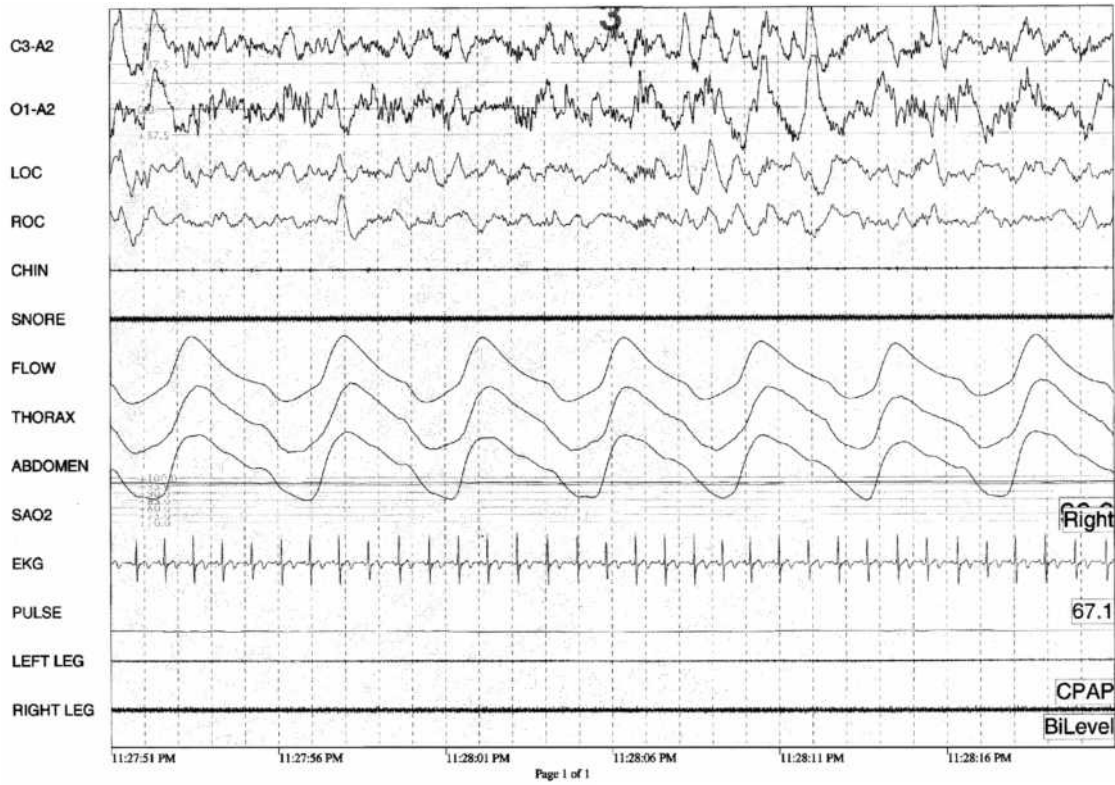


Figure 4. Illustration of Stage 3 sleep in patient C. Epoch width and recording channels are the same as those in Fig. 2. Note the high amplitude slow EEG waves and the sleep spindles in EEG channels 1 and 2 (arrow).

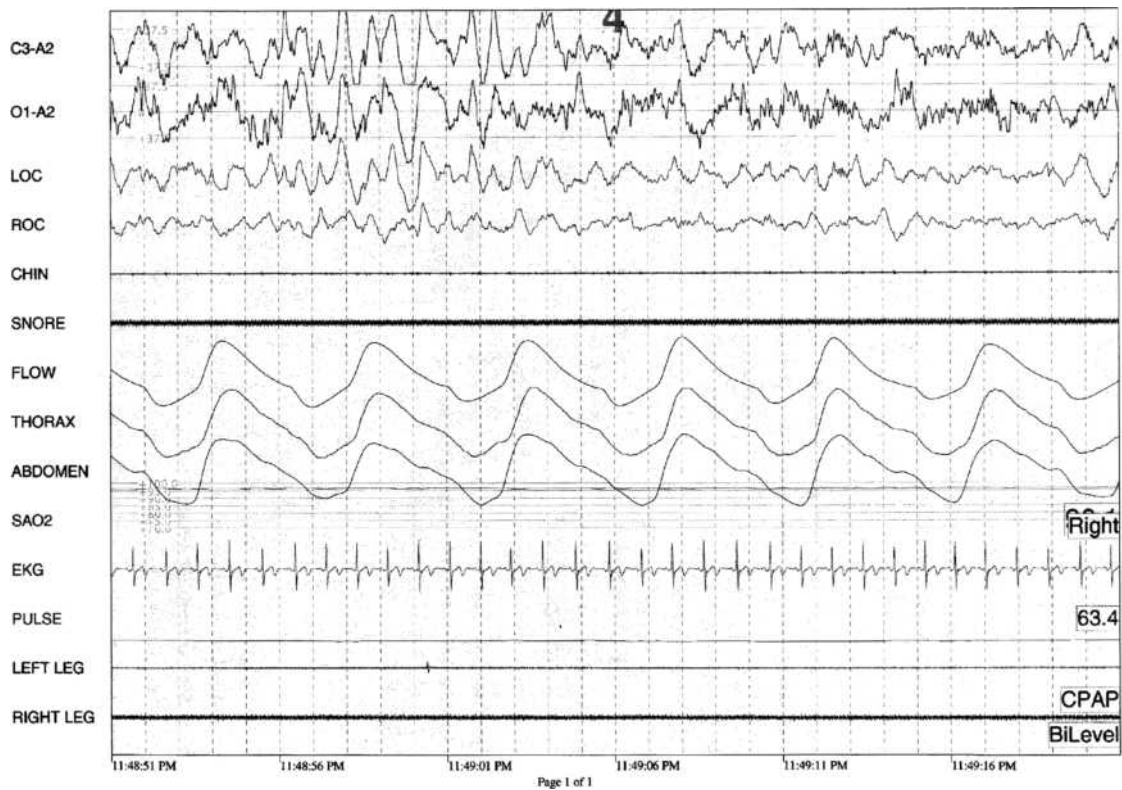


Figure 5. Illustration of Stage 4 sleep in patient C. Epoch width and recording channels are the same as in Fig. 2. Note the high amplitude slow EEG waves in EEG channels 1 and 2 (arrow). This slow activity occupies >50% of the epoch.

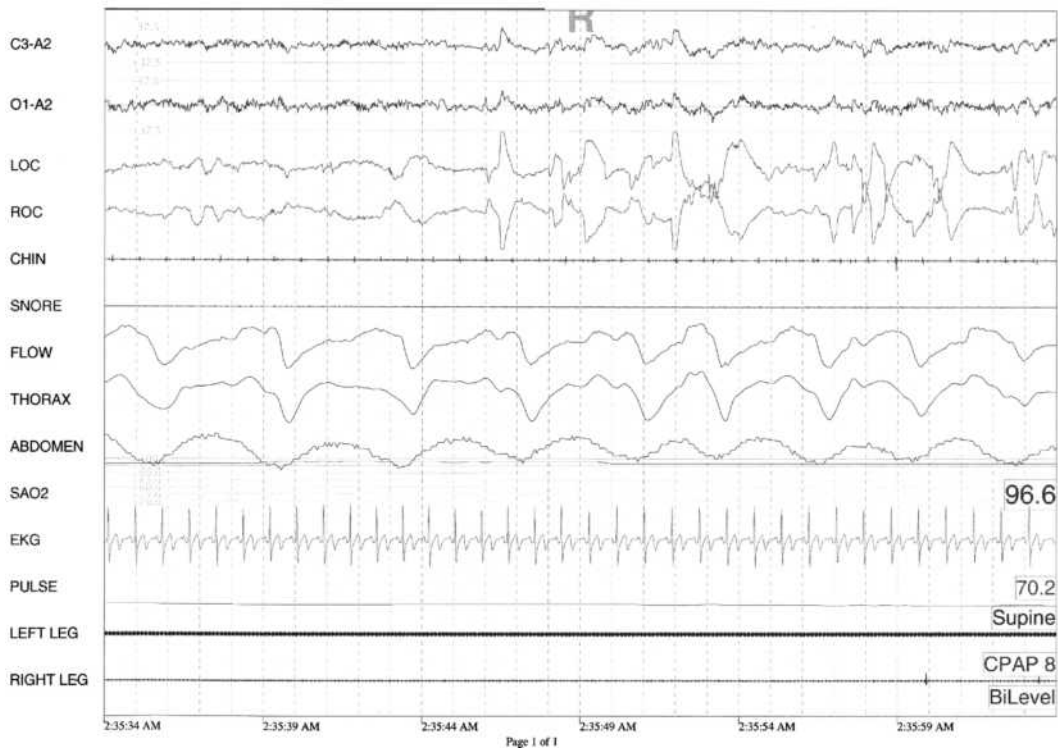


Figure 6. Illustration of Stage REM sleep in patient G. Epoch window and recording channels are the same as in Fig. 2. Note the low voltage EEG in EEG channels 1 and 2, the rapid eye movements in eye channels 3 and 4 (arrow), and the low amplitude chin EMG in channel 5.

amount of PSG data into a more comprehensible format while retaining events of diagnostic significance. The program allows gross distinction among sleep stages without manually scored data. However, because features associated with each sleep stage vary among individuals, the investigators noted that “it is unlikely that adequate sleep staging in the traditional sense can be accomplished by the PSGA or any other method except by visual analysis of the PSG itself” (p. 119). Nonetheless, the PSGA is able to present consistent patterns related to different sleep states, and with refinement may be useful for sleep staging. Two examples of how the data are consolidated are seen from the descriptions of leg and eye movements. Leg jerks appear as “abrupt peaks” (p. 121) in the EMG window assigned to leg movements. The EMG variables are represented by signal intensity or power. According to these investigators, periodic leg movements are more visible in the PSGA than in the PSG record. Additionally, the PSGA replaces the PSG EOG signals by markers that indicate the presence and intensity of eye movements. Thus, the PSGA transforms the data into events or markers appropriate for each type of measurement and displays these measures in a highly condensed format, allowing for up to 15 min of PSG data to be displayed on a single page. This method reduces the output from nearly 1000 pages to ~30. The purpose for the development of the PSGA was to “reduce the considerable human effort involved in scoring and evaluating PSG studies” (p. 124). This approach exemplifies how computers can assist in analyzing sleep, maximiz-

ing the amount of time polysomnographers can have for diagnosis and understanding unusual aspects of the sleep study that computers can only identify.

Agarwal and Gotman (13,15) described a computer assisted sleep staging method, which searches for clusters of wave patterns that have been identified and programmed as prototypes of waves or sleep stages. In this method, the PSG is first divided into multiple segments, containing at least 3 s of data. Second, prespecified features, such as sleep spindles and eye movements, are extracted by the computer from the PSG. Also, each epoch is examined for the different types of waves, including alpha and slow waves (delta, theta). After the data have been segmented and extracted, they are then “clustered” into groups with similar properties. These clusters can then be classified as particular sleep stages according to R&K or any other system preferred by the user. It appears that the approach taken by Agarwal and Gotman (13,14) relies on a self-organizational scheme that is based on relative differences in activity within a subject rather than fixed rules. So, for example, according to these investigators, stage REM sleep may be lost if hard thresholds for detection of atonia are used. This is the first study that has applied “self-organization[al] techniques to generate a hypnogram for an all night PSG in the context of R&K classification” (p. 1419). Agreement between automated staging and manual scoring according to R&K standards was 80.6% for 12 sleep records. This process uses the computer to initially analyze the data and identify various sleep

stages before it is inspected manually by a polysomnographer. Since this method is not fully computer automated, but is computer assisted, the investigators referred to this approach as "computer-assisted sleep staging (CASS)."

Because sleep stages are not distinct entities, computer errors are made. Empirical studies confirm that, despite our best effort to demarcate sleep stages for computer detection, errors occur, especially for the less clearly defined stages (13). For example, Agarwal and Gotman (13) reported that many errors occur when attempting to define Stage 1 sleep, perhaps the most ambiguous, which was often misclassified as either wakefulness or Stage 2 sleep. For computers to identify components of sleep, these components must be clearly defined. At present, the best defined features of sleep are the various stages according to the R&K rules. Some of these rules are specific enough for an algorithm to be programmed so that a computer can identify it on a polysomnogram. For example, Stage 3 sleep is defined by delta waves with amplitudes $>75 \mu\text{V}$, present between 20 and 50% of the time. Other definitions are less precise, such as the definition of Stage 2 sleep, which is marked by sleep spindles, K-complexes, and waves that lack the amplitude and slower activity seen in sleep Stages 3 and 4. Stage 2 sleep is more difficult to quantify and, therefore, less amenable to detection by computers.

Computer Analysis of Sleep Microstructure

From a more academic, less clinical, perspective, computers have been used to assist with analysis of the microstructure of sleep, such as parsing out the various components of specific waves (e.g., sleep spindles, K complexes) and exploring their relationships to various sleep problems. Although the presence of sleep spindles in Stage 2 sleep is well known, the different types of spindles and their subsequent spatial distribution are less well studied. With the aid of computers, two types of spindles have been identified: 12 spindles per second located primarily in the frontal region and 14 spindles per second located primarily in the centroparietal region (14). Our understanding of K complexes has also been improved with the aid of computers. Specifically, a classification of K complexes has been recommended, that includes K0 complexes (without sleep spindles), K1 complexes (spindles preceding the K complexes), K2 complexes (spindles occurring simultaneously with the K complexes), and K3 complexes (spindles following the K complexes (14)).

Computers have also been used to assess new components within the stages of sleep (e.g., slow wave sleep and REM sleep). Initially, slow wave sleep was defined as sleep marked by delta waves, occurring during stages 3 and 4. Delta waves were believed to be uniform, with no variable components. However, with the aid of computers, two types of these waves have been identified with scalp topographical differences. Sinusoidal $1\text{--}2 \text{ s}^{-1}$ delta waves tend to be located frontally, and have been compared to K complexes that are not precipitated by sensory stimuli. Polymorphic $<1 \text{ s}^{-1}$ delta waves, on the other hand, are located parietotemporally. The role of these components is presently unclear. Our understanding of the microelements of REM sleep is also being explored. For instance, although saw-

tooth waves have been linked to REM sleep, their function is poorly understood (14). Only since the advent of computers have such microelements of the various stages of sleep been able to be identified, and is it only possible to study these components with computers in order to understand their functions and their possible contribution to sleep disorders.

A PROTOTYPICAL COMMERCIAL SLEEP ACQUISITION, ANALYSIS, AND MANAGEMENT SYSTEM

Although, computers do not eliminate errors in scoring sleep stages or various waves, they do substantially increase our ability to measure more aspects of sleep and to organize and reduce the data so that they are more readily interpretable. Digital polysomnography has become an important means of sleep analysis today. An important advantage is probably that the tracings can be zoomed in and out. While the EEG is best displayed on a high resolution monitor with 10–15 s per page (s/p) and 20–30 s/p is sufficient for sleep staging, respiratory-related signals and body movements can best be recognized at 2–10 min/p. The correlations between slight respiratory changes and EEG arousals are sometimes best observed with a 2 min page. Thus events that often remain undetected by the conventional paper method can be visualized. Consequently, the paper method is not an optimal method or a good alternative for a "gold standard". The possibility of adjusting the gain off-line is also valuable. However, the low dynamic range provided by many manufacturers set unnecessary limits. Eight-bit A/D conversion is still used in many systems, which gives a resolution of only 256 points. The introduction of 24 bit resolution in modern equipment has allowed for same amplifiers and gains to be used for practically all signals.

Computers also can provide summary statistics from data obtained over the entire night of recording. One frequently used commercial computerized sleep system allows up to 64 channels of recording. Computer algorithms are available that will detect significant changes in EMG, EEG, respiration, and heart rate. These programs are dependent on people to set criteria based on specific standards of what constitutes an event, such as an arousal, an obstructive apnea, hypopnea, a bradycardia, and a tachycardia.

In order to provide the reader with an appreciation of the state of the art in sleep study computer technology, one of the systems used by many sleep laboratories, including our own will be described next. The software runs on most PCs and is designed to collect, store, analyze, summarize, and manage the large amount of physiological data obtained during an all night sleep study. This system, with its appropriate amplifiers, digitizes the electrical voltages obtained from the various sensing devices, such as electrodes applied to the scalp, face, and chest. After these data are collected and saved, the software provides a number of analytic tools and modules that allow the polysomnography technician/polysomnographer to reduce and quantify the data. As described by the manufacture (Sandman Sleep Diagnostic Systems), the software consists of four main

components: (1) *Data Collection*, for recording and saving the signals obtained from the patient. (2) *Analysis*, for scoring and analyzing the data saved from in the patient file. Computer assisted scoring modules are part of this component. (3) *Data Management*, used for file manipulation such as copying, backing up, and deleting patient data. (4) *Configuration*, provides the user with the ability to manipulate some of the features of the software, change the storage media, and so on.

Figures 2–6 illustrating the stages of sleep are examples of the data obtained in our laboratory with this system. These figures represent epochs or time periods of the actual, data collected during an all night polysomnogram. Figure 7 presents sleep data from a patient with obstructive sleep apnea. Apnea is characterized by a cessation of airflow for 10 s or longer. Obstructive sleep apnea in contrast with central sleep apnea is caused by an actual obstruction of the airway and is characterized by a cessation of air flow, arousal, and increased effort to breathe (Fig. 7). Hypopnea is also an obstructive event, but the obstruction is not complete. It is generally defined by a reduction, without complete cessation, in airflow or effort (16). There are several criteria that laboratories have used to define an event as hypopnea. For example, with respect to air flow, both a 50% and 30% decrease from baseline air flow have been used. The EEG arousals and amount of oxygen desaturation have also played a role in the definitions of obstructive respiratory events. These criteria, used to define respiratory events, are all entered into the sleep scoring program so that the computer using the appropriate definition can select these events automatically. In

addition, a polysomnographic technician reviews the entire sleep study and manually stages and confirms the relevant events. Once stages and events are reliably identified, a number of calculations are made via computer algorithms that summarize the information obtained over the sleep study. Figure 8 is an illustration of how data obtained over an entire night of sleep are summarized.

Software programs can now provide a report format that consolidates the information from the entire 6–8 h study once criteria are specified for relevant events by a sleep technician or polysomnographer. Such information can include measures of sleep efficiency, percentage of time in various sleep stages, the number of respiratory disturbances, the number of arousals, and other aspects of sleep, in order to accurately diagnose sleep disturbances. Figures 9 and 10 present a numerical summary and graphs of the data obtained for a patient being evaluated for sleep apnea. The system allows each laboratory to customize their own summaries and graphs with those variables they wish to send to the clinician along with the diagnostic interpretation. Note in Fig. 9 the patient’s respiratory disturbance index (RDI) during REM sleep is 19.6 h^{-1} . Thus, this patient, during their total time in REM sleep, had an average of 19.6 respiratory events in one hour of REM.

As indicated, all of these processes are semi-automated and under the watchful eyes of trained polysomnography technicians and sleep specialists. The degree of automation is dependent on the reliability of the measure. As would be expected, sleep staging is less automated than the detection of awakenings or arousals. The most automated

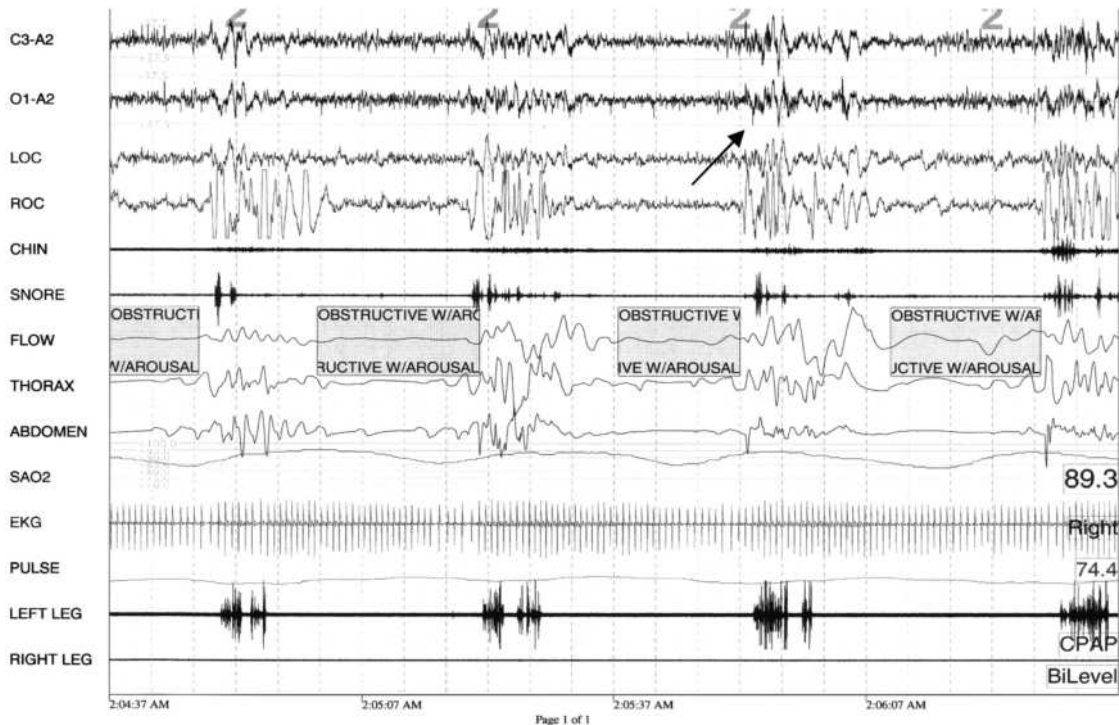


Figure 7. Illustration of obstructive sleep apnea in patient M. Epoch window is 120 s. Recording channels are the same as in Fig. 1–6. Arrow indicates EEG arousal following the event.

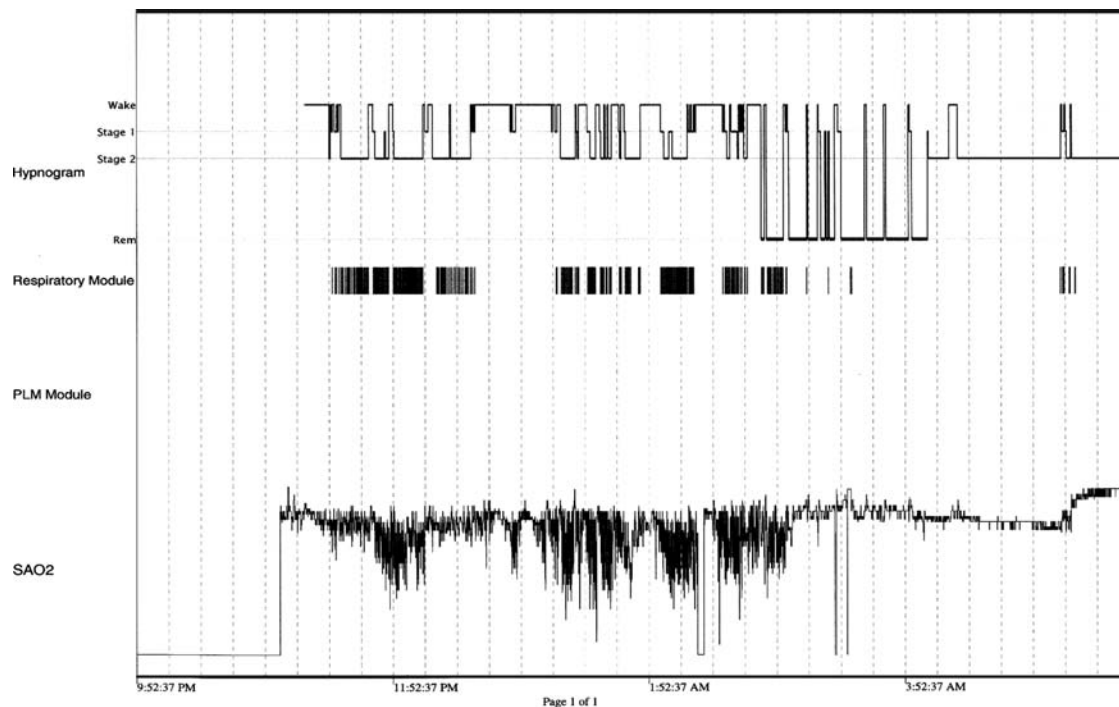


Figure 8. Illustration of computer summary of nocturnal polysomnogram for patient M. Note the oxygen desaturations (SaO_2) associated with respiratory events (Respiratory Module). Periodic limb movements (PLM Module) were not present.

processes are those that involve specific calculations derived from the data collected during the sleep study. For example, the calculation of the overall RDI is the total number of respiratory events that occurred over the entire sleep study divided by the number of hours of sleep.

A tool available on the system, more for research than clinical evaluation, is the fast fourier transform (FFT). FFT is a mathematical analysis of a wave form that derives a frequency spectrum over time. The software displays a signal as a function of amplitude on the y axis versus time on the x axis. The module will perform this transformation on data of interest and write the results to a specified file. Frequency spectral analysis is also available for data reduction of EEG activity. For example, sleep spindles and EEG activity, such as alpha, beta, delta, and theta, can be summarized by plotting the signal amplitude on the y axis versus the frequency of the signal on the x axis. However, the utilization of fixed frequency bands can give misleading results. The bands can be too wide, too narrow, or displaced. Alpha activity of drowsiness and sleep is often on the border between alpha and theta (17) and the frequency of the sigma spindles can have a frequency in both the alpha and the beta range (18). Ideally bands should not be preselected. Rather, bands should be adjusted according to the signal. Martens et al. (19) applied matched filtering based on information obtained by FFT for this purpose. Another attempt to solve the problem is to study very narrow frequency bands and their relationships (20).

The accuracy of frequency analysis is also dependent on sampling rate. Often frequency is understood as an equivalent to the inverse of the baseline-crossing interval of a

wave. However, in spectral analysis it is defined by the frequency content of the signal. With FFT the Nyquist theorem states that the sampling rate has to be at least twice the frequency of the fastest sine waves. However, EEG rhythms are often not sine waves and thus include frequency components that are both faster and slower than the frequencies calculated on the basis of their baseline-crossing intervals. Thus, in order to preserve the form of the waves, higher sampling rates have to be used. Also, the accuracy of baseline-crossing is dependent on the sampling rate. Thus, if the maximum frequency of a sigma spindle is 16 Hz, then a sampling rate of 1024 Hz is required in order to obtain an accuracy of at least 0.25 Hz.

LIMITATIONS OF THE USE OF COMPUTERS TO ANALYZE SLEEP

Despite the overwhelming contribution of computers in analyzing and furthering our understanding of sleep, limitations exist. Such limitations are expected when computers are relied on to explain a complicated, biological process. In fact, both intra- and interindividual variability contribute to the challenge of quantifying components of sleep, such as sleep stages or waveforms. The separation of wakefulness and stage REM is, in some systems, very dependent on the quality of the EMG recording. Because separation of REM and wakefulness is often difficult due to the considerable variation in the EMG activity level obtained for each stage, human supervision and adjustment of detection levels are necessary in practice.

SLEEP STUDY SUMMARY

Patient Name		Study Date		Subject Code	
D.O.B.		Height		Ref. Physician	
Sex		Weight		Scorer	
Sleep Architecture					
Lights Out:	10:54:42	Lights On:	05:48:12		
Total Recording Time:	413.5	Total Sleep Period:	396.0	Total Sleep Time:	341.0
Sleep Latency:	17.5	REM Latency:	122.0	Sleep Efficiency:	82.5
# REM Periods:	3	# Stage Shifts:	102	Awakenings:	34
Sleep Stage as % TST:					
Stage 1:	14.1 %	Stage 4:	0.1 %		
Stage 2:	46.8 %	REM:	26.1 %		
Stage 3:	12.9 %	MVT:	0.0 %		
Body Positions Slept:					
(%TST)	89.6 %	(%TST)	10.4 %	(%TST)	0.0 %
Supine:		Right:		Left:	Prone
Respiratory Analysis:					
	NREM	REM	TOTAL	INDEX	
Central Apneas	0	1	1	0.2	
Obstructive Apneas	0	7	7	1.2	
Mixed Apneas	0	0	0	0.0	
Hypopneas	41	21	62	10.9	
RERAs	0	0	0	0.0	
Respiratory Events	41	29	70	12.3	
RDI					
NREM			REM		
9.8			19.6		
Supine Events					
No. of Events	70		Non-Supine Events	No. of Events	0
Index	13.7		Index	0.0	
Oxygen Analysis:					
	Awake	NREM	REM	TRT	
Mean SaO2 (%)	96.2	95.3	95.5	95.5	
Min. SaO2 (%)	92.6	92.6	91.7	91.7	
Max. SaO2 (%)	99.0	98.0	97.1	99.0	
SaO2 (%TST)					
100-90	99.8 %	90-80	0.0 %	80-70	0.0 %
				70-60	0.0 %
				60-50	0.0 %
				<50	0.0 %
Movement Analysis:					
	NREM	Index	REM	Index	TOTAL
Total Arousals	55	9.7	39	6.9	94
PLM's	63	15.0	30	20.2	93
PLM Arousals	1	0.2	8	5.4	9
Respiratory Arousals	41	9.8	29	19.6	70
Spontaneous Arousals	13	3.1	2	1.4	15
Min. Pulse B.P.M.					
57.9					
Mean Pulse B.P.M.					
75.6					
Max. Pulse B.P.M.					
176.1					
Systolic Pressure:					
Diastolic Pressure:					
Diagnostic Code:					

Figure 9. Illustration of a sleep study summary. The values are calculated by specific computer algorithms and represent, in this case, data obtained over ~7 h of recording. This summary accompanies the diagnostic interpretation of the study that is forwarded to the patient's clinician.

Computer analysis is also often dependent on measurement of the wakefulness alpha activity, which causes problems with low alpha subjects. In this case, it is an advantage to rely on theta activity and non-EEG waveforms (21,22). Also, the distinction between REM and NREM sleep is difficult in patients with poor sigma spindles or if the spindles are of a frequency outside the usual range (18). Toussaint et al. (23) discussed the first-night effect in sleep studies, noted by lower sleep efficiency, increased wakefulness, and longer REM periods in patients. Computer algorithms have been difficult to derive that can take into account the first-night effect in individuals because it is relative to the person's normal sleep rather than referring to objective, specific changes in sleep across all individuals.

Interindividual differences in sleep measures have also been demonstrated. For example, variability in polysomnograms has been associated with both gender (24-26) and

age (27,28). For example, Stage 4 sleep presents differently in a young child than it does in an adult (28). Older adults have lower amplitude delta waves than children. Although this is a broader issue regarding the definition of Stage 4 sleep across the age range, it is also illustrative of the difficulties in standardizing automated computerized sleep staging systems. A computer must be programmed to specifically identify such sleep variability for fully automated scoring to be accurate. Furthermore, studies also show marked variability between various patient populations, with different polysomnograms seen for patients with schizophrenia (29,30), autism (31), and attention deficit/hyperactivity disorder (32). Thus, one issue with computer automated sleep scoring is how automated can or should the process be? Computers can only accomplish what they are programmed to do (27), and errors will occur due to artifacts or unexpected events (15). This issue was highlighted in a study by Cirignotla et al. (33) in which they

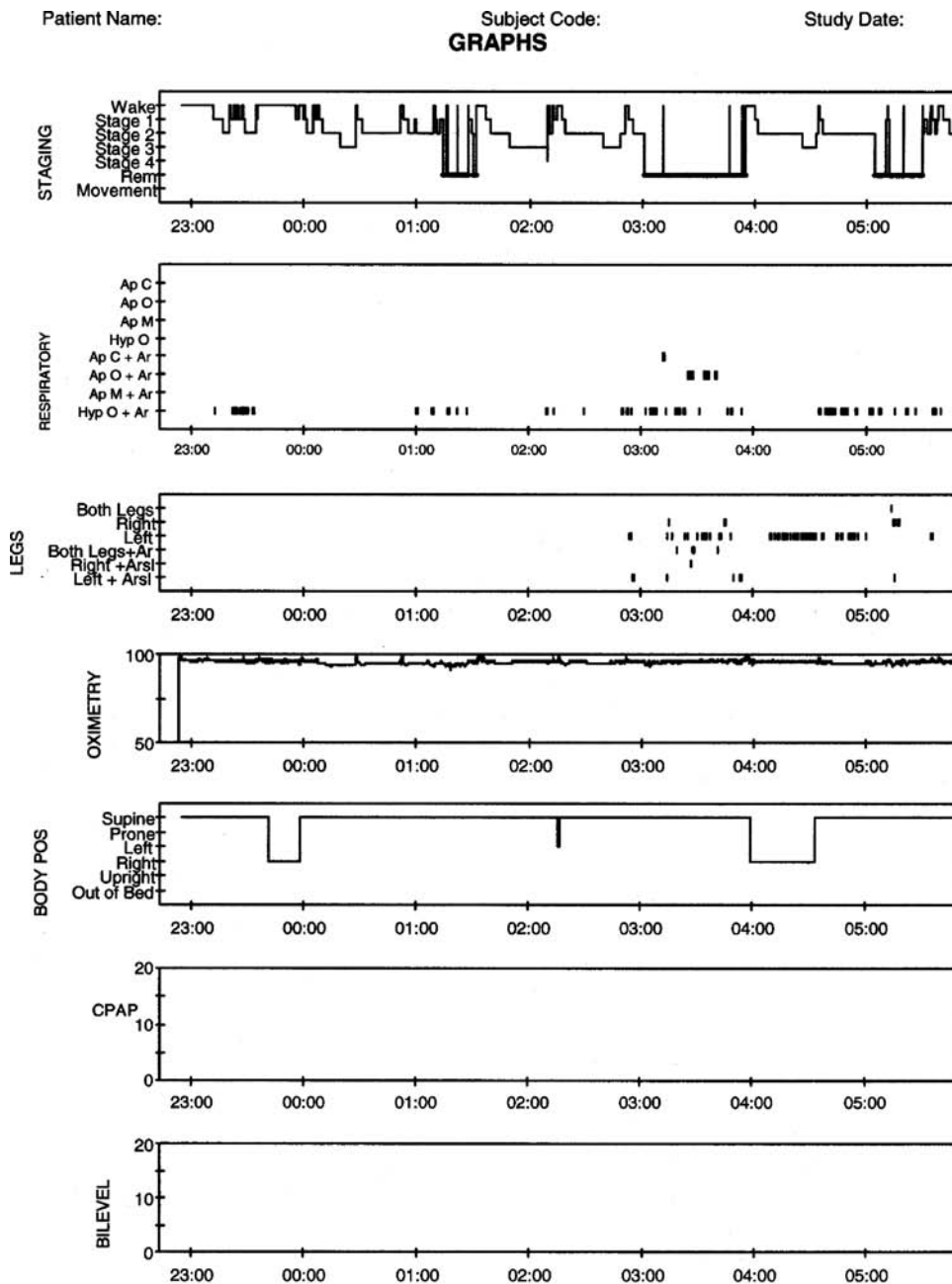


Figure 10. Illustration of graphs representing the numerical data presented in Fig. 9. These graphs are automatically plotted by the computer. The specific format is designed by the laboratory. Note that this patient, for example, has a significant positional component to their sleep apnea. As indicated in Fig. 9, the Supine RDI is 13.7 h^{-1} compared to a nonsupine RDI of 0.0 h^{-1} .

investigated the MESAM 4, a computerized, ambulatory cardiorespiratory monitor that can help diagnosis sleep apnea. However, the investigators found that this computerized system significantly underestimated the oxygen desaturation index in patients with complicated obstructive sleep apnea, or patients who suffered from sleep apnea along with another respiratory syndrome (e.g., chronic obstructive pulmonary disease). In other words, the MESAM 4 was unable to identify patients that did not fit the programmed definition of sleep apnea. This study highlights the current limitations of computers and the need for humans to be intimately involved in setting criteria for event identification and evaluating the reliability of the computerized measures obtained.

CONCLUSIONS

Advances in technology have contributed to major insights into the mechanisms and processes of sleep. However, computer programs can be written to recognize only what we understand and can clearly define (14). At present, reliable completely automated sleep scoring systems are not available. The EEG variables present particular difficulties for automated sleep staging systems. However, the ability of computers to store large amounts of data, to organize and consolidate these data, and to perform summary statistics has allowed the polysomnogram to become a significant clinical and research tool. It provides clinicians and scientists with increased power for the

evaluation and understanding of sleep pathology, as well as the basic mechanisms underlying sleep itself.

BIBLIOGRAPHY

- McNish R. *The Philosophy of Sleep*, 1st ed. New York: D. Appelton & Company; 1934.
- Kleitman N. *Sleep and Wakefulness*. Chicago: The University of Chicago Press; 1939.
- Aserinsky E, Kleitman N. Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science* 1953;118:273–274.
- Aserinsky E, Kleitman N. Two types of ocular motility occurring in sleep. *J Appl Physiol*, 1955;8:11–18.
- Dement W, Kleitman N. Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalogr Clin Neurophysiol* 1957;9:673–690.
- Hobson JA. *Sleep*. New York: Scientific American Library; 1989.
- Dement WC. Normal sleep and its variations. In: Kryger MH, Roth T, Dement WC, editors. *Principles and Practice of Sleep Medicine*, 2nd ed. Philadelphia: W.B. Saunders; 1994; p 3–15.
- American Electroencephalographic Society. Guideline fifteen: Guidelines for polygraphic assessment of sleep-related disorders (polysomnography). *J Clin Neurophysiol* 1994;11:116–124.
- Bartolo A, Clymer BD, Golish JA, Burgess RC. The polysomnogram assay: a method to represent the overnight polysomnogram in a condensed format. *Comput Biomed Res* 2000; 33:110–125.
- Rechtschaffen A, Kales A. *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. University of California: Brain Information Service; 1968.
- Siegel J. *The neural control of sleep and waking*. New York: Springer-Verlag; 2002.
- Hauri PJ. *The Sleep Disorders*. Kalamazoo (MI): The Upjohn Company; 1982.
- Agarwal R, Gotman J. Computer-assisted sleep staging. *IEEE Transactions on biomedical engineering* 2001;48: 1412–1423.
- Kubicki S, Herrmann WM. The future of computer-assisted investigation of the polysomnogram: Sleep microstructure. *J Clin Neurophysiol* 1996;13:285–294.
- Agarwal R, Gotman J. Digital tools in polysomnography. *J. Clin Neurophysiol*, 2002;19(2):136–143.
- Moser NJ, Phillips B, Berry DT, Harbison L. What is hypopnea anyway? *Chest* 1994;105(2):426–428.
- Broughton RJ, Hasan J. Quantitative topographic electroencephalic mapping during drowsiness and sleep onset. *J Clin Neurophysiol* 1995;12:372–386.
- Jankel WR, Niedermeyer E. Sleep spindles. *J Clin Neurophysiol* 1985;2:1–35.
- Martens WLJ, Declerck AC, Kums DJThM, Wauquier A. Considerations on a computerized analysis of long-term polygraphic recordings. In: Stefan H, Burr W, editors. *EEG monitoring*. Stuttgart: Gustav Fisher; 1982; p 265–274.
- Badia P, Wright KP, Wauquier A. Fluctuations in Single-Hertz EEG activity during the transition to sleep. In: Ogilvie RD, Harsh JR, editors. *Sleep Onset: Normal and Abnormal Processes*. Washington (DC): American Psychological Association; 1995. p 201–218.
- Hasan J. Differentiation of normal and disturbed sleep by automatic analysis. *Acta Phys Scand(Suppl)* 1983;526:1–103.
- Hasan J, Hirvonen K, Värri A, Häkkinen V, Loula P. Validation of computer analysed polygraphic patterns during drowsiness and sleep onset. *Electroencephalogr Clin Neurophysiol* 1993;87:117–127.
- Toussaint M, et al. Changes in EEG power density during sleep laboratory adaptation. *Sleep* 1997;20(12):1201–1207.
- Carrier J, et al. The effects of age and gender of sleep EEG power spectral density in the middle years of life (ages 20–60 years old). *Psychophysiology* 2001;38:232–242.
- Ehlers CL, Kupfer DJ. Slow-wave sleep: Do young adult men and women age differently? *J Sleep Res* 1997;6(3):211–215.
- Huupponen E, et al. A study on gender and age differences in sleep spindles. *Neuropsychobiology* 2002;45(2):99–105.
- Hirshkowitz M, Moore CA. Issues in computerized polysomnography. *Sleep* 1994;17(2):105–112.
- Tan X, Campbell IG, Feinberg I. Internight reliability and benchmark values for computer analyses of non-rapid eye movement (NREM) and REM EEG in normal young adult and elderly subjects. *Clin. Neurophysiol* 2001;112:1540–1552.
- Keshavan MS, et al. Delta sleep deficits in schizophrenia: evidence from automated analyses of sleep data. *Arch Gen Psychia* 1998;55(5):443–448.
- Keshavan MS, Reynolds CF, Miewald JM, Montrose DM. A longitudinal study of EEG sleep in schizophrenia. *Psychiat Res* 1996;59:203–211.
- Limoges E, et al. Atypical sleep architecture and the autism phenotype. *Brain* 2005;128(Pt5):1049–1061.
- Barry RJ, et al. Age and gender effects in coherence: II. Boys with attention deficit/hyperactivity disorder. *Clin Neurophysiol* 2005;116(4):977–984.
- Cirignotta F, et al. Unreliability of automatic scoring of MESAM 4 in assessing patients with complicated obstructive sleep apnea syndromexd *Chest*. 2001;119(5):1387–1392.

See also ANALYTICAL METHODS, AUTOMATED; ANESTHESIA, COMPUTERS IN; SLEEP LABORATORY.

SPECT. See COMPUTED TOMOGRAPHY, SINGLE PHOTON EMISSION.

SPECTROFLUORIMETRY. See FLUORESCENCE MEASUREMENTS.

SPECTROPHOTOMETRY. See COLORIMETRY.

SPEECH REHABILITATION. See LARYNGEAL PROSTHETIC DEVICES.

SPINAL CORD STIMULATION

TERRENCE L. TRENTMAN
KENT P. WEINMEISTER
Mayo Clinic Scottsdale
Scottsdale, Arizona

INTRODUCTION

Spinal cord stimulation (previously known as dorsal column stimulation) is a minimally invasive technique used primarily to treat chronic, refractory neuropathic pain. It is

based upon Melzack and Wall's gate control theory (1), and was first introduced by Shealy in 1967 (2). Neuropathic (nerve injury) pain has many etiologies, including trauma, stroke, diabetes, infection [e.g., human immunodeficiency virus (HIV), or shingles], and cancer. Unfortunately, nerve injury pain can be extremely difficult to manage. Many types of therapy have been used for neuropathic pain including medications such as antiinflammatories, opiates, and antiepilepsy drugs. Physical therapy and psychologically based approaches have also been tried with variable success. Spinal cord stimulation (SCS), transcutaneous electrical nerve stimulation (TENS), and peripheral nerve stimulation (PNS) are all forms of neuromodulation that are used for nerve injury pain.

Spinal cord stimulation is typically reserved for patients with refractory neuropathic pain, whereas deep brain stimulation is currently used for patients with movement and some pain disorders. In SCS, a lead is percutaneously inserted into the epidural space, and an electric field is applied in the vicinity of the spinal cord. The electric field depolarizes neural elements or in some way modifies the function of the nervous system. The goal is for the patient to experience a pleasant paresthesia, often described as "tingling", in the area of their pain. After an initial successful trial, a permanent stimulator can be implanted that the patient controls with a hand-held device.

This article reviews the equipment used in spinal cord stimulation, patient selection, and the possible mechanisms of this therapy. The process of inserting a stimulator will be described with possible complications, and the effectiveness of this therapy will be analyzed. Finally, possible future uses of SCS will be discussed.

EQUIPMENT

Medtronic, Inc. (Minneapolis, MN), Advanced Neuromodulation Systems, Inc. (Allen, TX) and Advanced Bionics (Valencia, CA) are the primary manufacturers of spinal cord stimulators. There are two types of implantable leads available: the paddle (surgical) lead and the tubular percutaneous lead (see Fig. 1). The percutaneous lead can

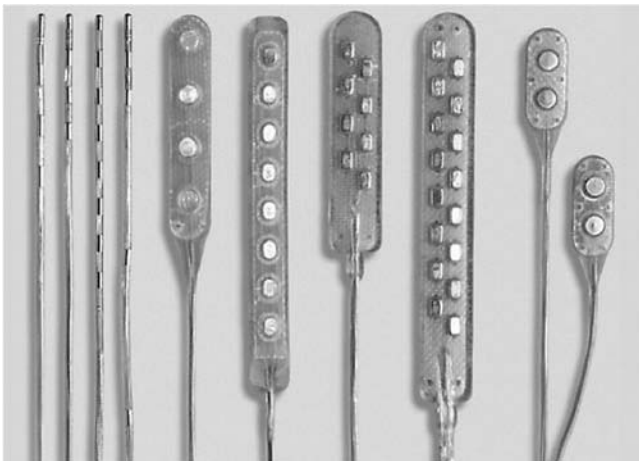


Figure 1. Various percutaneous and paddle leads. (Used courtesy of ANS, Inc.)

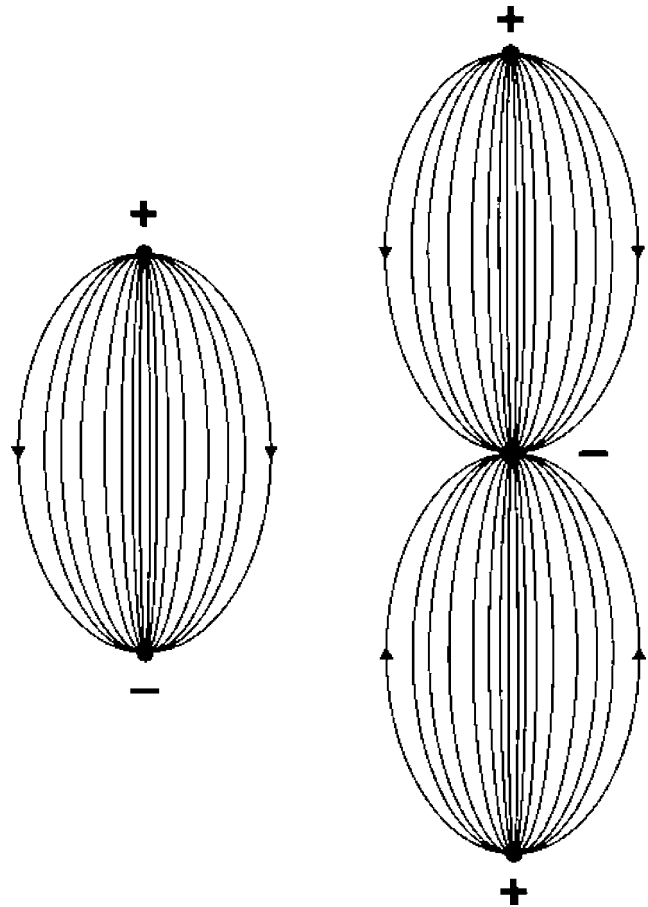


Figure 2. Current distribution between electrodes on SCS leads.

have four or eight contact points (electrodes), whereas the surgical leads are available with two, four, or eight electrodes. Each electrode can be programmed to function as an anode or cathode for the electrical current used in stimulation (see Fig. 2). The paddle lead is shielded on one side, such that stimulation is produced only on the side with the electrodes (see Fig. 3). This finding has the advantage of directing the entire electrical field toward the spinal cord, as opposed to the percutaneous lead that produces an electrical field circumferentially around the lead, including away from the spinal cord. Hence, the paddle lead can produce SCS at lower amperage, prolonging battery life. The surgical lead also has the potential advantage of greater stability (less likely to move postimplantation) as it is sutured to surrounding tissue (3). However, the paddle lead requires a minilaminotomy, whereas the percutaneous lead is placed less traumatically via a 15 gauge touhy needle.

The percutaneous lead is made of inert polyurethane with an outside diameter of ~ 1.3 mm. On its distal end, it has four or eight electrodes made of platinum iridium. These are spaced 4, 6, or 12 mm apart. The electrodes are 3–6 mm long. The plate lead has a two- or four-midline circular or eight parallel rectangular electrodes. There are several options to provide current to the electrodes. An implantable pulse generator (IPG) can be placed subcutaneously (usually in the low abdomen), similar to a

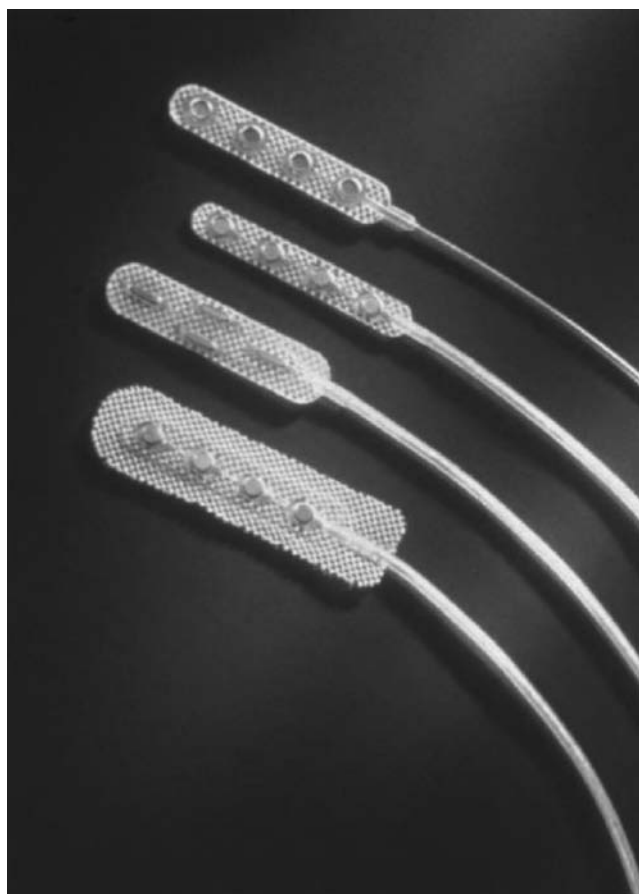


Figure 3. Examples of surgical (paddle) leads. (Used courtesy of Medtronic, Inc.)

pacemaker generator. Recently, a rechargeable IPG has become available. Depending on use, the Pt will percutaneously recharge the IPG every few days to weeks. Finally, a radio frequency (rf) receiver can be placed subcutaneously and powered by an external rf transmitter coil that is held over the device (see Fig. 4). In either case, a cable is tunneled subcutaneously from the power source to the lead in the spine.

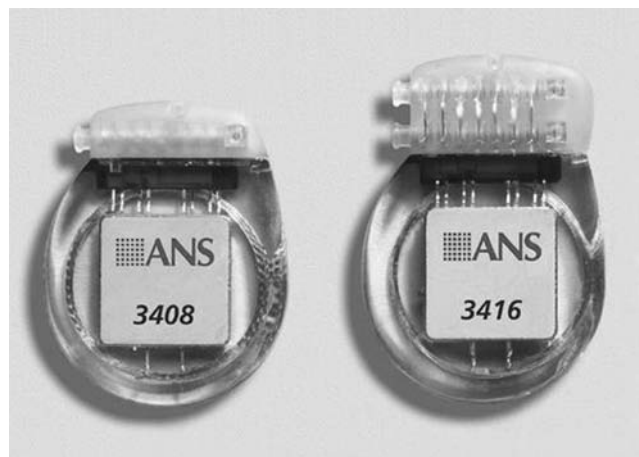


Figure 4. Implantable rf receivers. (Used courtesy of ANS, Inc.)

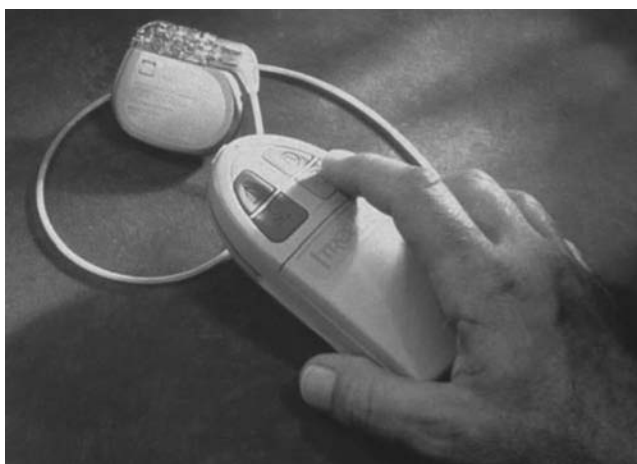


Figure 5. Implantable IPG with hand-held controller. (Used courtesy of Medtronic, Inc.)

The IPG provides pulses of electrical current that are rectangular and biphasic. Programmable features include pulse width, amplitude, and rate. Rates generally used are 30–80 Hz, amplitude is 0–12 V, and pulse width is 200–450 μ s. Battery life varies depending on rate, amplitude, and time the device is used. Unless the patient uses the stimulator constantly with high rate and amplitude, the battery should last at least several years. One obvious advantage of an rf receiver system is that there is no need to periodically replace the battery. The energy source of the IPG is a hermetically sealed silver vanadium oxide cell. The power source and electronics are sealed in an oval shaped titanium shield (see Fig. 5).

PATIENT SELECTION

As noted earlier, spinal cord stimulators are used primarily for neuropathic pain. Although there are many types of neuropathic pain, chronic unilateral lower extremity neuropathic pain seems to respond best to this type of therapy. A typical patient may have the failed back surgery syndrome with residual leg pain, or the patient may have some type of neural compressive lesion that is not operable and refractory to medical management. Spinal cord stimulation may also be indicated for chronic arachnoiditis, complex regional pain syndrome, peripheral neuropathy of the lower limb, and phantom limb syndrome. Patients with idiopathic pain, mechanical low back pain, or other forms of nociceptive pain have a lower success rate when compared to those with neuropathic conditions. Clinical experience and studies have shown that SCS is most efficacious when the entire painful area is covered with paresthesia. The more diffuse the patient's pain, the more difficult it will be to cover with SCS.

Selection for SCS often includes psychological screening. Patients with untreated depression, anxiety, or drug abuse issues are not good candidates. Obviously, the patient needs to be able to understand how to use the stimulator. A trial of stimulation is probably the best predictor of long term success (4). Although the value of

psychological testing in predicting success with SCS is controversial, there is no question that patients with chronic pain are best managed with a multidisciplinary approach. This may include physical therapy, psychological and spiritual support, medications, and surgical procedures as indicated.

MECHANISMS

Both animal and human research studies have provided a partial understanding of the mechanisms of SCS (5). Melzack and Wall's gate control theory (1) suggested that stimulation of large cutaneous A- β fibers would inhibit nociceptive input from the smaller A- δ and C fibers. Since SCS has been shown to be more effective for neuropathic pain than nociceptive pain, the mechanism must include more than simple inhibition of nociceptive input. Endorphins or other endogenous opiates do not seem to be involved. In patients with ischemic lower extremity pain or refractory angina, the mechanism of SCS appears to be an increased local blood flow (i.e., microcirculation). This may be due to both inhibition of the sympathetic nervous system and activation of vaso-active chemicals (6).

Animal studies have supported the contention that A- β fiber stimulation is one of the mechanisms of SCS. Animal models of neuropathic pain can be created by lesioning the sciatic nerve, which creates tactile allodynia in the animal, a phenomenon mediated by A- β fibers. Spinal cord stimulation has been seen to suppress this sign. Another effect of SCS is on wide-dynamic range neurons. Wide-dynamic range (WDR) neurons are second order neurons in the dorsal horn of the spinal cord. They receive input from a variety of sensory neurons. In the face of continuous stimulation from injured neurons, the WDR neurons will "wind-up," that is, fire at lower depolarization thresholds. Spinal cord stimulation may decrease this WDR response while simultaneously decreasing the central excitatory neurotransmitters glutamate and aspartate. γ -Aminobutyric acid (GABA), a central inhibitory neurotransmitter, is simultaneously released; therefore, SCS may have beneficial effects on both excitatory and inhibitory pain mechanisms (7).

Recent computer modeling of SCS has led to a greater theoretical and empirical understanding of the interaction of current with spinal structures (8). These models demonstrate how the depth of cerebral spinal fluid and the distance of the electrodes from both the dorsal columns and dorsal roots can affect the patient's paresthesia perception.

IMPLANTATION TECHNIQUE

Before a spinal cord stimulator is implanted, the patient needs to be informed of potential risks. These include infection, bleeding, nerve damage, allergic reaction, and failure of the stimulator to adequately cover or reduce the patient's pain. The patient may experience swelling around the site of the generator and a seroma may develop requiring drainage. If the lead or the generator becomes infected, it may have to be removed. Lead displacement, fracture, or movement can occur such that an initially adequate pat-

tern of stimulation becomes inadequate. Lead and battery revision may become necessary at some point. After placement of a SCS, the patient is instructed not to drive an automobile with the device turned on. Furthermore, they should not undergo a magnetic resonance imaging (MRI) scan or any type of diathermy.

Once consent has been obtained, a trial of SCS is performed. This consists of placing a trial lead in the epidural space, and if adequate coverage of the patient's area of pain can be achieved, allowing the patient to use the stimulator on an outpatient basis for 5–7 days. In 1993, Barolat et al. published a database of 106 patients in whom they had placed spinal cord stimulator leads (9). The electrodes were placed between the C1 and L1 spinal levels for chronic pain management, and the areas the patients felt stimulation were mapped. These maps provide a guideline as to which body areas will be stimulated by implanted electrodes. Barolat also noted that certain body areas were difficult to cover with paresthesia, including the low back, neck, and perineum. Clinical experience has shown that patients with bilateral extremity pain or pain in both the low back and legs may require bilateral lead placement to obtain adequate coverage. The placement of more than one lead in the epidural space allows not only wider paresthesia coverage, but also the use of complex stimulation programs that can be tailored to meet the patient's needs (see Fig. 6). With bilateral eight electrode leads, the possible stimulation combinations (anodes and cathodes) reach the thousands.

To insert the lead, the patient is placed in the prone position and sedated. The operative area is sterilely prepped and draped, and the skin is anesthetized with local anesthetic. When treating lower extremity pain, the puncture site is usually at the L1-2 level. The epidural space is entered with a Touhy needle, through which a lead is advanced in a cephalad direction. Using fluoroscopy, the lead is observed to move up the spinal canal until it reaches approximately the T9–T10 level. The lead can be manipulated to direct it slightly to the side corresponding to the patient's pain. For upper extremity pain, the skin is usually punctured at T1–T2, and the tip of the lead is placed at the C3–C4 level. The presence of scar tissue or other anatomic barriers can make lead placement difficult and occasionally impossible.

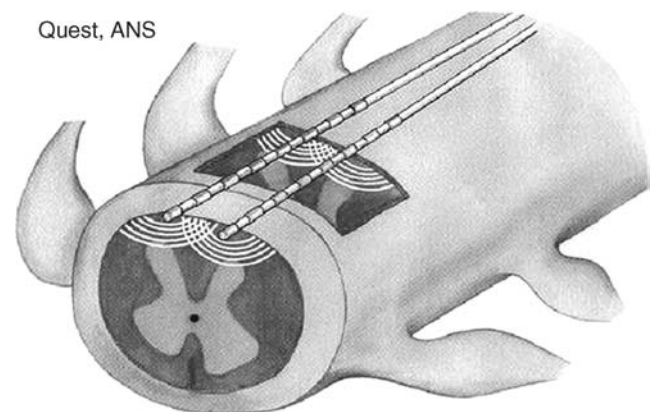


Figure 6. Dual leads allow wider parasthesia coverage. (Used courtesy of ANS, Inc.)

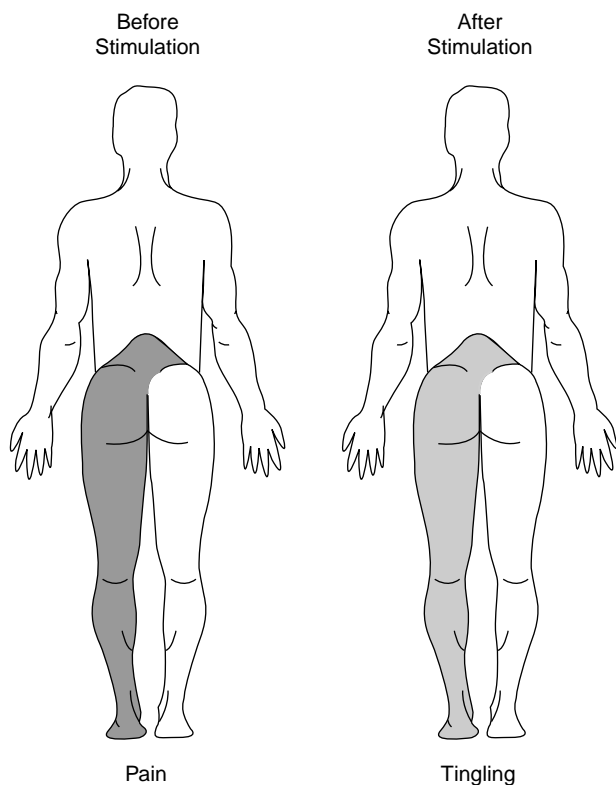


Figure 7. SCS covers painful area with pleasant paresthesias. (Used courtesy of Medtronic, Inc.)

Once the lead is felt to be in proper position, it is attached via a cable to the trial generator. This is an external programmer that allows various combinations of electrodes to be stimulated in an effort to cover the patient's pain with pleasant paresthesia, usually described as "tingling" (see Fig. 7). If the amplitude is set too high, the patient may experience discomfort or muscle stimulation. The patient must be awake enough at this point to answer questions and describe where they feel the stimulation. It is not unusual to need to adjust the position of the lead(s) several times before adequate coverage is obtained.

Once adequate coverage is obtained, the trial lead is secured with tape and/or suture. After recovery from anesthesia, the patient is given instructions as how to operate the stimulator. The patient is allowed to turn the device on or off, and can adjust the amplitude and rate to comfort. Reprogramming the pulse width and lead combinations is generally reserved for the pain specialist. The patient is told not to drive a car with the stimulator turned on, and excess twisting or raising the arms above the head is discouraged. As noted above, the patient will return to the clinic for removal of the trial lead in 5–7 days; however, the patient is encouraged to call sooner should anything change with the function of the device.

During the follow-up visit, several decisions are made. The patient is asked if the stimulator continued to cover their painful area, and if so, did it reduce the discomfort. Ideally, the patient obtained at least a 50% reduction in their pain during the trial. If the patient has received significant pain relief and they want to proceed with permanent implantation, the type of lead (surgical vs. percu-

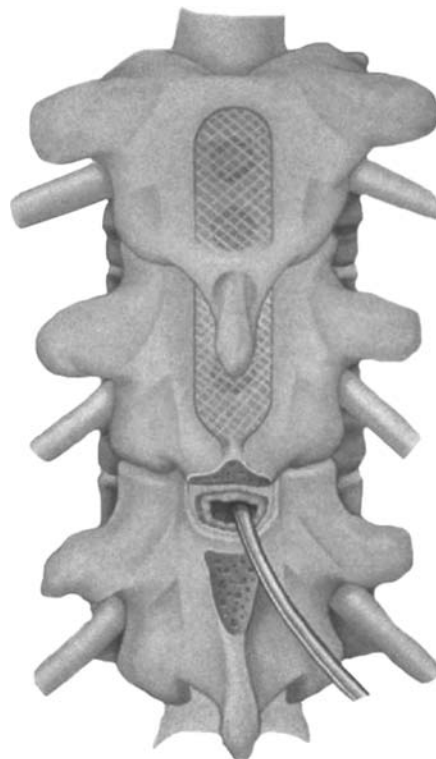


Figure 8. Paddle lead surgically inserted into the epidural space. (Used courtesy of Medtronic, Inc.)

taneous) is selected. Placement of the surgical lead requires a minilaminotomy, which includes removal of part of the inferior portion of the lamina and a portion of the *ligamentum flavum*, followed by insertion of the paddle lead into the exposed epidural space (see Fig. 8). Villavicencio et al. followed 27 patients who underwent placement of SCS leads (3). Patients who had electrodes placed via a laminectomy had significantly better long-term effectiveness than the patients with percutaneous leads. Nonetheless, permanent placement of percutaneous leads remains a viable and effective option that does not require a minilaminotomy. The placement of the permanent SCS lead requires that a generator (IPG) be inserted in a subcutaneous pocket, usually in the low abdomen or over the buttock. A cable is tunneled under the skin to the lead inserted in the spine. After permanent implantation, the spinal cord stimulator is controlled with a hand-held device.

OUTCOMES

A number of studies have looked at outcomes after SCS. Van Buyten et al. described 10 years of experience with SCS in 254 patients, 217 of whom had permanent stimulators placed (10). Before the study began, 10% of the patients had died and another 10% had undergone explantation. Reasons for explantation included ineffectivity (4.6%), infection, allergy, and recovery from pain. An independent review of the remaining patients who could be contacted and would participate in the study ($n = 123$) showed that 68% of them graded the treatment as excellent to good (excellent, very good, good, moderate, weak, no

improvement, worse). After excluding retirees and others not pursuing a career, 31% of the patients who had been working before their pain began had returned to work. The authors noted that their success rate is one of the highest reported.

Kay et al. published another retrospective study of SCS covering 13 years (11). Of 70 patients treated with SCS, there were 72 surgical revisions, including electrode (32), connecting cable (6), or generator revision (22). Battery depletion was the single most common indication for generator revision (16/22). Of the 72 revisions, 12 were for explantation. Of 48 patients who responded to a questionnaire, 60% rated their pain relief as substantial (>50%).

Bhadrakant et al. prospectively studied 29 patients over a 2-year period (12). The primary indication for SCS was failed back syndrome. Four of the 29 failed to obtain relief during the SCS trial. Of the 25 patients with permanent implants, SCS was beneficial in 50%. This result is similar to North et al. results for a large series ($n = 320$), where 52% of patients reported at least 50% pain relief (13).

Although these and other studies support the use of SCS for certain chronic pain syndromes, methodological problems preclude drawing final conclusions. Its retrospective nature, lack of controls, and heterogeneous patient populations flaw much of the research on SCS. More prospective studies, perhaps looking at SCS for individual pain syndromes, will be needed before this expensive technology becomes widely accepted.

FUTURE USES OF SCS

Greater understanding of both peripheral and central pain mechanisms combined with evolving technology have expanded the potential uses of SCS. Multiple-electrode configurations have allowed coverage of diffuse pain generators and made reprogramming simple when coverage is lost or new pain symptoms arise. Lumbosacral placement of SCS leads may allow treatment of refractory pelvic neuropathic conditions including sacral neuralgia, vulvodynia, or coccydynia. Urinary incontinence may also be treatable with this technique (14). Peripheral placement of SCS leads has been used for a number of conditions, including occipital neuralgia (i.e., spinally transformed migraine) and trigeminal neuralgia (15,16).

Other current and evolving uses for SCS include chronic regional pain syndromes (RSD and causalgia), postherpetic neuralgia, and postamputation pain. Patients with peripheral vascular disease suffering from rest and night pain seem to benefit from SCS; this indication is used more commonly in Europe than the United States. Spinal cord stimulation has also been shown to be effective in refractory angina (17). Other conditions treated with SCS include severe Raynauds phenomena, Buerger's disease, and diabetic neuropathy.

BIBLIOGRAPHY

1. Melzack R, Wall P. Pain mechanism: A new theory. *Science* 1965;150:951-979.

2. Shealy C, Mortimer J, Reswick J. Electrical inhibition of pain by stimulation of the dorsal columns: Preliminary clinical report. *Anesth Analg* 1967;46:489-491.
3. Villavicencio AT, Leveque J, Rubin L, Vulsara K, Gorecki JP. Laminectomy versus percutaneous electrode placement for spinal cord stimulation. *Neurosurgery* 2000;46:399-406.
4. Barolat G, Ketcic B, He J. Long term outcome of spinal cord stimulation for chronic pain management. *Neuromodulation* 1998;1:19-29.
5. Oakley J, Prager J. Spinal cord stimulation, mechanisms of action. *Spine* 2002;27:2574-2583.
6. Kumar K, Toth C, Nath RK, Verma AK, Burgess JJ. Improvement of limb circulation in peripheral vascular disease using epidural spinal cord stimulation: a prospective study. *J Neurosurg* 1997;86:662-669.
7. Meyerson BA, Linderhoth B. Mechanisms of spinal cord stimulation in neuropathic pain. *Neurolog Res* 2000; 22:285-292.
8. Alo KM, Holsheimer J. New trends in neuromodulation for the management of neuropathic pain. *Neurosurgery* 2002; 50:690-704.
9. Barolat G, Massaro F, He J, Zeme S, Ketcik B. Mapping of sensory responses to epidural stimulation of the intraspinal neural structures in man. *J Neurosurg* 1993;78:233-239.
10. Van Buyten J, Zundert JV, Vueghs P, Vanduffel L. Efficacy of spinal cord stimulation: 10 years of experience in a pain centre in Belgium. *Eur J Pain* 2001;5:299-307.
11. Kay AD, McIntyre MD, Macrae WA, Varma TRK. Spinal cord stimulation—a long-term evaluation of patients with chronic pain. *Bri J Neurosurg* 2001;15(4):335-341.
12. Bhadrakant K, Rosenfeld JV, Hutchinson A. The efficacy of spinal cord stimulation for chronic pain. *J Clin Neuro Sci* 2000;7(5):409-413.
13. North RB, Kidd DH, Zahurak M, James CS, Long DM. Spinal cord stimulation for chronic, intractable pain: Experience over two decades. *Neurosurgery* 1993;32(3):384-394.
14. Alo KM, Gohel R, Corey CL. Sacral nerve root stimulation for the treatment of urge incontinence and detrusor dysfunction utilizing a cephalocaudal intraspinal method of lead insertion: A case report. *Neuromodulation* 2001;4(2): 53-58.
15. Lou L. Uncommon areas of electrical stimulation for the relief of pain. *Curr Rev Pain* 2000;4:407-412.
16. Weiner L, Reed KL. Peripheral neurostimulation for control of intractable occipital neuralgia. *Neuromodulation* 1999; 2(3):217-221.
17. DeJongste MJL. Spinal cord stimulation for ischemic heart disease. *Neurolog Res* 2000;22:293-298.

See also BIOELECTRODES; BLADDER DYSFUNCTION, NEUROSTIMULATION OF; ELECTRONEUROGRAPHY; FUNCTIONAL ELECTRICAL STIMULATION; PERIPHERAL VASCULAR NONINVASIVE MEASUREMENTS; TRANSCUTANEOUS ELECTRICAL NERVE STIMULATION (TENS).

SPINAL IMPLANTS

MICHELE MARCOLONGO
Drexel University
Philadelphia, Pennsylvania
ABHIJEET JOSHI
Abbott Spine
Austin, Texas

INTRODUCTION

Spinal implants constitute the fastest growing segment of the orthopedic medical device industry. The area has until

the last 5 to 10 years been vastly under-studied for the proportion of patients who are afflicted with diseases and injuries to the spine. Consequently, new spine medical devices and medical device companies are emerging every day with new and better treatment strategies for prevalent spine disorders. This article will explore the physiological conditions and disease states that require treatment and then demonstrate some treatment strategies that are being used today. The devices in this text are not comprehensive (we would need a much larger space to do that) but do allow an understanding of the state-of-the-art in medical treatment of spinal disorders.

Human Spine

The human spine is a mechanical structure as it performs three fundamental biomechanical functions simultaneously (1). First, it transfers the weights (and resultant bending moments) of the head, trunk, and any weights being lifted to the pelvis. Second, it allows the sufficient physiological motion among the head, trunk, and pelvis. Third, and most important, it protects the delicate spinal cord from the potential damaging forces (and moments) resulting from the physiological motions and trauma (1).

Figure 1 show a schematic of the human spine, which is divided into three main regions: the upper region with 7 vertebrae (cervical spine), the middle region with 12 vertebrae (thoracic spine), and the lowermost with 5 vertebrae (lumbar spine). At the distal end of the spine, there is a basin-shaped structure, the pelvis, that supports the spinal column and is made of sacrum and coccyx with fused vertebrae. The human spine is not a straight structure, but it has specific curvature. The spine in the cervical and in the lumbar region is slightly convex anteriorly, whereas in the thoracic and sacral region, it is slightly convex posteriorly. The specific shape allows the increased flexibility while maintaining the overall spinal stability. It also facilitates increased shock-absorbing capacity along with adequate stiffness (1).

Each vertebra is made up of several parts. Figure 2 shows schematic of the vertebrae in a vertebral column. The body of the vertebra is the primary weight-bearing area. Between the vertebrae lie the intervertebral disks, which separate the adjacent vertebrae and act as cushions between them while allowing the movement of one vertebra relative to another. There is a large hole in the center part (spinal canal) that is covered by the lamina. The spinal cord runs through this spinal canal. There is a protruded bone in the central posterior region, called the spinous process. There are pairs of transverse processes that are orthogonal to the spinous process and provide attachment sites for the back muscles. Four facet joints are also associated with each vertebra. Four facet joints in two pairs (superior and inferior) interlock with adjacent vertebrae and provide the stability to the spine (1). An intervertebral disk is situated in between adjacent vertebrae. The disks are labeled with respect to the vertebrae levels, between which they are located. Thus, the T12/L1 disk is located between the twelfth thoracic and first lumbar vertebrae, whereas the L3/L4 disk is located between the third and fourth lumbar vertebrae.

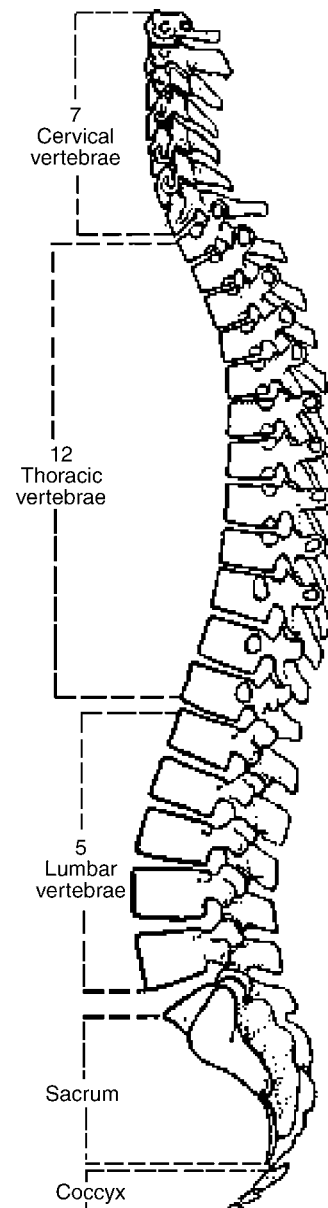


Figure 1. Schematic of human spine (2).

The intervertebral disk is basically a composite structure made up of three different tissues; the central core is called the nucleus pulposus (Fig. 3), which is attached radially to the multilayered fibers of the annulus fibrosus and attached superiorly and inferiorly to cartilaginous end plates (1). The nucleus is predominantly water in a matrix of proteoglycan, collagen, and other matrix proteins. The water content of the nucleus is very high at birth (approximating 90%) and then decreases through the aging cycle down to 70% or less. The annulus surrounds the nucleus with successive layers of tissue with collagen fibers oriented in alternating directions. The annulus is under tension when the nucleus absorbs water and swells. The cartilaginous end plates have multiple perforations that allow exchange of water and nutrients into the disk (4–6).

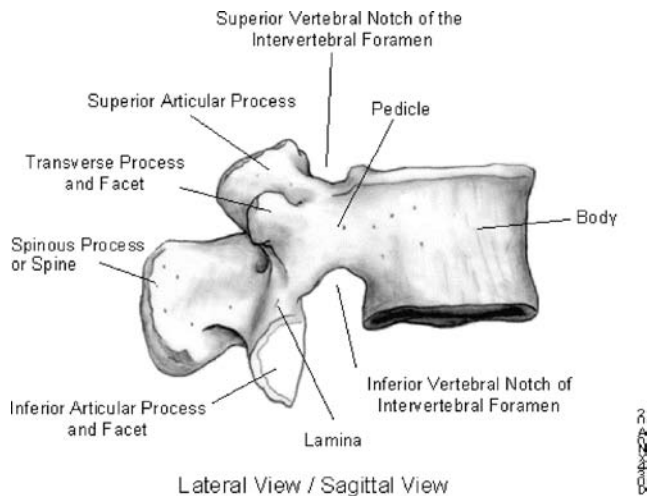


Figure 2. Spinal vertebrae (3).

SPINAL DISORDERS

Various spinal disorders are observed in humans; some manifest in pediatric patients, whereas others affect middle-aged and older patients. The most common spinal disorders can be generally described by three different categories: developmental bone deformities (scoliosis and kyphosis), bone degeneration (vertebral compression fractures), and degenerative disk disease (herniation, rupture and spinal cord stenosis).

Developmental Bone Deformities

The spine is a complex biomechanical structure and performs complex functions. This puts spine under greater strains, and bone deformities may develop during the course of time while performing the demanding functions such as supporting the cranium and trunk, or absorb stresses generated during daily physiological activities. Most common spinal disorders under this category include scoliosis and kyphosis.

Scoliosis. Scoliosis is a lateral curvature of the spine. The symptoms of the scoliosis include uneven waist,

different height shoulders, raised/prominent hip, and leaning of body to one side (7). Some causes of scoliosis include congenital deformity, cerebral palsy, atrophy, and neuromuscular problems.

Scoliosis can either be structural or functional. Structural scoliosis is referred in case of adjacent vertebrae rotation upon each other. This is generally followed by deformation of rib cage. In case of functional scoliosis, there is no fixed vertebral rotation or fixed deformity in the thoracic region. The rate of curve progression is not constant; however, the lumbar curves progresses more rapidly than thoracic curves. The scoliosis is generally classified as adolescent idiopathic scoliosis (AIS), adult scoliosis (with or without degenerative changes), and *de novo* scoliosis (which develops secondary to degenerative changes of the lumbar spine, especially in older age) (7).

The most common tools used for diagnosis of scoliosis include plane radiograph, computed tomography (CT) scans, and magnetic resonance imaging (MRI). Treatment options depend on the various factors, including the age, curvature angle, progress rate, location, flexibility, and spinal maturity. Conservative management (no treatment) is commonly incorporated when the curvature is mild (less than 20°).

Orthopedic braces are recommended in case of curvature angle of 25–40°, to prevent further spinal deformity, especially in children. The bracing, however, merely prevents the worsening of the existing curvature and does not restore normal alignment (8). Many types of braces are commercially available in the market. The brace, depending on the type and application, may extend from neck to pelvis (Milwaukee Brace) with plastic pelvic girdle, neck ring, and pressure pads (Fig. 4) (9), or it may just cover below the breast to the initial pelvic region only (Boston Brace). The use of braces has been generally effective in case of children, to prevent the further worsening of the scoliosis, but there is still a lack of consensus about the indications for the brace, type, and wearing time over the body.

Surgical options are used only in case of severe scoliosis (curvature angle greater than 45°) or for the curves that do not respond to nonsurgical treatments. The goals of the

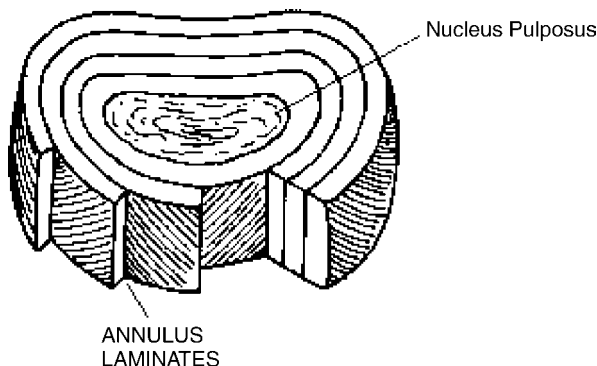


Figure 3. Schematic of an intervertebral disk (1).

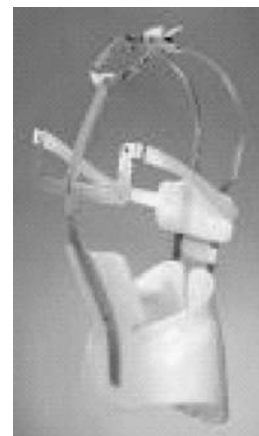


Figure 4. Milwaukee Brace (9).

surgical treatment are to prevent the progression of the curve and correct the deformity using instrumentation (7). The most common method to treat severe scoliosis is spinal fusion (anterior or posterior) and bone grafting/substitute. Bone graft can either be autologous iliac crest, from rib or allograft. In general, the facetectomy is followed by placement of bone graft and fixation. Various instrumentation options are available to surgeons for fixation. Basically, fixation is achieved by use of single/dual rods, posteriorly or anteriorly, along with screws and wires to the fixation point(s). The structures of the vertebral body, such as pedicles, sublaminar region, facets, and processes, may serve as fixation points for fusion (7). Anterior structural support is generally provided by mesh cages or ring allografts. Single thoracic curves are generally treated posteriorly using posterior instrumentation (hooks) and fusion.

Kyphosis. When viewed from the side, the normal spinal column is not completely straight. There are several gentle curves due to the shape and alignment of the vertebrae. Kyphosis is an exaggerated curvature of the spine or a rounded, "hunched" back. Most causes (metabolic, neuromuscular conditions, osteogenesis, spina bifida, among others) of the kyphosis are due to shortening of the anterior column, a weakening or lengthening of the posterior column, or both (7). The symptoms of kyphosis include difference in shoulder heights, forward bend of head compared with the rest of body, and tight hamstring (back thigh) muscles.

As in scoliosis, plane radiographs are also useful in diagnosing kyphosis. These help in defining the nature of sagittal deformities. Cobb (angle) measurements on these radiographs are performed to quantify the deformity in the sagittal and coronal plane. The angle is measured using the adjacent vertebral endplates (plane) as the basis of calculation. CT scans and MRI also find a place as useful diagnostic tools for better assessment of the spinal deformity.

The use of braces is recommended when the curve angle is between 40° and 60° on X ray. Surgical treatment is recommended when the curvature deformity is progressive, and the deformity may lead to neurological compromise. Again, spinal fusion (anterior or posterior) is referred for cases of severe deformity. In the case of young patients, posterior fusion might be considered, which would allow continuous anterior growth to partially correct the deformity (anterior release with posterior instrumentation).

Bone Degeneration

Compression fractures are generated in vertebrae when the bone tissue becomes weak due to degenerative changes. In most cases, the cause of the compression fracture is reduction in bone mineral density leading to weakening of bone (osteoporosis) (1,7). Osteoporosis causes both inorganic and organic phase bone loss. Loss of bone crystal weakens the bone to compressive loading, whereas loss of the organic matrix of bone makes the tissue more brittle, making the bony construct more susceptible to fracture. Other manifestations of osteoporosis include hyperkyphosis with chronic spinal pain and osteoporotic burst

fractures. However, the most common manifestation of the bone loss is a vertebral compression fracture (VCF).

To diagnose vertebral compression fractures, plane radiographs are used. To follow the progression of bone density loss throughout the osteoporotic disease process, dual photon absorptiometry (DPA) (which measures axial skeletal bone mass density) and dual energy X-ray absorptiometry (DEXA) (which measures baseline bone density with precision) are used. Quantitative CT can also be used in diagnosis of compression fractures.

Surgical treatment available for reduction of compression fractures is vertebral body augmentation: either kyphoplasty or vertebroplasty (7). Vertebroplasty is a procedure performed to relieve the pain and strengthen the weak vertebrae. During the procedure, an image-guided (X-ray) bone needle may be passed through the patient's back to have precise control over its location. Bone cement (polymethylmethacrylate or PMMA) is pushed through the needle to stabilize the fractured location of the vertebra. After curing, the PMMA biomaterial serves to stabilize the vertebra and to minimize the pain associated with the fracture.

Kyphoplasty is another method of vertebral augmentation, which uses bipedicular approach and balloon tamps to create voids in the bone. The instrumentation (cannula) used in kyphoplasty differs that from used in vertebroplasty. The void created by balloon is filled with PMMA. Other materials, such as calcium phosphate, hydroxyapatite, polymeric hydrogels, and combinations thereof, are being investigated as an alternative solutions to PMMA.

Degenerative Disk Disease

Lower back pain is one of the most prevalent socioeconomic diseases and one of the most important health-care issues today. Over five million Americans suffer from lower back pain, making it the leading cause of lost work days next only to upper respiratory tract illness (10–14). On an average, 50–90% of the adult population suffers from lower back pain (15), and lifetime prevalence of lower back pain is 65–80% (16). It is estimated that 28% experience disabling lower back pain sometime during their lives, 14% experience episodes lasting at least 2 weeks, whereas 8% of the entire working population will be disabled in any given year (16). The total cost of the lower back disabilities is in the range of \$50 billion per year in the United States (17) and £12 billion per year in the United Kingdom alone (18). The causes of lower back pain often remain unclear and may vary from patient to patient. It is estimated that 75% of such cases are associated with lumbar degenerative disk disease (DDD).

Many conservative treatment options exist for lower back pain. These generally aim at reducing the pain arising out of nerve root impingement and inflammatory response because of the migrated nucleus. The most commonly used surgical treatments include discectomy and spinal fusion and are sought when conservative treatments fail.

Progression of Degenerative Disk Disease. As the human life progresses, significant changes occur in the tissues of the intervertebral disk. DDD can be simply defined as the

loss of normal disk architecture accompanied by progressive fibrosis. At birth, the water content of the annulus fibrosus is about 80% and that of nucleus pulposus is about 90%. Through the degenerative process, this water content decreases to as low as 70% for the nucleus (19). Microscopic changes such as fragmentation of fibers, mucinous degeneration of fibers leading to cyst formation, and focal aggregation of the collagen to form round aggregates of amorphous material are observed in early stage of degeneration (20). The salient features of the DDD can be denoted as the loss of gelatinous nucleus pulposus, gradual disappearance of the originally well-defined border between the nucleus and the annulus, coarsening of the annulus lamellae, progressive fibrosis, and later fissuring of the annulus fibrosus with the deposition of the aging pigment (21–24).

The load transfer mechanism is clearly altered in the case of a dry nucleus. As a result, the end plates are subjected to reduced pressure at center and more pressure around the periphery. The stress distribution in the annulus is also altered significantly. Essentially, the nucleus does not perform its function of load transfer and the load transfer occurs through end plate—annulus—end plate route (1). The annulus is subjected to abnormal stresses and is more prone to injuries, and cracks/fissures first develop into the annulus.

With continued degeneration, the central nucleus may migrate through the crack developed in the annulus toward the periphery. The migration of the nucleus material is referred to as “disk herniation” (17). Approximately 90% of the disk herniation would occur at the L4-L5 and L5-S1 levels. The migrated material may impinge on the nerve root. The contact of the migrated nucleus with the nerve root irradiates debilitating back pain. Also, the herniated material elicits an inflammatory response because of the avascular nature of the nucleus. It is difficult to distinguish between the effects of aging from that of degeneration on the biomechanical behavior of the lumbar disk. The biomechanical behavior of the disk is dependent on its state of degeneration, which in turn depends on the age.

In case of the normal disk, any load acting on the disk is transferred to the annulus by means of swelling pressure (intradiscal pressure) generated by the nucleus (1). The water binding capability of the nucleus is a function of chemical composition of the nucleus. However, with aging and/or degeneration, changes occur to the proteoglycans as proteases and MMPs attack the molecules. The result is a decrease in the proteoglycan/collagen ratio, which leads to the lower water binding capability of the nucleus (25,26).

The load transfer mechanism in case of such a dehydrated disk is significantly altered (Fig. 5). The nucleus cannot generate sufficient intradiscal pressure to maintain disk height and normal mechanical function (25,27,28). Although it is not well understood, the consequence of the structural and mechanical changes to the disk may be a cause of lower back pain.

Stenosis. The reduction in the disk volume leads to instability, resulting in the growth of bone, end plates, and ligaments to compensate for this volume loss (stenosis). Stenosis is narrowing of the spinal canal (29). It occurs as a

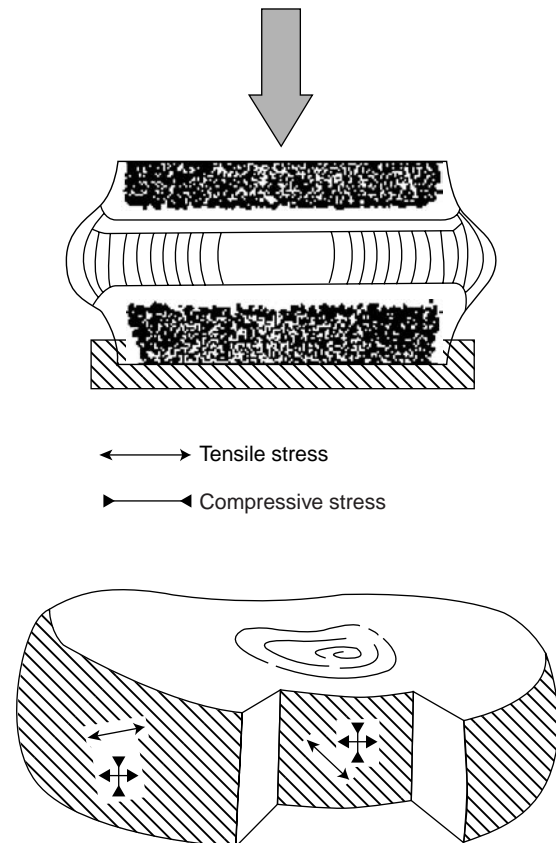


Figure 5. Degenerated disk (1).

result of aging and/or degenerative disk disease. The water content of the nucleus decreases, causing an abnormal load transfer mechanism within the disk. The disk height is reduced, and this dry/hardened disk may bulge into the spinal canal space. Additionally, the facet joints may become thick, thus narrowing the spinal canal further. Spinal stenosis in the lumbar spine may result in cauda equina syndrome and loss of bowel and bladder function. In general, the symptoms are not observed with stenosis. However, when present, the symptoms may include low back stiffness, leg weakness, numbness in the back/legs, and cramping.

Most common methods to diagnose and analyze the stenosis are plane X ray, MRI, and CT scan (7). These treatments, alone or in combination, provide valuable information about the patient's spine structure, location, and the extent of the disease. In particular, the following information can be revealed:

- Disk space narrowing
- Endplate osteophytes and sclerosis
- Facet enlargement and osteophytes formation
- Loss of lumbar lordosis

If conservative treatments such as medication, physical therapy, and spinal injections fail, a surgical approach may be recommended in the cases with persistent back pain

and/or progressive leg weakness. The indications for surgical treatment include radicular pain or neurogenic claudication with MRI or CT. In general, the goals of surgery are pain relief, increased mobility, and improvement in the patient's quality of life. Most common surgical treatments are laminectomy (in case of simple stenosis) and spinal fusion. Fusion is recommended when there is a stenosis in conjunction with

- Degenerative scoliosis or kyphosis
- Degenerative spondylolisthesis

The goal of the laminectomy or, lumbar decompression surgery, is to widen the spinal canal (30) to allow more space for spinal nerves. The treatment would ideally relieve the leg pain and, to a certain extent, back pain. When there is a vertebrae slippage relative to each other (spondylolisthesis), an abnormal motion would occur, which might require spinal fusion along with decompression.

Spondylolisthesis. Spondylolisthesis is defined as displacement or slippage of one vertebra on another (7). Osteoarthritis of the facet joints (degenerative arthritis that breaks the cartilage between the facet joints) can lead to instability of the vertebral segments. The L4-L5 motion segment has most flexion-extension movement and is more prone to such slippage, as a result of weakened facet joints.

The most common symptoms are pain irradiating in lower extremities and cauda equina compression along with incontinence of bowel or bladder. Like most other spinal disorders, surgical treatment is recommended only when the nonsurgical treatments such as activity modification and physical therapy show no significant improvement in the patient's quality of life. The goals of the surgical treatment are pain reduction, prevention of further slip, and stabilization of spine (7). Surgical treatments include spinal fusion (with or without decompression), slip reduction or instrumentation, and interbody fusion.

Recommended operative treatments of degenerative spondylolisthesis are decompressive laminectomy (removal of lamina and medial joints), decompression with postero-lateral fusion (complete laminectomy and partial facetectomy along with fusion of the transverse process), and decompression with instrumental fusion.

TREATMENT OPTIONS AND MEDICAL DEVICES

Most spinal medical devices involve permanently fusing vertebral bone to correct a deformity or to limit a motion segment to stabilize the joint segment and relieve pain. Spine medical devices have their origin in plates, screws, and rods made from stainless steel and titanium, and today, these components comprise the majority of the implantable devices today.

Implants for Developmental Spine Deformities

Implants for developmental bone deformities such as scoliosis and kyphosis, generally use metal rods, screws,

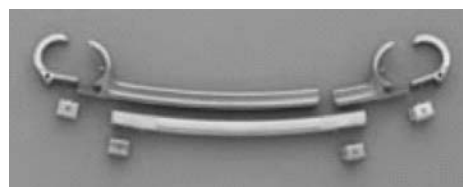


Figure 6. VEPTR for the treatment of pediatric scoliosis from Synthes Spine (41).

plates, and rib cages. For example, CD Horizon from Medtronic (Fig. 10) can be used for treatment of such deformities. Recently, the Food and Drug Administration (FDA) approved the Vertical Expandable Prosthetic Titanium Rib (VEPTR) from Synthes Spine (41,43), which is a surgically implanted device used to treat thoracic insufficiency syndrome (TIS) in pediatric patients. TIS is a congenital condition where severe deformities of the chest, spine, and ribs prevent normal breathing and lung growth and development.

The VEPTR device (Fig. 6) is a curved metal rod that is attached to ribs near the spine using hooks located at both ends of the device. The VEPTR device helps straighten the spine and separate ribs so that the lungs can grow and fill with enough air to breathe. The length of the device can be adjusted as the patient grows. It is hoped that the device will accomplish more normal growth pattern without spinal growth limitations, decreased chest, spine, and rib deformity and increased lung volume (43).

Implants for Degenerative Bone Disease

In case of kyphosis caused by osteoporosis (decrease in bone mass density with increased bone brittleness), minimally invasive methods such as vertebroplasty or (balloon kyphoplasty) are being used (7). Vertebroplasty involves the percutaneous injection of (PMMA) into a fractured vertebral body. Balloon kyphoplasty is another surgical approach to treat the kyphosis or deformity of the spine. In this procedure, an inflatable balloon is inserted between the vertebrae space to increase the disk height. Increase in disk height helps to reduce the deformity. The extra space created by balloon is filled with bone cement (PMMA), which when cured, binds the fracture. The hardened cement thus provides the strength to fractured/weak vertebrae and stability to the spine along with reducing the pain.

Implants for Degenerative Disk Disease

The treatment options for the patient would vary based on the age, pain history, and severity. When conservative treatments (such as rest, medications, physical therapy, etc.) fail, the patient is advised to undergo surgery. The goal of the surgery is to alleviate the pain. Most popular surgeries include lumbar microdiscectomy, lumbar laminectomy, microendoscopic surgery, and arthroscopic lumbar discectomy (30,31).

Lumbar microdiscectomy (or lumbar decompression) is a proven technique to reduce the back pain associated with herniated disks. In this treatment, a small portion of the

bone over the nerve root is removed to get relief from pain. A microscope is used to aid in visualizing the pinched nerve and subsequent microsurgical procedure to remove the excess portion of the herniated disk. Similarly, an open decompression (lumbar laminectomy) is another type of surgery that is performed to reduce the pain caused by neural impingement, which is particularly effective as a treatment for spinal stenosis. It is typically done with a posterior approach. The spine is approached by cutting the left and right back muscles off the lamina and removing the lamina itself. The facet joints are then trimmed to allow more space to nerve roots (30).

Even with these surgical treatments, pain may not be relieved for disks that are more severely degenerated. In these cases, fusion is required to restrict the motion of the segment and thus attempt to relieve pain. A discussion of fusion technologies and the associated implants follows. More recent advances in non-fusion technologies are aimed at preserving the motion segment while relieving pain. Such non-fusion technologies include total disk replacement, nucleus replacement, and annulus repair. Numerous new companies and new medical implant strategies are being explored currently and will hopefully prove to be clinically relevant pain relief and function restoring solutions to DDD.

Fusion Solutions. Spinal fusion is recommended when the discectomy approach may not be clinically relevant, and the goal is to relieve pain by stopping the motion of a spine segment. Spinal fusion instrumentation is essentially of three main types: pedicle screws, anterior interbody cages, and posterior lumbar cages. The bone generally fuses more effectively when its motion is minimized; hence, these devices are used to limit the motion of the fused segment. Similarly, the spinal fusion is based on the assumption that if the joint does not move, it will not create pain.

Pedicle screws (Fig. 7) are the means of providing anchor to the spine. They are used in combination with the short rod to grip the spine and are made from biocompatible metals such as medical-grade stainless steel or titanium. After a sufficient time, these screws can be removed by doing a surgery; however, most surgeons recommend keeping the screws unless it causes discomfort to patient.



Figure 7. Pedicle screws and instrumentation from Medtronic (32).



Figure 8. Jaguar Interbody Cage from Depuy Spine (33).

Anterior interbody cages (Fig. 8) have been recently approved by the FDA to use in the disk space. The cages are made from titanium and are porous, which allows the bone graft to grow. Cages are also made of novel composite materials (e.g., Jaguar from Depuy) such as carbon fiber-reinforced polymeric materials. The bone graft grows through the cage from one vertebra to another.

These cages are placed in front (anterior) of the lumbar spine and, hence, the name. The cages can be inserted using either mini-laparotomy or endoscopy, however, the former is preferred. In general, a 3 to 5 in. (7.6 to 12.7 cm) incision on the left side of the abdomen is made to approach the damaged disk.

Anterior lumbar interbody fusion (ALIF) surgery is often combined with posterior lumbar interbody fusion (PLIF) surgery to provide more rigid fixation. When the cages are placed from back of the spine, it is called posterior fusion. Coda (Fig. 9) is a titanium alloy device for PLIF with pedicle screws from Abbott Spine and features intraoperative adjustment for lordosis.

There is another form of the fusion surgery: transforaminal lumbar interbody fusion (TLIF), which is considered as an extended form of PLIF. In TLIF, an entire facet joint is removed to get a better access to disk space as compared with PLIF. This facilitates better visualization and more removal of the disk material and placement of larger bone graft/implant. The success rate of the cages almost entirely depends on the vertebrae condition. The surgery is not recommended in the case of osteoporosis because the vertebral body would not sustain the cage, leading to eventual failure of the end plates. In that sense, the pedicle screws are better than anterior cages as a



Figure 9. Coda PLIF device with pedicle screws from Abbott Spine (34).



Figure 10. CD Horizon Legacy from Medtronic (32).

fixation device. The anterior/posterior fusion is performed in the case of severe spinal instability or in revision surgery. The advantage of the anterior/posterior fusion is that it provides more surface area for the bone fusion to occur.

The gold standard in the case of fusion is considered to be postero-lateral gutter fusion surgery (30). A bone graft is placed in the postero-lateral portion of the spine. The transverse process of the vertebral body serves as an attachment site to the bone graft, which eventually grows to complete the fusion at the site.

The recent trend is to offer the spinal systems that can be used for multiple spinal treatments. For example, the CD Horizon Legacy Spinal System (Fig. 10) can be used as a posterior, noncervical, nonpedicle screw fixation system for treatment of DDD, spinal stenosis, spondylolisthesis, spinal deformities like scoliosis, and kyphosis. When used as a pedicle screw fixation system of the noncervical posterior spine, it may be indicated for degenerative spondylosthesis, kyphosis, and scoliosis.

Novel concepts for spine care (e.g., Dynesys from Zimmer Spine and Wallis from Abbott Spine) are recently introduced in the market. Dynesys (Fig. 11) is a posterior

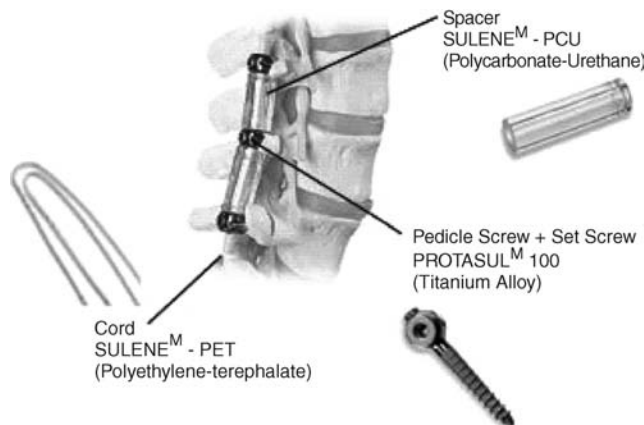


Figure 11. Dynesys from Zimmer Spine (35).



Figure 12. Wallis from Abbott Spine (34).

dynamics stabilization system, which is designed to bring the lumbar vertebrae back into more anatomical position while stabilizing the affected segments. The system used flexible materials threaded through pedicle screws, achieving dynamics stabilization.

The Wallis device (Fig. 12) from Abbott Spine is also another non-fusion spinal stabilization device that is under clinical trials in United States. The system is designed to treat the pain caused by initial stage DDD and aims to stabilize the lumbar spine without fusion, with a minimally invasive procedure.

The bone grafts used for the fusion can be either taken from patient's iliac crest (autograft) or from a human cadaver (allograft), such as Puros (Fig. 13) from Zimmer Spine (34). Autografts have the obvious advantage of compatibility with the patient's body. It helps in osteoconduction (bone growth) by means of providing calcium scaffold along with osteoblasts (bone growing cells) and morphogenic proteins (bone growing proteins). The allografts, in comparison, do not have osteoblasts and morphogenic proteins and merely provide the calcium scaffold for the fusion to occur. However, autografts lead to higher and longer postoperative pain as the bone graft is taken from the patient's own body.

Recently, synthetic bone grafts are introduced (e.g., Infuse), which represents an rhBMP-2 (recombinant human bone morphogenetic protein- 2) formulation combined with a bovine-derived absorbable collagen sponge (ACS) carrier. The INFUSE Bone Graft/LT-CAGE Lumbar Tapered Fusion Device, from Stryker Spine (36) (Fig. 14) is indicated for spinal fusion procedures in skeletally mature patients with DDD at one level from L4-S1, who may also have up to grade I spondylolisthesis at the involved level.

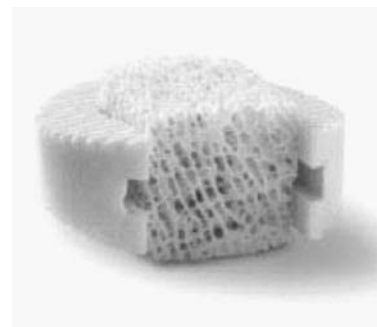


Figure 13. Puros allograft from Zimmer Spine (35).

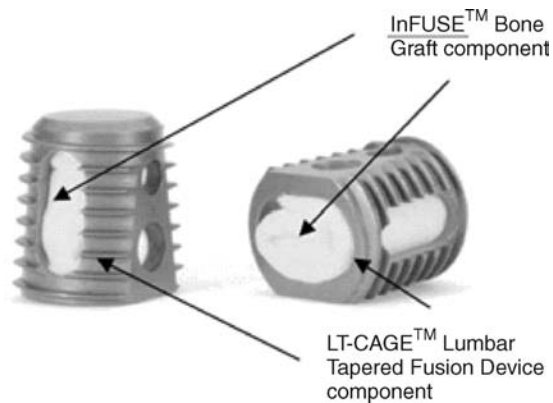


Figure 14. Infuse synthetic bone graft from Stryker Spine (36).

Bone stimulators offer another approach that potentially aid in spinal fusion. These externally applied devices emit low electrical current (30). This is aimed at facilitating stimulation of bone growth and increasing the chance of achieving spinal fusion. These are used in the case of patients who have a potentially very slow rate of obtaining solid fusion or in the case of the revision surgery.

The fusion and discectomy relieve pain but do not restore the normal spinal motion (37,38). The motivation behind exploration of the new and better solutions for the treatment of lower back pain is the failure of current treatments (conservative and surgical) in terms of restoring the motion and normal disk biomechanics. This is further aggravated by the complications that may occur after the surgical treatments, such as discectomy and/or spinal fusion.

Non-Fusion Solutions. Total disk replacement, where an entire diseased disk is removed and replaced by a synthetic implant, and nucleus pulposus arthroplasty, where only the nucleus of the disk is replaced either by a synthetic implant or recreated using tissue engineering approach, are the emerging approaches as alternatives to current surgical procedures for the treatment of the lower back pain (39). Annulus repair techniques, where defects in the annulus are either modified or repaired, also finds a place in emerging techniques and can be potentially used either alone or in combination with nucleus pulposus arthroplasty procedures, depending on the degenerative state of the intervertebral disk.

Total Disk Replacement. Total disk replacement targets later stages of disk degeneration (Galante grade IV), where the annulus is severely degenerated and is beyond repair (40). The diseased disk is entirely removed and replaced by a medical device that provides motion to the joint segment. Disk replacement may serve to eliminate the back pain and restore the physiological motion. A similar approach for total knee and hip replacement is highly successful. Total disk prostheses may be better options to spinal fusion and/or discectomy as it allows the physiological motion between the adjacent vertebrae. Another advantage would be that the effectiveness of the surgery will not be dependent on

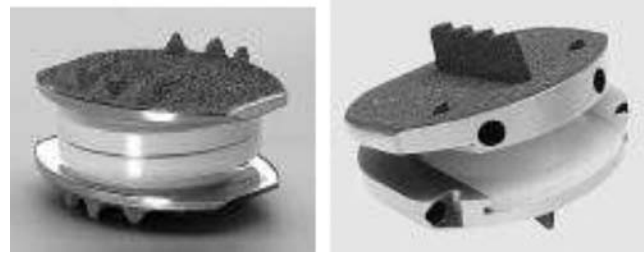


Figure 15. Charite from Depuy Spine (33) and ProDisc from Synthes Spine (41).

the integrity of the annulus or degeneration state. To simulate the natural structure and function of the spinal unit, total disk prostheses also provide adequate fixation to the vertebrae.

There are a variety of total disk replacement design strategies, but two of the concepts that are furthest along are the Charite and the ProDisc, which are each based on metallic end plates that are porous coated and allow fixation to the superior and inferior end plates as well as an ultra-high molecular weight polyethylene core, which provides a low friction articulation of the adjacent vertebrae. The use of artificial disks (Fig. 15) as a replacement to the damaged disk is currently in various phases of development and clinical trials. The Charite received FDA approval in 2004, and ProDisc, Maverick, and Flexicore are under clinical evaluation at the time of this writing (30).

Nucleus Replacement. The nucleus pulposus is a major component of the intervertebral disk and is actively involved in the disk function and load transfer mechanism. It is also involved with the pathologic changes of the disk. Researchers began to consider replacement of the nucleus alone because this tissue seems to degenerate before the annulus fibrosus. If this tissue alone can be replaced, preserving the annulus fibrosus, this may prolong the life of the disk and postpone or prevent the need for a more aggressive procedure such as fusion or total disk replacement. Nucleus replacement, as in case of total disk replacement, aims for restoration of the normal disk mechanics and functions, in contrast with the current surgical procedures of the discectomy and the spinal fusion.

There are several nucleus implants in the various phases of development and clinical trials. Some are already implanted in humans in Europe (e.g., RayMedica PDN, Disc Dynamic's DASCOR), whereas most other nucleus implants are undergoing bench-top testing and/or investigational device exemption (IDE) submissions. The Raymedica prosthetic nucleus device (Fig. 16) has the longest history of all nucleus implants on the market. The clinical results of the PDN have been promising for pain relief and disk height restoration (41), but they are troubled by expulsion of the device from the annulus. Alternative implant designs and surgical procedures have limited this complication, but it remains a major challenge for these types of devices.

To perform surgical intervention on the intervertebral disk (e.g., in the case of nucleus replacement), outer

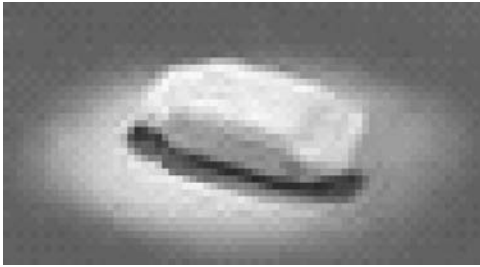


Figure 16. Single prosthetic nucleus device by Raymedica (42).

annulus fibrosus needs to be compromised. If the artificial incision in the annulus is not repaired, there is very high risk of nucleus implant expulsion, even under mild physiological loading. The main idea is to seal the annulus incision and/or prevent the expulsion of the nucleus implant from the created window. A barrier can be placed in between the nucleus and the annulus to prevent expulsion. These technologies are currently being explored in early clinical trials and in preclinical evaluations.

APPENDIX 1

Spinal Disorder	Treatment Options	Device/Implant
Stenosis	-Laminectomy -Spinal fusion	-Fusion Instrumentation <ul style="list-style-type: none"> • Cages • Pedicle screws • Metal rod • Autograft/Allograft
Spondylolisthesis	-Laminectomy -Spinal fusion	Fusion instrumentation <ul style="list-style-type: none"> • Cages • Pedicle screws • Metal rod • Autograft/Allograft
Scoliosis	-Bracing -Spinal fusion	-Braces -Fusion instrumentation <ul style="list-style-type: none"> • Cages • Pedicle screws • Metal rod • Autograft/Allograft
Kyphosis	-Balloon kyphoplasty	-Bone cement
Vertebral compression Fracture	-Vertebroplasty -Kyphoplasty	-Bone cement-Emerging materials
Disk degeneration and herniation	-Discectomy -Spinal fusion -Total disk prosthesis -Nucleus replacement	-Fusion instrumentation <ul style="list-style-type: none"> • Cages • Pedicle screws • Metal rod • Autograft/Allograft -Facet replacement -Nucleus implant device -Total disk arthroplasty

CONCLUSIONS

This is perhaps the most exciting time in the development of medical devices for spinal applications. Never before have so many academic researchers, clinicians, and corporations so aggressively pursued solutions to spinal conditions that have a potential to be solved with medical devices. Along with this tremendous interest is the interest in better understanding of the anatomical structure, biochemistry, and function of the spinal structures. As more information becomes available, further refinements in treatments through medical devices will be improved offering more tools to the surgeon and a better chance of relieving pain while preserving the function of the spine.

APPENDIX 2. TERMINOLOGY

Allograft The transplant of an organ or tissue from one person to another.

Autograft The transplant of an organ or tissue from one body site to another body site of the same person.

Compression Fracture Collapse of the bone of the vertebral body, mostly due to osteoporosis and trauma.

Disk Degeneration The loss of normal disk architecture accompanied by progressive fibrosis. This is seen as loss of hydration of the disk material and loss of disk height. This complex process alters the normal biomechanics of the spine and may cause back pain.

Disk Herniation Migration of the central nucleus pulposus of the disk toward disk periphery through cracks or fissures in outer annulus.

Discectomy A procedure in which an excess portion of the disk impinging on the nerve root is cut off.

Kyphosis An exaggerated curvature of the back bones (vertebrae) in the upper back area or a rounded, "hunched" back.

Kyphoplasty A procedure that combines the vertebroplasty technique with balloon catheter technology to treat the osteoporotic vertebral compression fractures.

Laminectomy A procedure in which the lamina (roof) of the vertebra is trimmed to create more space for the spinal nerves.

Nucleus Implant An artificial material, which can be used as a replacement to the degenerated nucleus of the intervertebral disk to relieve back pain and preserve the normal motion.

Osteoporosis A disease in which bones become fragile and brittle, making it prone to break easily.

Scoliosis A curvature of the spine.

Spinal Fusion A procedure in which an intervertebral disk between the adjacent vertebrae is replaced by bone graft. The procedure is performed to relieve the back pain and stabilize the spinal segment by fusing the vertebrae together, with or without spinal instrumentation.

Spondylolisthesis Slippage of one vertebra on another.

Stenosis Narrowing of the spinal canal.

Total Disk Prosthesis An artificial device, which can be used as a feasible replacement of the degenerated disk to relieve back pain and preserve the motion.

Vertebroplasty A procedure that stabilizes the collapsed vertebra with the injection of the medical-grade bone cement into the spine.

BIBLIOGRAPHY

- White AA, Panjabi MM. Clinical Biomechanics of the Spine. II ed. Philadelphia: J.B. Lippincott Company; 1990.
- <http://www.ab.ust.hk>.
- <http://www.spinalstenosis.org/>.
- Iatridis JC, et al. Is the nucleus pulposus a solid or a fluid? Mechanical behaviors of the nucleus pulposus of the human intervertebral disk. Spine 1996;21:1174–1184.
- Ayad S, Weiss JB. Biochemistry of the intervertebral disk. In: MIV J, editor. The Lumbar Spine and Back Pain. 3rd ed. New York: Churchill-Livingstone; 1987. 100–137.
- Buckwalter JA. Aging and degeneration of the human intervertebral disk. Spine 1995;20:1307–1314.
- Vaccaro A. Core Knowledge in Orthopaedics. Spine. St. Louis, MO: CV Mosby; 2004.
- <http://www.scoliosisrx.com>.
- <http://www.bostonbrace.com/superstructure.htm>.
- MedPro Month. 1998; V.VIII: Number 1.
- Andersson GBJ. Epidemiologic Aspects of low-back pain in industry. Spine 1981;6(1):53–60.
- Hedman TP, et al. Design of an intervertebral disk prosthesis. Spine 1991;16(6):S256–S260.
- Cats-Baril WL, Frymoyer JW. Identifying patients at risk of becoming disabled because of low-back pain—the Vermont Rehabilitation Engineering Center predictive model. Spine 1991;16(6):605–607.
- Sehgal N, Fortin JD. Internal disk disruption and low back pain. Pain Physician 2000;3(2):143–157.
- Heliövaara M, et al. Determinants of sciatica and low-back pain. Spine 1991;16(6):608–614.
- Manchikanti L. Epidemiology of low back pain. Pain Physician 2000;3(2):167–192.
- Bao QB, Yuan HA. Artificial disk technology. Neurosurgical Focus 2000;9(4):1–9.
- Bibby S, et al. The pathophysiology of the intervertebral disk. Joint Bone Spine 2001;68:537–542.
- Vernon-Roberts B. Age-related and degenerative pathology of intervertebral disks and apophyseal joints. In: Jayson MIV, editor. The Lumbar Spine and Back Pain. New York: Churchill Livingstone; 1992. 17–41.
- Vernon-Roberts B. Disk pathology and disease states. In: Ghosh P, editor. The Biology of the Intervertebral Disk. Boca Raton: CRC Press; 1988. 73–120.
- Coventry MB, Ghormley RK, Kernohan JW. The intervertebral disk: Its microscopic anatomy and pathology. Part III. Pathologic changes in the intervertebral disk. J Bone Joint Surg 1945;27A:460–474.
- Friberg S, Hirsch C. Anatomical and clinical studies on lumbar disk degeneration. Acta Orthop Scand 1949;19:222–242.
- Harris RI, Macnab I. Structural changes in the lumbar intervertebral disks. Their relationship to low back pain and sciatica. J Bone Joint Surg 1954;36B:304–322.
- Hoof VD. A: Histological age changes in the annulus fibrosus of the human intervertebral disk. Gerontology 1964;9:136–149.
- Bao QB, et al. The artificial disk: Theory, design and materials. Biomaterials 1996;17:1157–1166.
- Akeson WH, et al. Biomechanics and biochemistry of the intervertebral disk. Clin Orthop Rel Res 1977;129:133–139.
- McNally DS, Adams MA. Internal Intervertebral disk mechanics as revealed by stress profilometry. Spine 1992; 17:66–73.
- Osti OL, et al. Annular tears and disk degeneration in the lumbar spine. J Bone Joint Surg (Br) 1992;74B:678–682.

29. Snyder DL, Doggett D, Turkelson C. Treatment of degenerative lumbar spinal stenosis. *American Family Physician* 2004;70(3):517–520.
30. <http://www.spine-health.com>.
31. <http://www.texasspinecenter.com>.
32. <http://www.medtronic.com>.
33. <http://www.depuyspine.com>.
34. <http://www.abbottspine.com>.
35. <http://www.zimmerspine.com>.
36. <http://www.stryker.com/spine>.
37. Kambin P, Savitz MH. Arthroscopic microdiscectomy: An alternative to open disk surgery. *Mount Sinai J Med* 2000; 67(4):283–287.
38. Weber H. Lumbar disk herniation: A controlled prospective study with ten years of observation. *Spine* 1993;8:131–140.
39. Joshi A. Mechanical behavior of the human lumbar intervertebral disk with polymeric nucleus implant: An experimental and finite element study. Ph.D. Thesis. Drexel University, 2004.
40. Galante JO. Tensile properties of the human annulus fibrosus. *Acta Orthop Scand* 1967; (Suppl. 100):4–91.
41. <http://www.synthes.com>.
42. <http://www.raymedica.com>.
43. <http://www.fda.gov>.

See also HUMAN SPINE, BIOMECHANICS OF; SCOLIOSIS, BIOMECHANICS OF.

SPINE. See HUMAN SPINE, BIOMECHANICS OF.

SPIROMETRY. See PNEUMOTACHOMETERS.

STATISTICAL METHODS

ARNAUD DELORME
University of San Diego,
La Jolla, California

INTRODUCTION

Statistics can be called that body of analytical and computational methods by which characteristics of a population are inferred through observations made in a representative sample from that population. Since scientists rarely observe entire populations, sampling and statistical inference are essential. Although, the objective of statistical methods is to make the process of scientific research as efficient and as productive as possible, many scientists and engineers have inadequate training in experimental design and in the proper selection of statistical analyses for experimentally acquired data. Gill (1) states: "...statistical analysis too often has meant the manipulation of ambiguous data by means of dubious methods to solve a problem that has not been defined." The purpose of this article is to provide readers with definitions and examples of widely used concepts in statistics. This article first discusses some general principles for the planning of experiments and data visualization. Then, since it is expected that most readers are not studying this article

to learn statistics, but to find practical methods for analyzing data, a strong emphasis has been put on choice of an appropriate standard statistical model and statistical inference methods (parametric, nonparametric, resampling methods) for different types of data. Then, methods for processing multivariate data are briefly reviewed. The section following it deals with clinical trials. Finally, the last section discusses computer software and guides the reader through a collection of bibliographic references adapted to different levels of expertise and topics.

DATA SAMPLE AND EXPERIMENTAL DESIGN

Any experimental or observational investigation is motivated by a general problem that can be tackled by answering specific questions. Associated with the general problem will be a population. For example, the population can be all human beings. The problem may be to estimate the probability by age bracket for someone to develop lung cancer. Another population may be the full range of responses of a medical device to measure heart pressure and the problem may be to model the noise behavior of this apparatus.

Often, experiments aim at comparing two subpopulations and determining if there is a (significant) difference between them. For example, the frequency occurrence of lung cancer of smokers compared may be compared to nonsmokers or the signal/noise ratio generated by two brands of medical devices may be compared and determined which brand outperforms the other with respect to this measure.

How can representative samples be chosen from such populations? Guided by the list of specific questions, samples will be drawn from specified subpopulations. For example, the study plan might specify that 1000 presently cancer-free persons will be drawn from the greater Los Angeles area. These 1000 persons would be composed of random samples of specified sizes of smokers and nonsmokers of varying ages and occupations. Thus, the description of the sampling plan will imply to some extent the nature of the target subpopulation, in this case smoking individuals.

Choosing a random sample may not be easy and there are two types of errors associated with choosing representative samples: sampling errors and nonsampling errors. Sampling errors are those errors due to chance variations resulting from sampling a population. For example, in a population of 100,000 individuals, suppose that 100 have a certain genetic trait and in a (random) sample of 10,000, 8 have the trait. The experimenter will estimate that 8/10,000 of the population or 80/100,000 individuals have the trait, and in doing so will have underestimated the actual percentage. Imagine conducting this experiment (i.e., drawing a random sample of 10,000 and examining for the trait) repeatedly. The observed number of sampled individuals having the trait will fluctuate. This phenomenon is called the sampling error. Indeed, if sampling is truly random, the observed number having the trait in each repetition will fluctuate randomly ~10. Furthermore, the limits within which most fluctuations will occur are estimable using standard statistical methods.

Consequently, the experimenter not only acknowledges the presence of sampling errors, but he can estimate their effect.

In contrast, variation associated with improper sampling is called nonsampling error. For example, the entire target population may not be accessible to the experimenter for the purpose of choosing a sample. The results of the analysis will be biased if the accessible and nonaccessible portions of the population are different with respect to the characteristic(s) being investigated. Increasing sample size within the accessible portion will not solve the problem. The sample, although random within the accessible portion, will not be representative of the target population. The experimenter is often not aware of the presence of nonsampling errors (e.g., in the above context, the experimenter may not be aware that the trait occurs with higher frequency in a particular ethnic group that is less accessible to sampling than other groups within the population). Furthermore, even when a source of nonsampling error is identified, there may not be a practical way of assessing its effect. The only recourse when a source of nonsampling error is identified is to document its nature as thoroughly as possible. Clinical trials involving survival studies are often associated with specific nonsampling errors (see the section dealing with clinical trials below).

DESCRIPTIVE STATISTICS

Descriptive statistics are tabular, graphical, and numerical methods by which essential features of a sample can be described. Although these same methods can be used to describe entire populations, they are more often applied to samples in order to capture population characteristics by inference.

The two main types of data samples will be differentiated: qualitative data samples and quantitative data samples. Qualitative data arises when the characteristic being observed is not measurable. A typical case is the “success” or “failure” of a particular test. For example, to test the effect of a drug in a clinical trial setting, the experimenter may define two possible outcomes for each patient: either the drug was effective in treating the patient, or the drug was not effective. In the case of two possible outcomes, any sample of size n can be represented as a sequence of n nominal outcome x_1, x_2, \dots, x_n that can assume either the value success or failure.

By contrast, quantitative data arise when the characteristics being observed can be described by numbers. Discrete quantitative data is countable, whereas continuous data may assume any value, apart from any precision constraint imposed by the measuring instrument. Discrete quantitative data may be obtained by counting the number of each possible outcome from a qualitative data sample. Examples of discrete data may be the number of subjects sensitive to the effect of a drug (number of success and number of failure). Examples of continuous data are weight, height, pressure, and survival time. Thus, any quantitative data sample of size n may be represented as a sequence of n numbers x_1, x_2, \dots, x_n and sample statistics are functions of these numbers.

Table 1. Result of a Hearing Aid Device Satisfaction Survey in 1000 Patients Showing the Frequency Distribution of Each Response

Satisfaction rank	Number of Responses
0	38
1	144
2	342
3	287
4	164
5	25
Total	1000

Discrete data may be preprocessed using frequency tables and represented using histograms. This is best illustrated by an example. For discrete data, consider a survey in which 1000 patients fill in a questionnaire for assessing the quality of a hearing aid device. Each patient has to rank product satisfaction from 0 to 5, each rank being associated with a detailed description of hearing quality. Table 1 represents the frequency of each response type. A graphical equivalent is the frequency histogram illustrated in Fig. 1. In the histogram, the heights of the bars are the frequencies of each response type. The histogram is a powerful visual aid to obtain a general picture of the data distribution. In Fig. 1, notice a majority of answers corresponding to response type 2 and a 10-fold frequency drop for response types 0 and 5 compared to response type 2.

For continuous data, consider the data sample in Table 2, which represents amounts of infant serum calcium in $\text{mg}\cdot 100\text{ mL}^{-1}$ for a random sample of 75 week old infants whose mothers received vitamin D supplements during pregnancy. Little information is conveyed by the list of numbers. To depict the central tendency and variability of the data, Table 3 groups the data into six classes, each of width $0.03\text{ mg}\cdot 100\text{ mL}^{-1}$. The “frequency” column in Table 3 gives the number of sample values occurring in each class. The picture given by the frequency distribution in Table 3 is a clearer representation of central

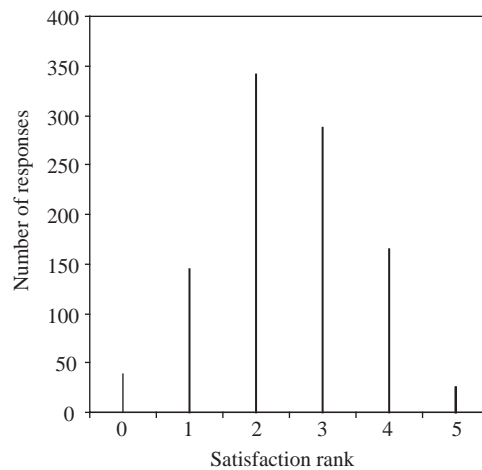


Figure 1. Frequency histogram for the hearing aid device satisfaction survey of Table 1.

Table 2. Serum Calcium (mg·100 mL⁻¹) in a Random Sample of 75 Week Old Infants Whose Mother Received Vitamin D Supplement During Pregnancy

9.37	9.34	9.38	9.32	9.33	9.28	9.34
9.29	9.36	9.30	9.31	9.33	9.34	9.35
9.35	9.36	9.30	9.32	9.33	9.35	9.36
9.32	9.37	9.34	9.38	9.36	9.37	9.36
9.36	9.33	9.34	9.37	9.44	9.32	9.36
9.38	9.39	9.34	9.32	9.30	9.30	9.36
9.29	9.41	9.27	9.36	9.41	9.37	9.31
9.31	9.33	9.35	9.34	9.35	9.34	9.38
9.40	9.35	9.37	9.35	9.32	9.36	9.35
9.35	9.36	9.39	9.31	9.31	9.30	
9.31	9.36	9.34	9.31	9.32	9.34	

tendency and variability of the data than that presented by Table 2. In Table 3, data are grouped in six classes of equal size and it is possible to see the centering of the data about the 9.325–9.355 class and its variability: The measurements vary from 9.27 to 9.44 with ~95% of them between 9.29 and 9.41. The advantage of grouped frequency distributions is that grouping smoothes the data so that essential features are more discernible. Figure 2 represents the corresponding histogram. The sides of the bars of the histogram are drawn at the class boundaries and their heights are the frequencies or the relative frequencies (frequency/sample size). In the histogram, the distribution of the data centered about the point 9.34 is clearly seen. Although grouping smoothes the data, too much grouping (that is choosing too few classes) will tend to mask rather than enhance the sample’s essential features.

There are many numerical indicators for summarizing and describing data. The most common ones indicate central tendency, variability, and proportional representation (the sample mean, variance, and percentiles, respectively). We assume that any characteristic of interest in a population, and hence in a sample, can be represented by a number. This is obvious for measurements and counts, but even qualitative characteristics (described by discrete variables) can be numerically represented. For example, if a population is dichotomized into those individuals who are carriers of a particular disease and those who are not, a 1 can be assigned to each carrier and a 0 to each noncarrier. The sample can then be represented by a sequence of zeroes and ones.

The most common measure of central tendency is the sample mean:

$$M = (x_1 + x_2 + \dots + x_n)/n \quad \text{also noted } \bar{X} \quad (1)$$

Table 3. Frequency distribution of infant serum calcium data

Serum Calcium mg·100 mL ⁻¹	Frequency
9.265–9.295	4
9.295–9.325	18
9.325–9.355	24
9.355–9.385	22
9.385–9.415	6
9.415–9.445	1
Total	75

where x_1, x_2, \dots, x_n is the collection of numbers from a sample of size n . The sample mean can be roughly visualized as the abscissa of the horizontal center of gravity of the frequency histogram. For the serum calcium data of Table 2, $M = 9.34$, which happens to be the midpoint of the highest bar of the histogram (Fig. 2). This histogram is roughly symmetric about a vertical line drawn through M , but this is not necessarily true of all histograms. Histograms of counts and survival times data are often skewed to the right (long-tailed with concentrated mass at the lower values). Consequently, the idea of M as a center of gravity is important to bear in mind when using it to indicate central tendency. For example, the median (described later in this section) may be a more appropriate index of centrality depending on the type of data and the kind of information one wishes to convey.

The sample variance, defined by

$$s^2 = \frac{1}{n - 1} [(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2] \\ = \sum_{i=1}^n \frac{(x_i - M)^2}{n - 1} \quad (2)$$

is a measure of variability or dispersion of the data. As such, it can be motivated as follows: $x_i - M$ is the deviation of the i th data sample from the sample mean, that is, from the “center” of the data; we are interested in the amount of

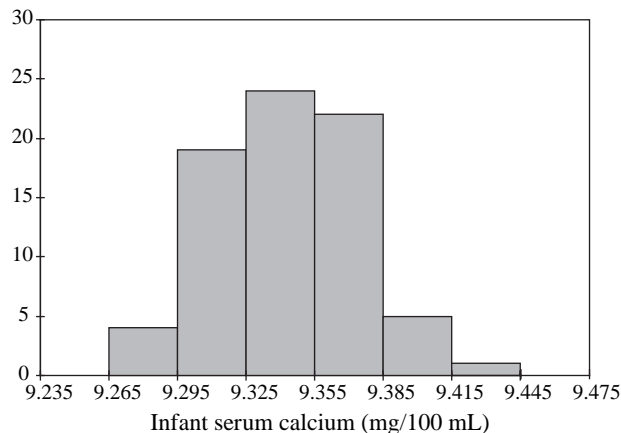


Figure 2. Frequency histogram of infant serum calcium data of Tables 2 and 3. The curve on the top of the histogram is another representation of probability density for continuous data.

deviation, not its direction, so the sign is disregarded by calculating the squared deviation $(x_i - M)^2$; finally, the squared deviations are averaged by summing them and dividing by the sample size $n - 1$. (Division by $n - 1$ ensures that the sample variance is an unbiased estimate of the population variance.) Note that an equivalent and often more practical formula for computing the variance may be obtained by developing Eq. 2:

$$s^2 = \frac{\sum x_i^2 - nM^2}{n - 1} \tag{3}$$

A measure of variability in the original units is then obtained by taking the square root of the sample variance. Specifically, the sample standard deviation, denoted s , is the square root of the sample variance.

For the serum calcium data of Table 2, $s^2 = 0.0010$ and $s = 0.03 \text{ mg}\cdot 100 \text{ mL}^{-1}$. The reader might wonder how the number 0.03 gives an indication of variability. Note that for the serum calcium data $M \pm s = 9.34 \pm 0.03$ contains 73% of the data, $M \pm 2s = 9.34 \pm 0.06$ contains 95% and $M \pm 3s = 9.34 \pm 0.09$ contains 99%. It can be shown that the interval $M \pm 3s$ will include at least 89% of any set of data (irrespective of the data distribution).

An alternative measure of central tendency is the median value of a data sample. The median is essentially the sample value at the middle of the list of sorted sample values. We say essentially because a particular sample may have no such value. In an odd-numbered sample, the median is the middle value; in an even-numbered sample, where there is no middle value, it is conventional to take the average of the two middle values. For the serum calcium data of Table 3, the median is equal to 9.34.

By extension to the median, the sample p percentile (say, e.g., 25th percentile) is the sample value at or below which $p\%$ (25%) of the sample values lie. If there is no value at a specific percentile, the average between the upper and lower closest existing round percentile is used. Knowledge of a few sample percentiles can provide important information about the population.

For skewed frequency distributions, the median may be more informative for assessing a population center than the mean. Similarly, an alternative to the standard deviation is the interquartile range: it is defined as the seventy-fifth minus the twenty-fifth percentiles and is a variability index not as influenced by outliers as the standard deviation.

There are many other descriptive and numerical methods (see, e.g., Ref. (2)). It should be emphasized that the purpose of these methods is usually not to study the data sample itself, but rather to infer a picture of the population from which the sample is taken. In the next section, standard population distributions and their associated statistics are described.

PROBABILITY, RANDOM VARIABLES, AND PROBABILITY DISTRIBUTIONS

The foundation of all statistical methodology is probability theory, which progresses from elementary to the most advanced mathematics. Much of the misunderstanding

and abuse of statistics comes from the lack of understanding of its probabilistic foundation. When assumptions of the underlying probabilistic (mathematical) model are grossly violated, derived inferential methods will lead to misleading and irrational conclusions. Here, only enough probability theory to provide a framework for this article is discussed.

In the rest of this article, experiments that have more than one possible outcome, the actual outcome being determined by some chance mechanism will be studied. The set of possible outcomes of an experiment is called its sample space; subsets of the sample space are called events, and an event is said to occur if the actual outcome of the experiment is a member of that event. A simple example follows.

The experiment will be the toss of a pair of fair coins, arbitrarily labeled coin number 1 and coin number 2. The outcome (1,0) means that coin No. 1 shows a head and coin No. 2 shows a tail. Then, the sample space by the collection of all possible outcomes can be specified:

$$S = \{(0,0)(0,1)(1,0)(1,1)\} \tag{4}$$

There are four ordered pairs so there are four possible outcomes in this coin-tossing experiment. Consider the event A "toss one head and one tail", which can be represented by $A = \{(1,0)(0,1)\}$. If the actual outcome is (0,1) then the event A has occurred.

In the example above, the probability for event A to occur is obviously 50%. However, in most experiments it is not possible to intuitively estimate probabilities, so the next step in setting up a probabilistic framework for an experiment is to assign, through some mathematical model, a probability to each event in the sample space.

Definition of Probability

A probability measure is a rule, say P , which associates with each event contained in a sample space S a number such that the following properties are satisfied:

1. For any event, A , $P(A) \geq 0$.
2. $P(S) = 1$ (since S contains all the outcomes, S always occurs).
3. $P(\text{not } A) + P(A) = 1$.
4. If A and B are mutually exclusive events (that cannot occur simultaneously) and independent events (that are not linked in any way), then

$$P(A \text{ or } B) = P(A) + P(B) \quad \text{and} \quad P(A \text{ and } B) = 0$$

Many elementary probability theorems (rules) follow directly from these definitions.

Probability and Relative Frequency

The axiomatic definition above and its derived theorems dictate the properties that probability must satisfy, but they do not indicate how to assign probabilities to events. The major classical and cultural interpretation of probabilities is the relative frequency interpretation. Consider an experiment that is (at least conceptually) infinitely

repeatable. Let A be any event and let n_A be the number of times the event A occurs in n repetitions of the experiment; then the relative frequency of occurrence of A in the n repetitions is n_A/n . For example, if mass production of a medical device reliably yields 7 malfunctioning devices out of 100, the relative frequency of occurrence of a defective device is 7/100.

The probability of A is defined by $P(A) = \lim n_A/n$ as $n \rightarrow \infty$, where this limit is assumed to exist. The number $P(A)$ can never be known, but if the experiment can in fact be repeated a large number of times, it can be estimated by the relative frequency of occurrence of A .

The relative frequency interpretation is an objective interpretation because the probability of an event is assumed to be independent of judgment by the observer. In the subjective interpretation of probability, a probability is assigned to an event according to the assigner's strength of belief that the event will occur, on a scale of 0–1. The assigner could be an expert in a specific field, for example, a cardiologist that provides the probability for a sample of electrocardiograms to be pathological.

Probability Distribution Definition and Probability Mass Function

It has been assumed that all data can be numerically represented. Thus, the outcome of an experiment in which one item will be randomly drawn from a population will be a number, but this number cannot be known in advance. Let the potential outcome of the experiment be denoted by X , which is called a random variable in statistics. When the item is drawn, X will be realized or observed. Although the numerical values that X will take cannot be known in advance, the random mechanism that governs the outcome can perhaps be described by a probability model. Using the model, the probability that the random variable X will take a value within a set or range of numbers can be calculated.

One such popular mathematical model is the probability distribution of a discrete random variable X . It can be best described as a mathematical equation or table that gives, for each value x that X can assume, the probability associated with this value $P(X = x)$. For example, if X represents the outcome of the tossing of a coin, there are two possible outcomes, tail and head. If it is a fair coin $P(X = tail) = 0.5$ and $P(X = head) = 0.5$. In statistics, the function $P(X = x)$ is called the probability mass function of X .

It follows from the relative frequency interpretation of probability that, for a discrete random variable or for the frequency distribution of a continuous variable, relative frequency histograms estimate the probability mass functions of this variable. For example, in Table 3, if the random variable X indicates the serum calcium measure, then

$$\hat{P}(X \text{ is in the first bin}) = \hat{P}(9.265 \leq X < 9.295) = 4/75$$

the $\hat{}$ symbol on P indicating estimated probability values, since actual probabilities describe the population itself and cannot be calculated from data samples. Similarly the probability that X is in the second bin,

the third bin, ... can be estimated and the collection of these probabilities constitute an estimated probability mass function.

Probability Density Function for Continuous Variables

The probability mass function above best describes discrete events, but what probabilities can be assigned to continuous variables? Since a continuous variable X can assume any value on a continuum, the probability that X assumes a particular value is 0 (except in very particular cases that will not be discussed here). Consequently, associated with a continuous random variable X , is a function f_X , called its probability density function that can be used to compute probability. The probability that a continuous random variable X assumes a value between values x_1 and x_2 is the area under the graph of f_X over the interval x_1 and x_2 ; mathematically

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f_X(x) dx \tag{5}$$

For example, for the infant serum data of Table 2 (see also Table 3), it can be estimated that the probability that an infant whose mother received a vitamin D supplement during pregnancy has between 9.35 and 9.38 mg·100 mL⁻¹ calcium is 22/75 or 0.293, which is the relative frequency of the 9.355–9.385 class in the sample. For continuous data, a smooth curve passing through the midpoint of a histogram bars' upper limit should resemble the probability density function of the underlying population.

There are many mathematical models of probability distribution. Three of the most commonly used probability distribution models described below are the binomial distribution and the Poisson distribution for discrete variables, and the normal distribution for continuous variables.

The Binomial Distribution

The scenario leading to the binomial distribution is an experiment that consists of n independent, repeated trials, each of which can end in only one of two ways arbitrarily labeled success or failure. The probability that any trial ends in a success is p (and hence $q = 1 - p$ for a "failure"). Let the random variable X denote the total number of successes in the n trials, and x denote a number in $\{0; \dots; n\}$. Under these assumptions:

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, \dots, n \tag{6}$$

with

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \tag{7}$$

where $n! = 1 * 2 * 3 * \dots * n$ is n factorial.

For example, suppose the proportion of carriers of an infectious disease in a large population is 10% ($p = 0.1$) and that the number of carriers follows a binomial distribution. If 20 individuals are sampled ($n = 20$) and X is the number of carriers (successes) in the sample, then the probability

that there will be exactly one carrier in the sample is

$$P(X = 1) = \binom{20}{1} (0.10)^1 (0.90)^{20-1} = 0.27$$

More complex probabilities may be calculated with the help of probability rules and definitions. For instance the probability that there will be at least two carriers in the sample is

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) \\ &\quad \text{(see third probability definition)} \\ &= 1 - P(X = 0 \text{ or } X = 1) \\ &= 1 - (P(X = 0) + P(X = 1)) \\ &\quad \text{(see fourth probability definition)} \\ &= 1 - \binom{20}{0} (0.10)^0 (0.90)^{20} - \binom{20}{1} (0.10)^1 (0.90)^{19} \\ &= 1 - 0.12 - 0.27 = 0.61 \end{aligned}$$

Historically, single trials of a binomial distribution are called Bernoulli variates after the Swiss mathematician James Bernoulli who discovered it at the end of the seventeenth century.

The Poisson Distribution

The Poisson distribution is often used to represent the number of successive independent events of a specified type (e.g., cases of flu) with low probability of occurrence (<10%) in some specified interval of time or space. The Poisson distribution is also often used to represent the number of occurrence of events of a specified type where there is no natural upper limit, for example, the number of radioactive particles emitted by a sample over a set time period. Specifically, X is a Poisson random variable if it obeys the following formula:

$$P(X = x) = e^{-\lambda} \lambda^x / x! \quad x = 0, 1, 2, \dots \quad (8)$$

where $e = 2.178 \dots$ is the natural logarithmic base and λ is a given constant. For example, suppose the number of a particular type of bacteria in a standard area (e.g., 1 cm²) can be described by a Poisson distribution with parameter $\lambda = 5$. Then, the probability that there are no more than 3 bacteria in the standard area is given by

$$\begin{aligned} P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= e^{-5} 5^0 / 0! + e^{-5} 5^1 / 1! + e^{-5} 5^2 / 2! + e^{-5} 5^3 / 3! \\ &= 0.265 \end{aligned}$$

Note that the Poisson and the binomial distributions are closely related. In the case of a rare event ($p < 10\%$), the binomial distribution (described by probability p and n events) is well approximated by the Poisson distribution with the constant $\lambda = np$. The Poisson distribution was named after the French mathematician Siméon-Denis Poisson, who discovered it in the early part of the nineteenth century.

The Normal Distribution

The binomial and Poisson distributions describe discrete events, but there are also many distributions describing

continuous variables. The most important one is the normal distribution (also called Laplace–Gauss distribution as it was discovered by the French astronomer Pierre–Simon Laplace and the German mathematician Karl Friedrich Gauss in the early nineteenth century). Normal distributions arise as a result of many small random fluctuations about some general average (e.g., repeated recordings of a constant body temperature using a noisy electronic thermometer). A random variable X is said to be a normal or Gaussian random variable with mean parameter μ and standard deviation parameter σ if its probability density function is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad -\infty < x < \infty \quad (9)$$

The normal probability density function graphed in Fig. 3, is bell shaped with tails rather rapidly receding to zero height. Because f_X represents probability density, the total area bounded by the curve is 1 (see Eq. 9). The area between two values of variable X (x_1 and x_2 where $x_1 < x_2$) represents the probability that X lies between x_1 and x_2 , (Eq. 5).

As shown in Fig. 3, if X is normal (μ, σ), it can be calculated that $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.997$, which, according to the relative frequency interpretation of probability, states that $\sim 99.7\%$ of a large sample from a “normally distributed population” will be contained in the interval mean plus or minus three standard deviations ($\mu \pm 3\sigma$).

Note that there is a relation between the normal and the binomial distribution. Using the same notation as in Eq. 6, if n , the number of samples, is large enough then the variable z defined as

$$z = \frac{x - np}{\sqrt{npq}} \quad (10)$$

is approximately normally distributed with mean 0 and standard deviation 1. In a coin throwing experiment, throwing the coin a large number of times and counting

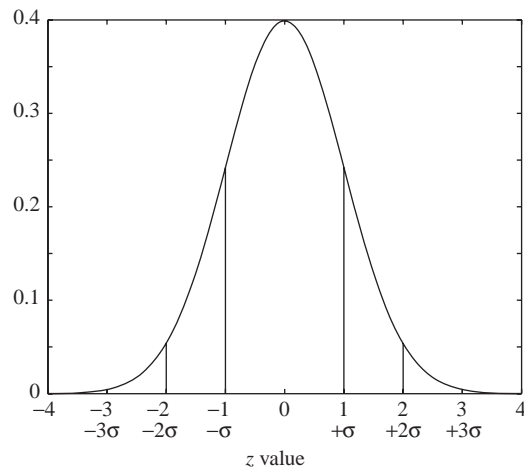


Figure 3. The normal probability density function showing symmetry about a vertical line through μ , and the role of σ as a variability parameter. Vertical bars indicates $\pm\sigma, \pm 2\sigma, \pm 3\sigma$.

the number of heads x , then building a histogram for the value z , the histogram will be close to a normal distribution (as shown in Fig. 3). Similarly, there is a relation between the Poisson and the normal distribution, the variable z defined as $z = (x - \lambda) / \sqrt{\lambda}$ is normally distributed for large values of λ .

Many statistical inferential methods described in the next section assume that the data is approximately normally distributed. Much abuse occurs, however, when these methods are applied blindly with no verification of the normality assumption. Incidentally, methods that incorporate assumptions of normality often can be applied to non-normal situations because under certain conditions, the normal distribution can approximate other distributions, such as the binomial and the Poisson distributions. Sometimes, the data can also be preprocessed to fit the normal distribution. For example, a histogram might indicate nonnormality, while a histogram of the logarithms of the data would fit the normal distribution, indicating that normal-based models can be applied to the log-transformed data. These transformations are discussed in most experimental design textbooks.

The importance of the normal distribution in statistics is also due to the central limit theorem in statistics that states that the distribution of any linear mixture of two or more independent random variables is more normal (has a shape closer to the normal distribution) than the distribution of the random variables themselves. This property is used by some algorithms processing multivariate data (as described in a later section).

There are many other continuous probability distributions besides the normal distribution. For example, the most commonly used distribution in survival analysis is the Weibull distribution. The von Mises distribution allows parametric statistical tests for periodic data (i.e., seasonal).

Characteristics of Probability Distributions

Just as there are numerical indexes for sample description, for example, sample means, variances, and percentiles, there are numerical characteristics of probability distributions. The expectation or mean (not sample mean) of an random variable X is

$$\begin{aligned}
 E(X) &\equiv \sum_x xP(X = x) && X \text{ discrete} \\
 &\equiv \int_{-\infty}^{\infty} x f_X(x) dx && X \text{ continuous}
 \end{aligned}
 \tag{11}$$

The expectation E is a measure of central tendency for a population (i.e., the center of gravity of the probability distribution about the y axis). The variance of X is defined in terms of expectation by

$$\text{Var}(X) \equiv E\{[X - E(X)]^2\}
 \tag{12}$$

In words, $\text{Var}(X)$ is the expected squared deviation of X from $E(X)$, and in this sense is a measure of variability or dispersion for a population. The standard deviation of X is the square root of its variance. Table 4 indicates mean and variance for the binomial, the Poisson, and the normal distribution.

Numerical descriptors of populations are often the very things we want to know about populations. They should not

Table 4. Mean and Variance for Standard Distributions (see text for details)

	Binomial	Poisson	Normal
Mean	$\mu = Np$	$\mu = \lambda$	μ
Variance	$\sigma^2 = Npq$	$\sigma^2 = \lambda$	σ^2

be confused with their sample counterparts; the sample numerical descriptors are the basis for drawing inferences regarding their population counterparts, which are of primary interest.

Statistical Inference

A statistical hypothesis is a statement about the probability distribution of populations using one or more data samples. Typical questions are Is this single data sample consistent with this theoretical distribution of values?, Are these two data samples originating from the same population?, Are these n data samples originating from the same population?. Associated with each of these questions, in statistics, two hypotheses are usually formulated.

Hypothesis H_0 : All data samples originate from the same population (or the single data sample is consistent with a given theoretical distribution).

Hypothesis H_1 : Some data samples do not originate from the same population (or the single data sample is not consistent with the given theoretical distribution).

The test is called significant if hypothesis H_0 is rejected with respect to a user-defined confidence interval (e.g., 5% of chance of wrongly rejecting H_0). It is important to remember that inference tests can never disprove hypothesis H_0 . Instead, based on the significance threshold and on the inference test chosen, it can be said that the data support rejecting H_0 . The test is called nonsignificant if the hypothesis H_0 and reject hypothesis H_1 is accepted. Accepting H_0 means that we failed to find any significant difference with respect to our user-defined confidence interval. Because the error in accepting H_0 is usually large (see error types below), in general we should avoid drawing any conclusion about the experiment when accepting H_0 .

Degree of Freedom. Elementary tests usually depend on the data sample size as well as the number of parameters (e.g., mean or variance) that have to be estimated from the sample, to run the test. Specifically, the number of degrees of freedom of a statistics is defined as the number of independent observations minus the number of population parameters, which must be estimated from sample observations. Details will be provided for each test.

p -Values. Once hypotheses H_0 and H_1 have been defined, that a test has been chosen to address these hypotheses (see below), and that parameters for this test have been calculated, one must choose a level of significance. The term $p < 0.05$ is the arbitrary value that is generally accepted to be significant. This means that there must be $< 5\%$ possibility of falsely detecting a significant difference. Now describe how the p value relates to the different types of errors associated with elementary tests.

Type I and Type II Errors. If a hypothesis H_0 is rejected when it should be accepted, it can be said that a type I error has been made. If a hypothesis H_0 is accepted when it should be rejected, it can be said that a type II error has been made. In either case, a wrong decision or judgment has occurred. This is not a simple matter because decreasing one error type usually leads to increasing the other error type. One way of getting around this problem is just to set your significance level at 0.05 (and not at 0.01 or 0.001). In this way, you are balancing between type I and type II errors in your decision making process. One way to decrease both error types is to increase the size of the sample. However, two ways of analyzing the same size dataset (i.e., two types of inference test) might have different efficiency, so that the more efficient might give better performance on both error types. As an example of type I and type II errors, let us imagine that there is a significant difference between the average of blood pressure measured from a population of patients and the general population at $p = 0.05$. Then there will be a 5% chance that our statement is false (type I error). This means that if we repeat the test 100 times, when in fact no real effects are present, we will draw a wrong conclusion ~5% of the time that we observe a significant difference. In contrast, if we state that there is no such difference between population of patients at $p = 0.05$, there is not a 5% chance of being wrong, but usually more (type II error). This is why, in general, when accepting hypothesis H_0 , no conclusions should be drawn about the results of an experiment. The exact calculation of type II error usually depends on the size of the actual effect in the population, hence it is usually described by curves as a function of effect magnitude.

Correction for Multiple Comparisons. When multiple tests are performed, the probability that one of them is significant by chance becomes larger. As for type I error, if 100 tests are performed with significance threshold of $p = 0.05$, when in fact no real effects are present, then on average about five of them will indicate significance, but will be false positives. This is the case, for example, when processing biophysical images, such as magnetic, resonance imaging data: a collection of values is acquired for each coordinate on a three-dimensional (3D) grid and a statistical test must be performed on this data. The same problem may arise when processing time series data. The standard conservative approach developed by Bonferroni (3) consists of dividing the p -value threshold by the number of comparisons performed. For example, for 100 tests performed at $p = 0.05$, the corrected p value is $0.05/100 = 0.0005$. This is a conservative approach and a less stringent method has been developed by Holm (4): first choose a significance level $p = \alpha$ (e.g., $p = 0.05$). Then compute the exact p -value for each test, which is usually possible using modern computerized approaches. Rank the collection of p -values from smallest to largest. The smallest p -value is tested against α/N , where N is the number of tests. If the smallest p -value is not $< \alpha/N$, stop the procedure. However, if it is $< \alpha/N$, proceed to test the second smallest p -value against $\alpha/(N - 1)$, and so on. A variant of the Holm's procedure consist of testing the first p -value against α/N , the second one against $2\alpha/N$, the third one against $3\alpha/N$, and so on. Technical details and theory about multiple comparisons may be found in (5).

Paired/Unpaired Samples. Table 5 distinguishes between paired versus unpaired data samples. For

Table 5. Which Statistical Inference Test to Use for Which Type of Data

Goal	Dataset		
	Binomial or Discrete	Continuous measurement (from a normal distribution)	Continuous measurement, Rank, or Score (from non-normal distribution)
Example of data sample	List of patients recovering or not after a treatment	Readings of heart pressure from several patients	Ranking of several treatment efficiency by one expert
Describe one data sample	Proportions	Mean, SD	Median
Compare one data sample to a hypothetical distribution	χ^2 or Binomial test	One-sample t test	Sign test or Wilcoxon test
Compare two paired samples	Sign test	Paired t test	Sign test or Wilcoxon test
Compare two unpaired samples	χ^2 Fisher's exact test	Unpaired t test	Mann-Whitney test
Compare three or more unmatched samples	χ^2 test	One-way ANOVA ^a	Kruskal-Wallis test
Compare three or more matched samples	Cochrane Q test	Repeated-measures ANOVA ^b	Friedman test
Quantify association between two paired samples	Contingency coefficients	Pearson correlation	Spearman correlation

^aAll statistical tests in this table are described in the text and often instantiated using a numerical example.

^bAnalysis of Variance = ANOVA

unpaired data samples, there is no direct correspondence between values. This may be the case when a specific measure (e.g., blood pressure) is taken from two distinct populations of patients (e.g., patients suffering from heart failure and control patients). The two data samples corresponding to the two groups of patients are said to be unpaired because there is no relationship between them. In contrast, for paired samples, each value in one sample corresponds to a value in the other sample. In the previous example, it could be the case if each patient tested had a twin volunteering to be a control patient. This would also be the case if two assessments were performed on the same patients (e.g., measure of blood pressure before and after taking a drug). Note that paired groups must necessarily be of the same size. Matched/unmatched data samples are an extension of paired/unpaired data samples when there are more than two samples.

Sampling With or Without Replacement. Sampling with replacement means that each item is put back in the data sample after being sampled (so it may be sampled more than once and appear twice or more in a data sample). Sampling with replacement satisfies the requirement that the trials are independent, but when the sample size is small relative to the size of the population, sampling with or without replacement makes little difference. In elementary statistics, a representative sample is synonymous with the concept of a random sample. When sampling from a population of finite size, a sample of n items is a random sample if it is chosen in such a way that any other sample of size n would be equally likely to be chosen. Sampled items can be chosen with or without replacement. Although impractical in many situations, sampling with replacement leads to easier mathematical analysis. When the population is large relative to the sample size, the analytical methods developed for sampling with replacement yield good approximations. A random sample can be chosen by assigning a number to each member of the population, and then choosing at random n numbers (with or without replacement). This can be done by the so-called Monte Carlo method (consisting of random draws) that uses computer-generated (pseudo)-random numbers.

Table 5 indicates which statistical test should be used depending on data type and question type. Most types of questions have already been described when hypotheses H_0 and H_1 were defined. The last row of Table 5, which is concerned with the relationship between data samples (or more specifically the relationship between variables underlying two paired data samples) was not covered. The corresponding question may be formulated as ‘Is there any relationship between the two variables (e.g., two paired measurements)?’ The H_0 hypothesis is that there is no relationship between the two variables.

Columns of Table 5 contain elementary tests for different types of variables. Elementary tests cover confidence interval estimation and parametric hypothesis testing for situations involving normally distributed samples, including two-sample situations where the purpose is to compare two populations with respect to their means or variances. Other types of elementary confidence intervals are for proportions and difference of proportions, usually based

on the binomial distribution or based on the normal approximation to the binomial distribution. Confidence interval estimation for parameters of nonnormal distributions are much more difficult and closed form formulas often do not exist. In these cases, experimenters must use nonparametric statistical tests that only take into account rank ordering of data samples. They may also use resampling statistical tests, which estimates confidence intervals using many computer-generated random resamplings. For practicality, in Table 5, hypothesis testing was divided into three main categories: hypothesis testing on discrete variables, parametric statistical testing on continuous variables, and nonparametric statistical testing on continuous variables. In a separate section, resampling methods will be dealt with, since it may be applied to any type of data. The list of tests is not exhaustive but instead seeks to provide, within the limited scope of this short article, a range of methods to perform statistical inference on different types of data.

Which type of test to use is often one of the most delicate choices an experimenter has to make. For continuous data, for example, one could use at least three tests: a parametric, a nonparametric, or a resampling inference test. Different tests make different assumptions: parametric test, such as the t -test, make the hypothesis that the data is normally distributed. Nonparametric tests make fewer assumptions about the population distribution but require more data samples. Resampling tests make the assumption that the data samples are an accurate representation of the population. There is no ideal test (although some applied statisticians would argue that resampling methods are indeed superior to other methods), and the test to choose often depends on the type of data being processed or common usage in one specific field of research.

Testing Hypothesis on Discrete Variables

For discrete variables, data is most often represented by proportions of different outcomes. As shown in the first column of Table 5, specific tests have been designed to deal and compare proportions between data samples. Some of these tests (as indicated below) can only deal with binomial data samples (success or failure).

Goodness of Fit to Distribution for One Data Sample. A goodness of fit test may be used to compare one data sample to a hypothetical value or distribution. In a goodness-of-fit test the hypotheses are concerned with the distribution itself. For example, a drug has been repeatedly tested on adults and has shown minor side effects in 2.5% of the cases in which it was administered. To validate this drug for treating children, it is given to a sample of 300 children. The goal of this study is to determine if children showed more or less side effects than adults. The hypothesis H_0 is that the distribution of sample data values for children is generally the same as the hypothetical distribution for adults. The hypothesis H_1 is that the distribution of sample data values for children generally differs from the hypothetical distribution for adults. Table 6 indicates that 13 out of 300 children showed an abnormal reaction to the drug. The

Table 6. Measured and Expected Frequencies of Side Effect for 300 Children Treated With a Test Drug

	Children	Expected value
Side effect	13	7.5
No side effects	287	292.5
Total	300	300

second column in Table 6 indicates the expected values from the theoretical distribution (2.5% of cases for 300 subjects is 7.5 individuals; it is not so important that the expected value is not a whole number since this distribution is only theoretical).

The χ^2 value is then simply calculated by comparing the expected frequencies e_1 (7.5 individuals showing side effects) and e_2 (292.5 individuals showing no side effects) to the observed frequencies O_1 (13) and O_2 (287) using the formula:

$$\chi^2 = \frac{(O_1 - e_1)^2}{e_1} + \frac{(O_2 - e_2)^2}{e_2}$$

or more generally

$$\chi^2 = \sum_i \frac{(O_i - e_i)^2}{e_i} \tag{13}$$

where O_i is the frequency observation in row i and e_i is the corresponding expected frequency. The degrees of freedom is equal to $(n - 1)$, where n is the number of rows in the table. Once the χ^2 value and the degrees of freedom have been calculated, the critical value for χ^2_{crit} can be looked up in Table 7 for a given level of significance. If $\chi^2 > \chi^2_{crit}$, we reject hypothesis H_0 in favor of hypothesis H_1 and conclude that the data support the hypothesis that there is a difference between the sample data and the theoretical distribution at the 5% level of significance.

In the example shown in Table 6,

$$\chi^2 = \frac{(13 - 7.5)^2}{7.5} + \frac{(287 - 292.5)^2}{292.5} = 4.13$$

with 1 degree of freedom (2 rows minus 1). For a test at the 5% level of significance ($p = 0.05$) with 1 degree of freedom, χ^2_{crit} in the χ^2 table (Table 7) is equal to 3.84. Since $4.13 > 3.84$, the hypothesis that the proportion of children having side effects in the same as that of adults is rejected. Comparing actual and expected frequencies in Table 6, we conclude that children have higher occurrences of side effects than adults for this drug.

Note that the construction of the χ^2 table is relatively simple. One can simply assume that a known population (whose expected distribution is known) is sampled several times and that the χ^2 value is computed for each of these samples. The histogram of these observed χ^2 values, when in fact no real effects are present, is an approximation to the χ^2 distribution for the null hypothesis (Fig. 4). The tails of this distribution may be used to set thresholds for significance testing (if an observed χ^2 value ends up in the tail of the distribution, then it is likely that it does not originate from the known population). For example, the χ^2

Table 7. χ^2 Distribution of Critical Values^a

	$p = 0.05$	$p = 0.01$	$p = 0.001$
1df	3.84	6.64	10.83
2	5.99	9.21	13.82
3	7.82	11.35	16.27
4	9.49	13.28	18.47
5	11.07	15.09	20.52
6	12.59	16.81	22.46
7	14.07	18.48	24.32
8	15.51	20.09	26.13
9	16.92	21.67	27.88
10	18.31	23.21	29.59
11	19.68	24.73	31.26
12	21.03	26.22	32.91
13	22.36	27.69	34.53
14	23.69	29.14	36.12
15	25.00	30.58	37.70
16	26.30	32.00	39.25
17	27.59	33.41	40.79
18	28.87	34.81	42.31
19	30.14	36.19	43.82
20	31.41	37.57	45.32
21	32.67	38.93	46.80
22	33.92	40.29	48.27
23	35.17	41.64	49.73
24	36.42	42.98	51.18
25	37.65	44.31	52.62
26	38.89	45.64	54.05
27	40.11	46.96	55.48
28	41.34	48.28	56.89
29	42.56	49.59	58.30
30	43.77	50.89	59.70
35	49.80	57.34	66.62
40	55.76	63.69	73.41
50	67.51	76.15	86.66
60	79.08	88.38	99.62
70	90.53	100.42	112.31
80	101.88	112.33	124.84
90	113.15	124.12	137.19
100	124.34	135.81	149.48

^aTo use this table, choose a p value (column) and read the value for your calculated degrees of freedom (df). If your calculated χ^2 value is larger than the one you read in the table, the test you performed is significant (see text for details).

value for a data sample is significantly different from the χ^2 standard distribution at $p = 0.05$ if it lies in the lower or upper tails each containing only 2.5% of the values of the standard χ^2 distribution.

Binomial Test for Binomial Variables. For data samples that are assumed to be obeying the binomial distribution, it is possible to compute exact p values as explained in a previous section. For example, a coin is tossed 10 times to determine if it returns fair results or not. It returns 9 heads. The hypothesis H_0 is that the coin is fair and that the probability of obtaining a head is 0.5. The H_1 hypothesis is that the coin is biased toward head. Using the binomial distribution, the probability of obtaining an equal or more extreme number of heads than the one measured needs to be computed. The probability of obtaining 9 heads

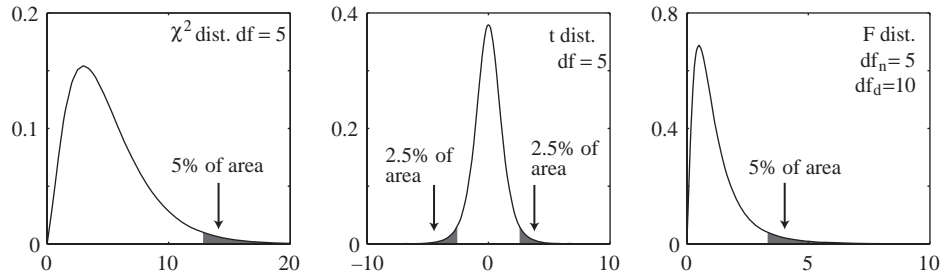


Figure 4. Standard distributions (χ^2 , t , and F). Tails of these distributions are used to determine significance thresholds (see text).

or more is

$$P(X \geq 9) = P(X = 9) + P(X = 10) = \binom{10}{9} 0.5^9 (1 - 0.5)^{10-9} + \binom{10}{10} 0.5^{10} = 0.011$$

It appears that this result would appear by chance in only 1.1% of coin tossing trials if the coin is returning fair results. If $p < 0.05$ is considered to be the standard threshold for significance, it can be concluded that the coin does return more heads than a fair coin at the 5% level of significance. Note that this was a one-sided test, assuming that there was prior knowledge that the coin will be biased toward heads (based, e.g., on the aspect of the coin): for a two-sided test that would assess if the coin is fair in returning both faces and heads, it would be necessary to add the probability of obtaining both 9 to 10 heads and 9 to 10 faces.

Sign Test to Compare Paired Samples. This test is best illustrated by an example. To determine if drug *A* is more effective than a drug *B* for pain control, 10 patients are tested with these two drugs (with an interval of 7 days to prevent carry over effects) and asked if the drug was effective in controlling their pain. Hypothetical results are shown in Table 8, with + signs indicating a positive effect of the drug and - signs indicating no effect of the drug. The last row indicates the sign of the difference between the first two rows: a + sign indicates that drug *A* is performing better than drug *B* and a - sign indicates that drug *B* is performing better than drug *A*. When the outcome is the same, the cell is left empty. If the two drugs are equally effective, and if the sample is large enough, then there should be approximately equal numbers of + and - signs in the last row. The expected number of + signs (6 out of 7 nonempty cells) using binomial probability (note that for a large number of values, the approximation of the binomial distribution by the normal distribution may be used). We need to compute the probability of obtaining an

equally or more extreme number of + or - than the one obtained, hence to compute $P(0, 1, 2, 5, 6, 7)$:

$$P(+ = 0, 1, 2, 5, 6, 7) = \binom{7}{0} 0.5^0 (1 - 0.5)^{7-0} + \binom{7}{1} 0.5^1 (1 - 0.5)^{7-1} + \dots = 0.45$$

Although it seems that drug *A* is a better pain killer than drug *B*, the p value did not reach significance ($p > 0.05$). In other words, H_0 , the hypothesis that the two drugs are performing equally well cannot be rejected. This type of test applied to binomial variables is also sometimes called the Mc Nemar's test.

χ^2 Test to Compare Two or More Unpaired Samples. The χ^2 test allows the comparison of proportions observed in several groups under two or more conditions. Suppose that we wish to determine which of four prosthetic devices perform better for improving muscular response. Each of the four devices is implanted in four random samples of 100 patients each. For each patient, a clinician then estimates if there has been no improvement or partial to full restoration. Data is cross-classified as shown in Table 9. The test described here is usually called the χ^2 test of independence because it aims at finding if results from different groups can or cannot originate from the same population.

Here, the objective is to determine whether improvement is independent of the type of device. If it is the case, then the proportion of responses with no improvement and partial to full restoration should be similar for all four types of devices. The χ^2 test allows the comparison of the actual proportion of responses to each type of device to the idealized proportions, where all types of devices perform equally well. These proportions (also called expected frequencies) are calculated by pooling the responses for all types of devices. For instance, in Table 9, irrespective of the device type, there are 120 patients showing no restoration and 280 patients showing some degree of restoration, so the expected frequency for no restoration is 30% and the expected frequency for partial to full restoration is 70%.

Table 8. Success (+) or Failure (-) of Drug A and B for Reducing Pain in 10 Patients

Patient	1	2	3	4	5	6	7	8	9	10
Drug A	+	+	+	-	+	+	+	+	+	+
Drug B	+	-	-	+	+	-	-	+	-	-
Sign		+	+	-		+	+		+	+

Table 9. Results of Improvement in Muscular Response Following Implantation of an Electronic Device Available in Four Types

	Type of Device				Total
	1	2	3	4	
No improvement	35(30) ^a	40(30)	35(30)	10(30)	120
Partial-to-full restoration	65(70)	60(70)	65(70)	90(70)	280
Total	100	100	100	100	400

^aNumbers are observed frequencies and number in parentheses are expected frequencies.

As for the simpler example earlier in this section comparing a sample data distribution to a theoretical distribution, the χ^2 is simply calculated by comparing the expected frequencies, denoted by $e_{i,j}$ for device i (where i ranges from 1 to 4) and outcome j (where $j = 1$ indicates no restoration and $j = 2$ indicates partial to full restoration), to the observed frequencies $O_{i,j}$ using the formula:

$$\chi^2 = \sum_{i,j} (O_{i,j} - e_{i,j})^2 / e_{i,j} \tag{14}$$

The degrees of freedom is equal to $(r - 1)(c - 1)$, where r and c are the number of rows and columns in the table. Once the χ^2 value and the degrees of freedom have been calculated, the critical value for χ^2_{crit} can be looked up in Table 7 for a given level of significance. If $\chi^2 > \chi^2_{crit}$, then there is a significant difference between the groups being compared.

For the example shown in Table 9,

$$\chi^2 = \frac{(35 - 30)^2}{30} + \frac{(65 - 70)^2}{70} + \frac{(40 - 30)^2}{30} + \dots = 26.2$$

The degrees of freedom is $(4 - 1)(2 - 1) = 3$. In this example, for a test at the 5% level of significance ($p = 0.05$) and three degrees of freedom, Table 7 indicates that $\chi^2_{crit} = 7.82$. Since $26.2 > 7.82$, the hypothesis that all four devices are equally effective is rejected. It can be seen that device type 4 is most effective. In fact, further analysis supports the conclusion that differences between the other device types can be explained by sampling variation, and that there is a statistically significant difference between the first three device types taken together and the fourth device type. The additional analysis is sensible because the first three types are different vintages of essentially the same design, whereas the type four device is an experimental version of a fundamentally different design.

The χ^2 test may be used on a table of any size and not necessarily on binomial variables. For the example shown in Table 9, three possible outcomes could be imagined: no improvement, partial restoration, and full restoration. This would have added a row to Table 9 but the χ^2 formula (Eq. 14) would still apply.

Quantify Relationship between Variables. Classification in a table often reflects characteristics of individuals or objects, so they are often referred to as attributes. A measure of the degree of relationship, association, or dependence of two attributes (and the associated variables in the population) is called the coefficient of correlation. It

is given by

$$r = \sqrt{\frac{\chi^2}{N(\min(\text{No. rows}, \text{No. columns}) - 1)}} \tag{15}$$

where χ^2 represents the value computed from the χ^2 table; N is the total number of observations, and $\min(\text{No. rows}, \text{No. columns})$ represents the smaller number between the number of rows (No. row) and the number of columns (No. columns). r can only take values between 0 and 1. The closer r is to 1, the greater the association between the two (or more) columns of the table. To determine if a value of r is significant or not, χ^2 tests previously described in this section may be used.

Parametrical Hypothesis Testing on Continuous Variables

A parametric statistical hypothesis assumes that the data sample originates from a population that fits a specific model (most often the normal model). This is usually the case when recording a measure that fluctuates around a fixed mean because of environmental noise. Before running any statistical tests, one must verify that the data distribution is consistent with the normal distribution. First, plot the histogram to check that the distribution’s overall shape is similar to that of the normal distribution. You may then perform a goodness of fit test with the normal distribution. In a goodness-of-fit test, the hypotheses are concerned not with the parameters, but with the distribution itself. For example, H_0 : X has a normal distribution; H_1 : X does not have a normal distribution. This may be done using the χ^2 goodness of fit test (mentioned in the previous section) applied to the data histogram frequencies compared to expected values calculated from the normal distribution (by integrating Eq. 9 using Eq. 5). Other goodness-of-fit tests are the Kolmogorov–Smirnov, Cramer–Von Mises, and Anderson–Darling. There are also tests when H_0 involves some specific distribution, for example, the Shapiro–Wilk test for normality. Most computer packages incorporate such tests.

One-Sample t-Test to Compare One Data Sample to a Hypothetical Distribution. This test is used to determine if a data sample belongs to a population with mean μ and standard deviation σ (the hypothesis H_0 is that it does belong to this population). This test applies to continuous or noncontinuous data that have a distribution that is not significantly different from normal. First, check that the standard deviation of the data sample is similar to the

Table 10. The *t* Distribution of Critical Values^a

One-Tailed	0.1	0.05	0.025	0.01	0.005	0.0001
Two-Tailed	0.2	0.1	0.05	0.02	0.01	0.0002
df						
1	3.078	6.314	12.71	31.82	63.66	318.3
2	1.886	2.920	4.303	6.965	9.925	22.33
3	1.638	2.353	3.182	4.541	5.841	10.21
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.611
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
50	1.295	1.676	2.009	2.403	2.678	3.261
60	1.296	1.671	2.000	2.390	2.660	3.232
80	1.292	1.664	1.990	2.374	2.639	3.195
100	1.290	1.660	1.984	2.364	2.626	3.174
1000	1.282	1.646	1.962	2.330	2.581	3.098
inf.	1.282	1.64	1.960	2.326	2.576	3.091

^aTo use this table, find your degrees of freedom in the *df* column (or a lower one if yours is not present in the table). Then, look up the probability in the top row ($p = 0.05$ is a test of significance at 5%). If your calculated *t* value is larger than the one you read in the table, the test you performed is significant (see text for details).

population’s standard deviation σ (within a twofold range). For a data sample containing N values that has a mean M and standard deviation SD , the variable t is defined as

$$t = \frac{M - \mu}{SD} \sqrt{N} \tag{16}$$

The degrees of freedom associated with t is equal to $df = N - 1$. After calculating t and df , set up a threshold for significance (i.e., $p < 0.05$) and look up t_{crit} critical value in Table 10. In the t -test table, you may choose either one- or two-tailed t -test critical values. One-tailed t -tests are used when there is some prior knowledge to predict the direction of the difference. Most commonly, two-tailed t -tests are used when there is no such knowledge. If the calculated t value is $> t_{crit}$, there is a statistically significant difference between the data sample and the hypothetical distribution (the null hypothesis H_0 is rejected).

As for the χ^2 table, building the t -test table is straightforward. One may assume that, for a given degree of freedom, a known population (with a normal distribution) is sampled several times and that the t value is computed for each of these samples (M should on average be the same as μ since it is the theoretical population which is being sampled). The histogram of these observed t values, obtained when in fact no real effects are present, is an approximation to the t distribution (Fig. 4, middle panel). The tails of this distribution allow for threshold setting for significance (as for the χ^2 , if an observed value ends up in the tail of the distribution, then it is likely that it does not belong to this distribution). Note that for an infinite number of degrees of freedom, the t distribution is equal to the normal distribution.

For example, in the past, a machine has been producing washers having a thickness of 0.06 in. (0.15 cm) on average.

To test if the machine is still working properly, 10 washers of size (0.065; 0.062; 0.060; 0.059; 0.061; 0.064; 0.067; 0.064; 0.061; 0.062) are produced. The sample mean is 0.0625 and the sample standard deviation is 0.0025. The t value is equal to

$$t = \frac{0.0625 - 0.06}{0.0025} \sqrt{10} = 3.16$$

A two-tailed t -test is used since there is no *a priori* knowledge about the sampled distribution. At the 5% significance level, $t_{\text{crit}5\%}$ is equal to 1.83. Since $t > 1.83$, it can be concluded that there is a significant difference between the expected washer thickness and the observed one (reject hypothesis H_0 , which assumes that the sample distribution has a mean of 0.06 in., 0.15 cm). However, at 0.5% significance level, $t_{\text{crit}1\%}$ is equal to 3.25. Since $t < 3.25$, it cannot be concluded that such a difference exists at this level of significance (hypothesis H_0 cannot be rejected).

Paired t-Test to Compare Paired Data Samples. This test applies to two paired samples of continuous or noncontinuous data that have a distribution nonsignificantly different from normal and similar standard deviations (with less than twofold difference). First calculate the difference between each pair and average them (D_{av}) (note that differences in values also have to be normally distributed). Then calculate the value of t using

$$t = \frac{D_{\text{av}}}{SD} \sqrt{N} \tag{17}$$

where SD is the standard deviation of the difference between each pair. Since the accuracy of a statistic is influenced by the population size, the degrees of freedom (df) or the number of independent parameters used in the calculation of the test statistic must then be calculated. The degrees of freedom is equal to the degrees of freedom used in calculation the sample SD , that is, the number of pairs minus 1: $df = N - 1$.

Finally, as for the one sample t -test, set up a threshold for significance, look up t_{crit} critical value in Table 10, and compare it to the calculated value. If the calculated t value is $> t_{\text{crit}}$, there is a statistically significant difference between the two groups (the null hypothesis H_0 is rejected).

For example, to test if a newly designed electronic blood pressure (BP) device returns similar (hypothesis H_0) or different (hypothesis H_1) readings compared to an old manual blood pressure device, readings on 10 patients are performed and presented in Table 11 (only systolic pressure in Hgmm is reported in the table).

First, ensure that the two standard deviations are similar (14.1 for the electronic BP device and 13.5 for the manual BP device). To calculate the t value, compute the

difference between each pair, check that their distribution is normal, and then average them, $D_{\text{av}} = ((121 - 115) + (130 - 131) + \dots) / 10 = 2.8$. The standard deviation of the difference is $SD = 2.57$, and the degrees of freedom is 9 (10 readings minus 1). Thus the t value is equal to

$$t = \frac{2.8}{2.57} \sqrt{10} = 3.44$$

At the 5% level of significance, for 9 degrees of freedom, $t_{\text{crit}5\%}$ is equal to 2.26. Since $t > 2.26$, it can be concluded that the two devices return different averages (reject hypothesis H_0). The newly devised electronic BP device probably has to be recalibrated to better match the readings of the manual one.

Unpaired t-Test to Compare Unpaired Data Samples. An unpaired t -test aims to compare two unpaired data samples and applies to continuous or noncontinuous data that have a distribution not significantly different from normal. Sample sizes should be similar (with less than twofold difference) for the two groups and, if $n < 30$, variances should also be similar (with less than twofold difference). If the t -test is used in other circumstances, the results will have no meaning.

The most common way of calculating the t -statistics for unpaired data samples is to use the pooled variance estimate (it is also possible to use unpooled variance estimates, but this is less common and will not be presented here). First, calculate the unbiased pooled variance estimate:

$$V = \frac{V_A(N_A - 1) + V_B(N_B - 1)}{N_A + N_B - 2} \tag{18}$$

Then estimate the standard error of the difference of the means:

$$SE = \sqrt{V(1/N_A + 1/N_B)} \tag{19}$$

Then the t statistics is the difference of the means divided by its estimated standard error:

$$t = \frac{M_A - M_B}{SE} \tag{20}$$

where M_A , and M_B are the means of groups A and B, respectively, and where V_A and V_B are the variances of groups A and B, respectively. For this test, the number of degrees of freedom is equal to the total number of points minus 2, because two means are estimated.

$$df = (N_A + N_B) - 2$$

Finally, set up a threshold for significance ($p < 0.05$, e.g.), and look up the critical value t_{crit} in Table 10 (see the section above on one sample t -test for the difference between one- and two-tailed t -tests). If the calculated t

Table 11. Systolic Blood Pressure in Hgmm Measured in 10 Patients Using Either a New Electronic Device or an Old Manual Device

Patient	1	2	3	4	5	6	7	8	9	10
Electronic BP device	121	130	129	113	145	132	110	116	125	155
Manual BP device	115	131	127	111	140	131	111	111	121	150

Table 12. Heart Rate in Beats per Second of Control and Test Patients Suffering from Heart Failure

HF patients	78	81	88	76	93	112	83	96		
Control patients	80	71	68	80	95	67	85	69	85	77

value is $> t_{crit}$, there is a statistically significant difference between the two groups (the null hypothesis H_0 is rejected).

For example, to test if patients diagnosed with heart failure have similar (hypothesis H_0) or higher (hypothesis H_1) heart rates than control patients, 15 readings are performed at rest for these two groups of patients A and B of matched age, sex, and ethnicity. Heart rate is reported in beating per minutes in Table 12.

After testing for normality (see for how to test for normality at the beginning of this section), it is ensured that standard deviations for the two data samples are similar ($SD_A = 11.5$ and $SD_B = 9.1$). To calculate the t -value, it is necessary to compute the mean heart rate for each group. For patients suffering from heart failure, $M_A = 88.4$, and for control patients, $M_B = 77.7$ (variances are $V_A = 140.3$ and $V_B = 82.9$). Thus the pooled variance estimate is $V = 108$, the standard error of the mean is 4.93 and the t value is equal to

$$t = \frac{88.4 - 77.7}{4.93} = 2.17$$

At 5% significance level for 16 degrees of freedom (10 heart failure patients plus 8 control patients minus 2), $t_{crit1\%}$ is equal to 2.12. Since $t > 2.12$, the data support the fact that patients with heart failure have higher heart rate than controls (hypothesis H_0 cannot be rejected).

One-Way ANOVA for Unmatched Samples. One-way ANOVA is used to test the hypothesis that two or more samples are drawn from the same distribution of values and have the same mean and variance. Unpaired student t -test is a particular case of one-way ANOVA applied to two data samples. As for t -test, ANOVA test applies to continuous or noncontinuous data that have a distribution that is not significantly different from normal. Sample sizes should be similar (with less than twofold difference) for all sample groups and, if $n < 30$, variances should also be similar (less than twofold difference). If the test is used in other circumstances, the test outcome will lead to erroneous conclusions. The basis of ANOVA is the F (Fisher) variable, which combines the unbiased variance between sample groups ($V_{interGroup}$) and the variance within sample groups ($V_{withinGroup}$).

$$F = \frac{V_{interGroup}}{V_{withinGroup}} \tag{21}$$

For several data samples A, B, C, \dots of the same size, intergroup variance is defined as

$$V_{interGroup} = \frac{N_A(M_A)^2 + N_B(M_B)^2 + N_C(M_C)^2 + \dots - N_T(M_G)^2}{N_G - 1} \tag{22}$$

where $M_A, M_B,$ and M_C are the means of sample A, B, C, \dots and N_A, N_B, N_C, \dots are the number of values in samples A, B, C, \dots M_G is the average of all values from all sample groups and N_G is the number of samples. The within sample group variance is defined as

$$V_{withinGroup} = \frac{(N_A - 1)(SD_A)^2 + (N_B - 1)(SD_B)^2 + (N_C - 1)(SD_C)^2 + \dots}{N_T - N_G} \tag{23}$$

where SD_A, SD_B, SD_C, \dots are the standard deviations of group A, B, C, \dots and N_T represents the total number of observations (for all data sample pooled together). Degrees of freedom for the numerator of F and the denominator of F are defined as

$$df_{numerator} = N_G - 1$$

$$df_{denominator} = N_T - N_G$$

Note that each variance in Eqs. 22 and 23 is divided by the appropriate degrees of freedom to give unbiased estimate of population variance (assuming the null hypothesis H_0 is true). As for other inference tests, the computed F value is tested against critical F values (Table 13) obtained from the tail of null-hypothesis F distribution (Fig. 4, right panel).

For example, a clinician planning to purchase equipment for electroencephalography compares the signal to noise ratio for three sets of electroencephalographic equipment. For each system that has been made available to him, he records 10 new patients performing standard psychophysical tasks and measures the signal to background noise ratio of the encephalographic equipment (Table 14).

After testing for normality, we must ensure that standard deviations are similar (i.e., no twofold differences). Standard deviation for Brand A is equal to $SD_A = 1.11$; Brand B: $SD_B = 0.75$; Brand C: $SD_C = 0.94$. After calculating $V_{intergroup} = 0.44$ and $V_{withinGroup} = 0.89$, F may be calculated using Eq. 21

$$F = \frac{0.44}{0.89} = 0.49$$

The degrees of freedom for the numerator is $df_{numerator} = N_G - 1 = 2$. The degrees of freedom for the denominator is $df_{denominator} = 30 - 3 = 27$. Reading $F_{crit} = 2.95$ in Table 13, we may conclude that there is no significant difference (since $F < 2.95$) in terms of signal/noise ratio between the three sets of EEG equipments (accept hypothesis H_0).

One-Way ANOVA for Matched Samples. One-way ANOVA may also be used to compare paired sample groups. In fact, since for matched samples, one may

Table 13. The F Distribution of Critical Values at $p = 0.05$ for ANOVA tests^a

$df_2 \backslash df_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	30	40	60	100	∞
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.62	8.59	8.57	8.55	8.54
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.50	4.46	4.43	4.41	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.81	3.77	3.74	3.71	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.86	2.83	2.79	2.76	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.70	2.66	2.62	2.59	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.57	2.53	2.49	2.46	2.41
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.47	2.43	2.38	2.35	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.38	2.34	2.30	2.26	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.31	2.27	2.22	2.19	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.25	2.20	2.16	2.12	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.19	2.15	2.11	2.07	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.15	2.10	2.06	2.02	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.11	2.06	2.02	1.98	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.07	2.03	1.98	1.94	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.04	1.99	1.95	1.91	1.84
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	1.98	1.94	1.89	1.85	1.78
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.94	1.89	1.84	1.80	1.73
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.90	1.85	1.80	1.76	1.69
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.87	1.82	1.77	1.73	1.66
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.84	1.79	1.74	1.70	1.62
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	2.04	1.96	1.88	1.79	1.74	1.68	1.63	1.56
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.74	1.69	1.64	1.59	1.51
45	4.06	3.20	2.81	2.58	2.42	2.31	2.22	2.15	2.10	2.05	1.97	1.89	1.81	1.71	1.66	1.60	1.55	1.47
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.87	1.78	1.69	1.63	1.58	1.52	1.44
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.65	1.59	1.53	1.48	1.39
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.89	1.81	1.72	1.62	1.57	1.50	1.45	1.35
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.88	1.79	1.70	1.60	1.54	1.48	1.43	1.33
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.85	1.77	1.68	1.57	1.52	1.45	1.39	1.28
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.80	1.72	1.62	1.52	1.46	1.39	1.32	1.19
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.77	1.69	1.59	1.48	1.42	1.35	1.28	1.12
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.76	1.68	1.58	1.47	1.41	1.33	1.26	1.08
∞	3.84	3.00	2.61	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.46	1.40	1.32	1.25	1.03

^aTo use this table, read the value at the intersection of the numerator's degrees of freedom (df_1) and the denominator's degrees of freedom (df_2). If your calculated F value is larger than the one you read in the table, the test you performed is significant (see text for details).

analyses either the rows or the columns of a table, the formula given here may be used both for rows or columns, and are usually associated with two-way ANOVA. The formula for the F (Fisher) variable is now equal to

$$F = \frac{V_{\text{interGroup}}}{V_{\text{error}}} \tag{24}$$

The variance due to error or chance is defined as

$$V_{\text{error}} = \frac{\sum_{j,k} (x_{jk} - M_j - M_k - M)^2}{(N_R - 1)(N_C - 1)} \tag{25}$$

where x_{jk} are all the elements in the array, M_j are the row means, M_k are the column means, M is the global array mean, N_C is the number of columns, and N_R the number of rows. The degrees of freedom for the numerator and

denominator are now defined as

$$df_{\text{numerator}} = N_R - 1 = N_G - 1$$

$$df_{\text{denominator}} = (N_R - 1)(N_C - 1)$$

Using the same example as shown in Table 14, and now assuming that the data samples are paired (EEG systems were tested with the same patients), the intersubject variance $V_{\text{error}} = 1.04$ can be computed, and

$$F = \frac{0.44}{1.04} = 0.42$$

The degrees of freedom for the numerator is $df_{\text{numerator}} = N_G - 1 = 2$. The degrees of freedom for the denominator is $df_{\text{denominator}} = (N_R - 1)(N_C - 1) = (3 - 1)(10 - 1) = 18$. For a test at 5%; significance, reading

Table 14. Signal/Noise Ratio for 10 Patients and for Three Brands of EEG Systems

Brand A	1.87	3.88	2.68	1.19	0.93	0.38	2.69	1.8	0.39	1.62
Brand B	2.48	1.71	3.05	1.58	1.7	3	0.47	2.11	2.18	2.22
Brand C	2.29	1.49	2.52	1.26	3.71	2.14	2.33	2.79	2.61	0.29

Table 15. Example of Table for a Two-Factor Experiment

	Protocol 1			Protocol 2			Protocol 3		
Brand A	6	8	8	1	0	2	1	4	4
	10	8	2	2	0	1	4	2	2
	10	6	2	1	3	3	3	5	0
Brand B	2	2	10	4	9	8	3	3	6
	6	10	10	5	5	9	5	4	2
	6	2	6	3	7	7	3	5	6
Brand C	6	0	4	5	3	2	6	6	8
	6	4	4	1	1	1	6	0	10
	4	8	8	3	3	2	4	8	6

$F_{crit} = 3.55$ in Table 13, it can be concluded that there is no significant difference (since $F < 3.55$) in terms of signal to noise ratio between the three sets of EEG equipments.

Note that one could argue that instead of using ANOVA analysis, t -tests could be performed between each pair of samples. Although this is possible, the ANOVA test is more sensitive than a series of paired t -tests because it processes all data samples simultaneously.

Two-Way ANOVA for Two-Factor Experiments. This type of test is being used for experiments with two factors or two attributes. In the example above, to test the reliability of the EEG equipment, the clinician might want to perform three experimental protocols and measure the signal to noise ratio in each of these protocols. The two factors are now the three sets of EEG equipment and the three protocols as shown in Table 15.

In each of the cells of Table 15, the clinician recorded nine values. In the case of only one value per cell, the analysis would be similar to the one-way ANOVA (row and column data may be analyzed separately using one-way ANOVAs for matched samples). However, if several values are recorded for each cell (several subjects, e.g.), one must use the repeated measures two-way ANOVA test. This test is especially interesting because it is possible to test for interaction between variables. Hypothesis H_0 would be that there is no significant relationship between brands and type of protocol and Hypothesis H_1 would be that there is indeed such a relationship. Running a repeated measures two-way ANOVA test under any software will return 3 p -values: the first value is for significant differences between rows; the second value is for significant differences between columns; the last p -value is for the interaction between columns and rows. In the case of Table 15, the p -value for the columns (protocol) is 0.0004 indicating a significant difference between protocols. As observed in Table 15, the values for the first protocol are indeed higher than the values for other protocols. The p -value for the different rows (device brand) is not significant ($p = 0.22$). The p -value for the interaction between brand and protocol is 0.0006. In fact, it appears that the device of brand *B* returns higher values for protocol 2 than other brands, and that the device of brand *C* returns higher values for protocol 3 than other brands.

Experimental design and ANOVA in its many variations is perhaps the most important statistical methodology for experimenters, and the literature is immense. Extreme care should be taken when choosing an ANOVA

test. For example, there are different ways to treat multi-factor ANOVAs analytically when the number of observations is unequal among the treatment combinations (called unbalanced designs). A nontechnical discussion is the classic Planning of Experiments by Cox (6). Other general introductions are Refs. (1,7–12).

Regression and Correlation. Regressions and correlations aim at determining relationship between variables. We may wish to determine if there is a significant correlation between independent and dependent variables, the independent variable being set by the experimenter, and the dependent variable being measured. For example, to test the reliability of a device, an experimenter may change the temperature of the room where the device is being tested (independent variable), and see if this change affects measures returned by the tested device (dependent variable). Regression and correlation can also be used to estimate the relationship between two (or more) dependent variables.

The first step in determining the relation between two variables is to plot values of one variable versus values of the other variable. This is usually called a scatterplot (Fig. 5). From the scatterplot it is often possible to visualize a smooth curve that approximates the data. If it is a straight line, then the least-squares regression method may be used. Otherwise, other curve fitting procedures may be used. It is sometimes useful to plot scatterplots of transformed variables (e.g., log transformation of values in first variable versus values of the second variable).

The method of least-squares computes the best linear regression between two variables. Specifically, for two variables X and Y , the data consist of n pairs $(x_1, y_1), \dots, (x_n, y_n)$. For all values of X and Y , we wish to find the parameter a and b such that

$$Y = aX + b \tag{26}$$

Assuming the jittering of points along the straight line is normally distributed, parameters a and b may be obtained using the formula

$$a = \frac{N \sum x_i y_i - (\sum x_i)(\sum y_i)}{N \sum x_i^2 - (\sum x_i)^2} \tag{27}$$

$$b = \frac{1}{N} \left(\sum y_i - a \sum x_i \right) \tag{28}$$

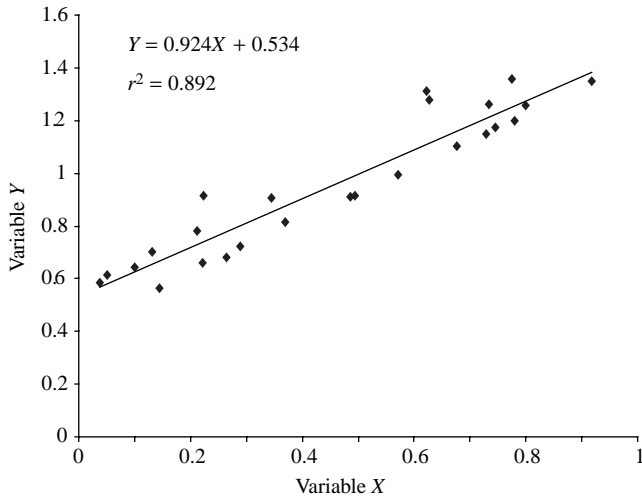


Figure 5. A typical scatterplot with the least-square line drawn through the data points. The r^2 value as well as the best fit equation is indicated on the diagram. The t -value is equal to 9.24 and indicate a significant relationship between X and Y (at $p = 0.05$, for 22 degrees of freedom, $t_{crit} = 2.07$).

To draw the linear regression line, y_i^{est} values may be calculated using Eq. 44 for all values of X . A sample-based measure of the strength of the linear association between the X and Y variables is the sample correlation coefficient (also known as the Pearson correlation coefficient) defined by

$$r = \pm \sqrt{\frac{\text{explained-variation}}{\text{total-variation}}} = \pm \sqrt{\frac{\sum (y_i^{est} - M_Y)^2}{\sum (y_i - M_Y)^2}} \quad (29)$$

r may also be expressed using the original variables X and Y .

$$r = \frac{\text{cov}(X, Y)}{SD_X \cdot SD_Y} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - M_X)(y_i - M_Y)}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - M_X)^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - M_Y)^2}} \quad (30)$$

where M_X and M_Y (SD_X and SD_Y) are the mean (the standard deviation) for X and Y , respectively, and $\text{cov}(X, Y)$ is the covariance between X and Y (the numerator on the right of Eq. 30 is equal to $\text{cov}(X, Y)$) and the denominator is equal to $SD_X \cdot SD_Y$). Necessarily $-1 \leq r \leq 1$. Positive (respectively negative) values of r indicate that large values (respectively small values) of X are associated with large values (respectively, small values) of Y . Values of r near 0 indicate little or no linear association. Interpretation must be done with care because there are many reasons for the presence or absence of a correlation. Also, comparing r values may be misleading as a value of $r = 0.6$ does not mean that the linear relationship is twice as strong as $r = 0.3$. On the other hand, r^2 , called the sample coefficient of determination, represents the proportion of the total variation in the sample values of Y that can be “explained” by a linear relationship as in Eq. 26. Thus $r^2 = (0.3)^2 = 0.09$ versus $r^2 = (0.6)^2 = 0.36$ indicates a 9%

versus 36% accountability for total variability by the proposed linear relationship.

To test if the linear correlation between the two variables is significant, different tests may be used. The null hypothesis H_0 states that there is no relationship between the two variables. A t -test (with degrees of freedom equal to $N - 2$) may be used if the expected population correlation coefficient between variable X and Y is 0 and if we expect the correlation coefficient to be normally distributed when random samples of X and Y are drawn. The variable t is defined as

$$t = \frac{r\sqrt{N - 2}}{\sqrt{1 - r^2}} \quad (31)$$

More details for determining if correlation coefficients are significant or to compare between correlation coefficients may be found in Spiegel and Stephens (2).

As shown in Fig. 5, most regression computer packages will output scatterplots, and correlation coefficients. Residuals plots (not shown here) indicate if the distribution of distance between estimated and actual values of Y . A histogram of these residuals should be normally distributed (computing the parameter a , b , and the coefficient of correlation r requires that these residuals are normally distributed with mean 0 and a constant standard deviation irrespective of the X values).

A comprehensive presentation of regression methods for linear and nonlinear regression is given in Refs. 2, 7, 13.

Nonparametric Testing

Elementary tests mentioned in the previous section require that the distribution of values in the population be normally distributed. In practice, this assumption may not hold so statisticians have devised tests that are less dependent of population distribution. Nonparametric or distribution-free statistical methods generally are not concerned with inferences about parameters of distributions and assume little or no knowledge about the distributions of the underlying populations. Their primary advantage is that they are subjected to less restrictive assumptions than their parametric counterparts. Moreover, the data need not be quantitative (data values may indicate ranks on an ordinal scale). However, a disadvantage of nonparametric methods is that they may not utilize all the information in a sample, consequently requiring a larger sample than the parametric version to attain the same Type II error (see error types).

The χ^2 goodness-of-fit tests previously mentioned is an example of a nonparametric test. Other nonparametric tests make various hypotheses for medians (or means of a symmetric distribution) and differences in location and/or variability of two populations. There are also tests for randomness, independence, and association among random variables. Relatively elementary texts that give a fairly broad and complete coverage of nonparametric methods are Refs. 14 and 15.

Compare Sample Distribution to a Hypothetical Distribution. As for binomial and discrete data, a χ^2 goodness-of-fit test may be performed. For continuous data, a χ^2

Table 16. Heart Rate Variability for 10 Patients While Their Pacemaker Is Switched On or Off, and Calculation of Signed Rank for Wilcoxon Test

Patient	1	2	3	4	5	6	7	8	9	10	Sum
Pacemaker off	0.15	0.32	0.25	1.1	0.82	0.83	0.94	0.42	0.48	0.21	
Pacemaker on	0.12	0.19	0.28	0.56	0.37	0.52	0.24	0.73	0.81	0.13	
difference	0.03	0.13	-0.03	0.54	0.45	0.31	0.70	-0.51	-0.43	0.08	
abs difference	0.03	0.13	0.03	0.54	0.45	0.31	0.70	0.51	0.43	0.08	
Rank of abs difference	1.5	4	1.5	9	7	5	10	8	6	3	
Signed rank	1.5	4	-1.5	9	7	5	10	-9	-6	3	23

goodness-of-fit test may be used on the frequency distribution (histogram) of the data compared to a hypothetical distribution.

Sign Test and Wilcoxon Test for Paired Samples. As for binomial and discrete data, a sign test allows the comparison of paired samples (see the beginning of the section for a definition of paired and unpaired samples). A sign test simply involves pair-wise comparisons of measures between the two sample data sets (see sign test for binomial and discrete data). A variation of this test is called the Wilcoxon test, which takes into account the signed rank of the difference between each pair (instead of using all the signs). This is best illustrated using an example. To test if a pacemaker device has any effect on heart rate variability (defined as the standard deviation of heart beat intervals in seconds), 10 patients' heart rate variability are measured while the pacemaker was either switched on or off (Table 16).

The Wilcoxon test begins by taking the difference in heart rate variability between the two conditions for each patient (forth row of Table 16). If a difference is equal to 0 it is eliminated from further consideration, since it provides no useful information. The second step consists of taking the absolutes of the differences, which is accomplished simply by removing all the positive and negative signs (fifth row of Table 16), then ranking these absolute differences from lowest to highest, with tied ranks included where appropriate. Tied rank means that if two values are equal they are first-ordered randomly and then assigned their average rank (see the first and third columns of the sixth row in Table 16). Finally, reattach to each rank the positive or negative sign that was removed from the difference in the transition from row four to row five, and sum up these values. In our case, $W=23$ and the number of values used in this sum is 10 (degrees of freedom).

If two sets of sample values from the same distribution (which verify hypothesis H_0 that the two samples belong to the same distribution) were to drawn repeatedly and W values were calculated, it would be realized that the distribution (histogram) of W values is close to normal.

In fact,

$$z = \frac{W}{SD_W} \tag{32}$$

may be defined, where z is normally distributed with mean 0 and variance 1, and SD_W is the standard deviation of W , which can be shown to be equal to

$$SD_W = \sqrt{\frac{N(N+1)(2N+1)}{6}} \tag{33}$$

For $N=10$ values, $SD_W=19.6$, so $z=23/19.6=1.17$. As mentioned earlier, the t -distribution is equal to the normal distribution for infinite degrees of freedom. Looking in the last row of the t -table (Table 10), for a significant threshold at $p=0.05$ (two-tailed), $z_{crit}=1.64$ is obtained. Since $z < 1.64$, hypothesis H_0 cannot be rejected. Although it seems that heart rate variability is higher when the pacemaker is switched on, the difference did not reach significance.

Mann-Whiney U Test for Unpaired Samples. The Mann-Whitney U test is similar to the Wilcoxon test. Once more, this test will be illustrated using an example. To compare sensitivity of two hearing aids, the minimum sound a patient can hear using each brand is measured (in dB) and reported in Table 17, where 10 different patients tested each prosthetic device (unpaired samples).

To perform a Mann-Whitney test, first combine all values in an array and assign a rank from 1 to 20 to all these values, assigning tied ranks where appropriate (see Wilcoxon test). The rank for each value is indicated in Table 18.

Then, sum up the ranks for each brand, where $R_A=80$ is the sum for brand A and $R_B=130$ is the sum for brand B. A significant difference between the two rank sums implies a significant difference between the two samples. Calculate the U statistic to test the difference between the ranks:

$$U = N_A N_B + \frac{N_A(N_A+1)}{2} - R_A \tag{34}$$

Note that the formula above is symmetrical with respect to A and B. In the hearing aid example, $N_A=10$ and

Table 17. Patient Maximal Sensitivity (in dB) for Two Brands of Hearing Aids

Brand A	0.1	-1	4.1	2.4	-2.3	3.8	0.9	1.4	0.4	1.2
Brand B	2.7	3.1	5.2	2.1	4.7	1.5	-1.2	3.7	2.8	3.1

Table 18. Rank of Measures for Table 17

Brand A	4	3	18	11	1	17	6	8	5	7
Brand B	12	14.5	20	10	19	9	2	16	13	14.5

$N_B = 10$, so

$$U = 10 * 10 + \frac{10(10 + 1)}{2} - 80 = 75$$

There is no table for U values. Instead, as for the Wilcoxon test, the table for z values is used because of a property of the U distribution. When calculating the U value repeatedly on samples known not to be statistically different (e.g., two data samples drawn from the responses of the same device), then it can be shown that the repeated U values (U_1, U_2, U_3, \dots) have a Gaussian distribution with mean M_U and standard deviation SD_U defined as

$$M_U = \frac{N_A N_B}{2} \tag{35}$$

$$SD_U = \sqrt{\frac{N_A N_B (N_A + N_B + 1)}{12}} \tag{36}$$

This means that the U distribution can be normalized and that

$$z = \frac{U - M_U}{SD_U} \tag{37}$$

is normally distributed with mean 0 and variance 1.

In the example above, $M_U = 10 * 10 / 2 = 50$ and $SD_U = 13.2$, so $z = 3.78$. Looking up the last row of the t -table (Table 10) for a significance level of 5%, we read $z_{crit} = 1.64$. Since $z > 1.64$, hypothesis H_0 can be rejected and it can be concluded that one hearing aid performs better than the other one. Looking at the mean or median for each brand, or for this simple example simply at Table 17, brand A clearly allows patients to hear sounds of smaller amplitudes than brand B. Note that the calculations above are usually not necessary since most statistical software will return the value of U along with its significance level.

Kruskal–Wallis Test for Unmatched Samples. The Kruskal–Wallis H test is a generalization of the Mann–Whitney U test to more than two samples (e.g., three brands A, B, and C of sample sizes N_A, N_B, N_C, \dots with the total number of samples equal to N). As for the Mann–Whitney test, values from all distributions are sorted and once the sum of the rank for each sample is calculated R_A, R_B, R_C, \dots the value of H is given by

$$H = \frac{12}{N(N + 1)} \left(\frac{R_A}{N_A} + \frac{R_B}{N_B} + \frac{R_C}{N_C} + \dots \right) - 3(N + 1) \tag{38}$$

It can be shown that, after collecting repeated measures of H from several samples from the same population (verifying the hypothesis H_0 that they originate from the same population), the histogram of H values is very close to a χ^2 distribution with degrees of freedom equal to the number of groups minus one (so the χ^2 table may be used for H). Thus,

to use the Kruskal–Wallis test, first calculate H , then compute the degrees of freedom (number of groups minus one), and look up the χ^2 critical value in Table 7. If the calculated H value is larger than the critical value, reject hypothesis H_0 .

Friedman Test for Matched Samples. Suppose it is wished to determine if three spectroscopy machines A, B, and C returns the same hematocrit density (density of blood cells in a blood sample). We test the three machines using 20 blood samples (the same blood sample is used for all machines). Since preliminary analysis shows that the readings are not normally distributed, nonparametric test will have to be used. To do so, for each blood sample, rank the machines (from 1 to 3) and compute the total rank for each machine T_A, T_B , and T_C . The parameter T_{all} being the sum of all the ranks, the squares deviate SS is equal to

$$SS = \frac{(T_A)^2 + (T_B)^2 + (T_C)^2}{N_G} - \frac{(T_{all})^2}{N_G N} \tag{39}$$

where N_G is the number of groups and N is the number of samples in each group. As for the Kruskal–Wallis test, we may use the χ^2 distribution with degrees of freedom equal to $df = N_G - 1$. In the Friedman test, simply refer to this value as χ^2

$$\chi^2 = \frac{SS}{N_G(N_G + 1)/12} \tag{40}$$

If the calculated χ^2 value is larger than the critical value for the specified degrees of freedom, reject hypothesis H_0 .

The Spearman’s Rank Correlation Test. Rank methods may also be used to determine the correlation between two variables. Instead of using exact variable values, their ranks may be used. For two sample A and B of the same size, corresponding to two variables X and Y (e.g., lifespans and prices of a family of devices), rank each sample value from 1 to N separately for A and B. Then calculate the difference D_1, D_2, D_3, \dots between the sorted rank for A and B and compute

$$r_S = 1 - \frac{6((D_1)^2 + (D_2)^2 + (D_3)^2 + \dots)}{N(N^2 - 1)} \tag{41}$$

If r_S is close to 0, there is no correlation between the two variables, whereas if it is close to 1 or -1 , there is a strong correlation between the two variables. To test if r_S is significantly different from 0, the same t -test as for the Pearson correlation coefficient may be used (replacing r by r_S and using the same degrees of freedom $df = N - 2$).

Resampling Methods

Resampling methods help provide confidence intervals for parameters in situations where these are difficult or impossible to derive analytically. Resampling methods also help

perform statistical inference without assuming a known probability distribution for the data. The bootstrap method consists of drawing random subsamples and the randomization method consists of shuffling data samples.

Bootstrap Method. The bootstrap method is the most recently developed method to estimate errors and other statistics. It is not primarily aimed at performing inference although it may be used to do so, since it provides confidence intervals for the measure of interest. The term “bootstrap” derives from the phrase “to pull oneself up by one’s bootstrap” (Adventures of Baron Munchausen, by Rudolph Erich Raspe). Suppose we have a data sample and an estimator (e.g., mean). The basic idea involves sampling with replacement to produce random samples of size N from the original data sample (of size larger than N). Each of these samples is known as a bootstrap sample and provides an estimate of the parameter of interest. Repeating the sampling a large number of times provides information on the variability of the estimator and help define confidence limits. There are N to the power of N , N^N , possible samples, called the ideal bootstrap samples. It is important to emphasize that subsamples are drawn with replacement: for example, for an empirical distribution composed of 2 values (5 and 8), the bootstrap samples are (5,8), (5,5), (8,8), and (8,5) (note that there are $2^2 = 4$ of them). Getting all ideal bootstrap samples becomes unrealistic as N becomes larger, so the Monte Carlo approach (which consists of random draws) is used. The sampling is said to be balanced if each sample value is drawn the same number of times. For each bootstrap sample, let us suppose that the mean is calculated. The standard deviation of the bootstrap distribution for the mean correspond to the standard error (Eq. 19) and may be used in parametrical t -test to compute the t value (Eq. 20), and perform inference testing (assuming normality of the distribution of course). However, this mixture of bootstrap and parametric t -test is relatively unconventional, and it is better to estimate the bootstrap distribution of t -values as explained below.

To perform a statistical inference test using bootstrap, first state a null hypothesis H_0 . Null hypotheses for resampling tests are usually vague because there may be many reasons (based on the shape of the distribution) why two samples may differ (whereas when performing a parametric t -test, the nonnull hypothesis states clearly that the means are nonequal). Moreover, bootstrap statistics use the implicit assumption that data samples are representative of the underlying population and in fact do as if the data samples were the population itself. Therefore it is not possible to draw direct conclusions about the underlying population either.

In the case of the heart rate study of Table 12, for example, where comparing a measure (i.e., heart rate) for patients suffering from heart failure (sample A) and control subjects (sample B), the null hypothesis would be “patient suffering from heart failure have abnormal heart rate”. One way to test this hypothesis is to perform a bootstrap t -test. Two bootstrap samples are first drawn from the pooled distribution of A and B: sample A' and B' of the same size as A and B, respectively. The t -value is then

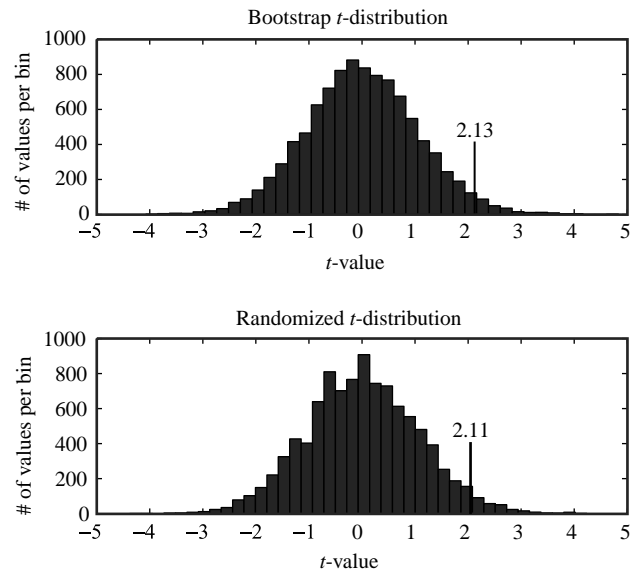


Figure 6. Bootstrap t -distribution for Table 12 (top) and randomized t -distribution for Table 12 (bottom). Since the actual t -value obtained from the original samples in Table 12 ($t = 2.17$) belong to the rightmost 2.5% value in both the bootstrap and the randomized distribution (the 2.5% limit being indicated by a vertical line), it may be considered significant at 5%.

computed using the two bootstrap samples as in Eq. 16. The operation is repeated m times to obtain the distribution of t -values for the null hypothesis. Note that, even if a t -value, is computed it is not assumed normality for the data samples since the distribution of t -values for the null hypothesis is estimated using bootstrap samples. The actual t -value is calculated for the original data samples A and B and tested against the bootstrap t -distribution. If it lies in the lower 2.5% or upper 2.5% tails, then the bootstrap test may be considered to be significant at the 5% level of significance. In Fig. 6 (top), 10,000 bootstrap t -values were accumulated for the two samples in Table 12. Since the original t -value for Table 12 is equal to 2.17 (see the t -test section) and since it lies in the upper 2.5% of the bootstrap t -value distribution, it may be concluded that the data support the hypothesis that heart rate is affected in patients suffering from heart failure at the 5% significance level.

There are other ways to test for significance using bootstrap, such as the bootstrap-percentile method, or the bootstrap-bca method (see Ref. 16 for a comprehensive reference). In general, it should be remembered that bootstrap methods are designed primarily for estimating characteristics of data samples, not for performing inference tests. Resampling methods specializing in statistical inference are called randomization methods and are describe below.

Randomization Methods. For the purpose of performing paired or unpaired comparisons, randomization methods consist of random permutations of data. Randomization methods are also often called permutation methods or surrogate methods. Specifying the null H_0 hypothesis is

the same as for the bootstrap and involves a vague formulation about the result of the experiment, such as “patient suffering from heart failure have abnormal heart rate” or “the drug treatment does not have an effect on blood pressure”.

Randomizing the data is straightforward. Using the same example as for the bootstrap distribution with two unpaired samples A and B of sizes N_A and N_B , a randomization method consists of pooling the data of A and B together (into C), then randomly drawing from C (without replacement) two groups A' and B' that have the same size as A and B, respectively (17). Then, compute the estimator (e.g., t -value) for each randomized pair of samples. Repeat this procedure a large number of times to obtain the distribution of the estimator (e.g., t -value) for the null hypothesis. Significance is assessed as for the bootstrap t -test. For example, in Fig. 6 (bottom), 10,000 randomized t -values have been accumulated for the two samples in Table 12. Note that irregularities in the distribution are due to the fact that we are randomizing a relatively small number of values. As for the bootstrap, since the original t -value ($t = 2.17$) lies in the upper 2.5% of the randomized t -values, it may be concluded that the data support the hypothesis that heart rate is affected in patients suffering from heart failure at the 5% significance level. It is reassuring to notice that the upper 5% significance threshold t -value for the bootstrap ($t_{crit} = 2.13$), the randomized ($t_{crit} = 2.11$), and the normal distribution ($t_{crit} = 2.12$) are all similar.

For paired comparisons, the principle is slightly different since we are now randomizing not the sample values but the pairs. For example, for the data of Table 11, one-half of the pairs are selected randomly then shuffled (the value for the first device is now attributed to the second device and vice versa) and the paired t -test value is recalculated (Eq. 17). This procedure is repeated many times. To assess significance, as in the previous paragraph, the original t -value computed using the nonrandomized samples is compared against the distribution of randomized t -value.

This procedure may be generalized to compare an arbitrary number of samples. For example, to compare several unpaired sample, data sample values may be randomized among groups and one-way ANOVA values may be calculated repeatedly. The ANOVA value for the nonrandomized groups is then compared against this ANOVA randomized distribution. Web Ref. 18 provides a clear introduction to resampling methods.

MULTIVARIATE METHODS

Previously the probability distributions involving one variable were discussed, but in many situations there are two or perhaps many interdependent variables, for example, height, weight, daily caloric intake, genetic strain. Data samples involving several variables are called multivariate. Many multivariate analytical methods involve inference for the parameters (means, variances, and correlation coefficients) based on multivariate normal distribution. One such method is known as discriminant analysis and is concerned with the problem of distinguishing between two or more populations on the basis of observations of a multivariate nature. Principal components analysis, cluster, and factor analysis seek to determine relatively few out of possibly many variables that will serve to explain the variability or the interrelationships in the variables. Principal component analysis (PCA) would specifically make each successive component account for as much as possible of the remaining variability uncorrelated with previously determined components. In Fig. 7, data points from two variables are represented. Coordinates of data points on the abscissa axis correspond to values of the first variable and coordinates on the ordinate axis correspond to values of the second variable. The PCA is able to find a first principal axis (labeled one) that accounts for most of the variance of the data. The second principal axis (labeled two) has to be perpendicular to the first principal axis and accounts for the remaining of the variance.

Recent progresses in signal processing and information theory have seen the development of blind source

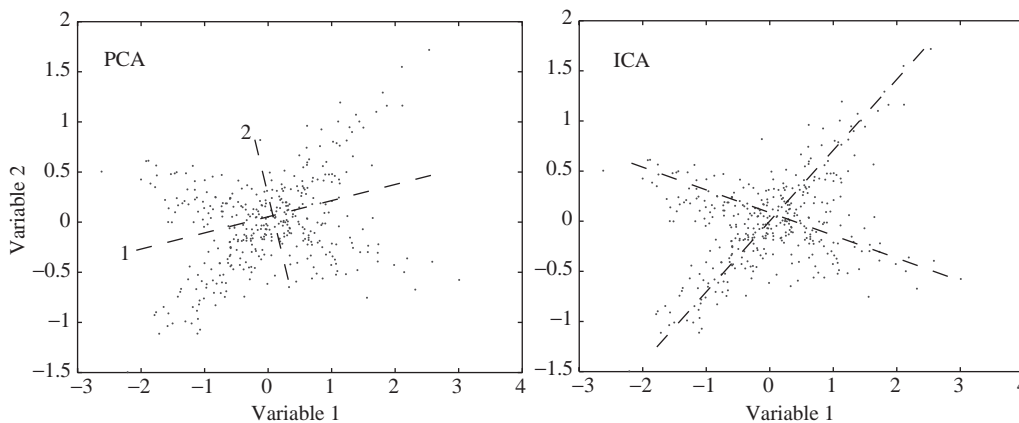


Figure 7. Illustration of PCA and ICA algorithms. The PCA finds axis with maximum variance. By contrast in ICA, the projection of data point on ICA axis is maximally independent.

separation methods, which attempt to find a coordinate frame onto which the data projections have minimal overlap. For example, if two sources of sounds (e.g., a conversation and a CD player) are recorded simultaneously in the same room on two microphones, the sound signal from the two sources are mixed on both microphones. Coordinates of data points in Fig. 7 could represent the signal recorded from the two microphones. Separating the two sound sources from the microphone signal is called blind source separation. Independent component analysis (ICA) is a family of linear blind source separation methods. The core mathematical concept of ICA is to minimize the mutual information among the data projections. PCA components are orthogonal as shown in Fig. 7, which is usually not a realistic assumption for biophysical data. To find biologically plausible sources, PCA must be followed by an axis rotation procedure, and ICA can be viewed as a powerful rotation method. The ICA seeks to find axes for which the projection of data is maximally nonnormal (i.e., contains the maximum amount of information). It uses the property of the central limit theorem in statistics, stating that any linear mixture of two or more source activities is more normal than the original source activities, so, by finding axes that maximize nonnormality, source separation may be achieved. As can be seen in Fig. 7, ICA is free to adapt to the actual projection patterns of source generators, if their activity time courses are (near) independent of one another. Performing ICA decompositions is most appropriate when sources are linearly mixed in the recorded signal, without differential time delays.

The ICA is being applied to various biomedical signal processing problems that include performing speech and noise separation (19), decomposing functional resonance imaging data (20), and separating brain area activities and artifacts mixed in electroencephalography scalp sensors (21).

Texts that give broad coverage of multivariate analysis are (22–24).

CLINICAL TRIALS

A clinical trial is not a method *per se*, but is a term applied to any form of planned experiments that involves human patients. The purpose of a clinical trial is to evaluate and verify the efficacy and safety of a new treatment or sets of treatments for a given medical condition. Although most of the analytical methods employed for clinical trials are the same as in other contexts, there is a special effort to avoid bias, which leads to some unique designs. Another distinguishing characteristic of clinical trials is the constraint imposed by studying living patients and the often difficult ethical considerations that must be addressed.

Double Blind. The usual method to avoid bias in experimental designs is the random allocation of experimental subjects to treatments, but this will generally not suffice in clinical trials. A major potential source of bias is when subjects or evaluators in a trial know which treatment (e.g., placebo or active) is being received. In double-blind trials, neither the subject nor the evaluators are

aware of which treatment is being received. Sometimes ethical or practical considerations make double-blinding infeasible, and sometimes partial blinding, for example, independent blinded evaluators only, may be sufficient to reduce bias in treatment comparison.

Within Patient Studies versus Across Patient Studies. Most clinical trials are conducted as parallel studies in which two or more treatments are evaluated concurrently in separate groups of patients. As many researchers remain reluctant to assign patients randomly to new or standard treatments, current patients on the new treatment may be compared with data external to the study containing patients who had received standard treatments. Such an approach invites severe bias, since there is no assurance that treatment and control groups do not differ with respect to some factors other than the treatment itself. In crossover studies, each patient receives in succession two or more treatments. When feasible, such within-patient studies require smaller sample sizes than between-patient studies to achieve the same level of significance.

Lifetime Variables. Some clinical studies are conducted as life data analysis and survival studies, and require specific statistical tools. In such studies, a variable represents the time to the occurrence of some event of interest, and is called a lifetime variable. In the engineering context, a life test consists of monitoring the operation of a sample of devices and to observe causes of and times to failure for all or some of the devices. In the clinical context, a survival study may involve observing cause of death (and time from entry to the study until death occurs) for some potentially fatal or, in the case of animal studies, induced disease. Alternatively, the event of interest may be time to relapse or time to remission for some diseases or conditions. The purpose of life tests or survival studies is to estimate or to compare lifetime or survival between different treatment groups.

Statistical Test for Lifetime Variables. Since a lifetime variable must be positive (number of remissions, e.g.), the normal distribution is not usually a suitable model. The normal-based methods of multiple regression and analysis of variance cannot be used in the usual manner and in general requisite mathematical and computational methods are much less tractable than normal-based methods. Consequently, a nonparametric, partially parametric, or nonnormal distributional analytic approach is taken. Data is usually visualized using Kaplan–Meier survival curves where censored patients (patients that have left the study) are explicitly indicated on the curve. Comparing between unpaired groups usually involve a log-rank test or a Mantel–Haenszel test. Conditional proportional hazards regression may be used to compare between two or more paired groups. Finally, Cox proportional hazard regression may be used to compare between more than two unpaired groups and perform regression analysis.

Censoring. As mentioned above, a further complicating factor for survival studies is censoring. Under censoring,

exact lifetimes are known only for a portion of the experimental units, the remainder known only to exceed certain censoring times. Censoring is usually a practical necessity and must be preplanned. For example, a life test on a random sample of 100 devices that has median time-to-failure of 2500 h will likely take over a year to complete if the tests were to continue until all devices fail. Instead, the test might be terminated at some predetermined time (e.g., 1000 h), or immediately upon achieving some predetermined number of failures (e.g., 30). These are called Type I and Type II censoring, respectively, and are the simplest to deal with. A distinguishing characteristic of survival studies involving human patients is that censoring times are often random. For example, suppose patients with a certain cancer are undergoing different chemotherapy treatments. Patients may enter the study in a random manner and patients may survive the termination time of the study or may die due to causes unrelated to the cancer. There are probability models that incorporate these data and lead to appropriate statistical inferential techniques. For example, some techniques assess the effectiveness of different treatments by comparing estimated mean survival times with the effect of unrelated causes of death removed. Other methods used for dealing with censoring will not be discussed. It is sufficient to say that the special problems of statistical inference in the presence of censoring necessitate the use of large sample approximations and computer-aided numerical solutions. Some of these methods incorporate strong assumptions that users should be aware of.

Extensive treatment of methods for censoring and the analysis of survival data is given in Refs. 25–27. Nontechnical discussions of clinical trials and the special statistical treatments they require are given by Pocock (28) and Shapiro (29).

STATISTICAL COMPUTING AND SOFTWARE

Standardized computer programs aiming at performing a variety of statistical analyses were developed through the 1960s at several universities and became widely available in the 1970s. There is now a large number of them and the one to use will depend on the users' expertise in statistics and field of research. For infrequent usage on small data samples and testing of simple hypothesis (χ^2 , t -test, ANOVA), MS Excel, which is usually already installed on many computer desktops, may be sufficient. Note the availability of extra statistical functions when one selects the Analysis Toolpack add-in (installed but inactive by default). However, MS Excel is not a statistical software *per se*, so to go beyond exploratory analysis stages it is better to rely on professional statistical software.

The best known and most comprehensive of these, now all under privately managed companies, are the Statistical Package for the Social Sciences (www.spss.com), the Statistical Analysis System (www.sas.com), and JMP (www.jmp.com). As its name suggests, SPSS, was developed primarily for use by social scientists and is relatively easy to learn by individuals with limited statistical and computer backgrounds. The SPSS graphical interface is

organized as tabular spreadsheets similar to MS Excel. The programs comprising SPSS, their output format, and the examples in the manuals retain a social science flavor. The SAS has evolved into a widely utilized and extremely flexible package that is generally regarded to be more statistically sophisticated and complete than SPSS. JMP, also developed by the SAS institute, is a user-friendly graphical interface that sequentially guides the user through all stages of the experimental design and data analysis.

Apart from the graphical packages mentioned above, most other statistical softwares rely on command line calls, where users call functions from a prompt (note that most of these softwares also include menus). The free R software (www.r-project.org) offers powerful functions contributed by leading statisticians in the world. Because it is an open source project, it is used by many scientists and its extensive libraries are probably the place to look for rare statistical procedures. The Biomedical Programs (BMDP) (www.statsol.ie) contains a large variety of elementary and advanced statistical procedures. The programs are widely applicable, but some are particularly appropriate in biomedical contexts, such as repeated measures ANOVA designs (see ANOVA). The S-plus software is also very popular (www.insightful.com) and very similar to R. It is based on the S language developed at AT-T. Finally, a widely used package in academia, as well as in industry is a package called MINITAB (www.minitab.com), which is one of the most user-friendly command line software.

There are many smaller, less comprehensive statistical analyses packages available for computers. These range from packages that perform elementary, mostly descriptive analyses, to some that are rather sophisticated. For bootstrap and surrogate statistics, SAS software is preferred among graphical software, although it is possible to program bootstrap and surrogate data routines in SPSS. The R software contains the majority of such user-contributed routines and S-Plus also contains a few of them. Finally, MATLAB (www.mathworks.com), an interpreted language widely used in engineering, also has a large number of user-contributed bootstrap and surrogate statistics routines available.

Caution against the ignorant use of computerized statistical analyses cannot be overemphasized. In planning studies, the methods of analysis and the constraint they impose on experimental designs should be taken into consideration in advance. If not, much work and data collection efforts could be wasted. Worse still, misleading and even meaningless results are often given undeserved weight merely because they represent the voluminous output of computer programs. How often do we hear that “a computer analysis shows...”, but such programs can be totally inappropriate. For example, the mathematical methods underlying repeated-measures ANOVA incorporate restrictive assumptions on the normality of the data and the experimental design for appropriate randomization of events. Although these considerations are often ignored, researchers should systematically assess the degree to which test-related assumptions are satisfied. These facts notwithstanding, computer-aided data

management and analysis can be of great benefit if used properly and wisely.

REFERENCES USED

This list is not meant to be comprehensive. For the naïve reader, a basic introduction to statistics with a plethora of exercises is given in the Schaum's outline series on statistics (2). For the nonnaïve reader in statistics, a more technical yet still accessible reference is Ref. 30. Other texts dealing with general statistical methods, particularly regression and analysis of variance are Refs. 31 and 32. Comprehensive web references are Refs. 18, 33, and 34.

Statistical books have also been written for specific research topics. For example, see Ref. 35 for a beginner's reference in designing biology experiments and Refs. 6, 8–10 for more detailed references. As already mentioned, see Refs. 28, 29, 36, and 37 for clinical trials. Finally, a recent development in statistics is statistical process control that deals with optimizing production and quality in the industry (38).

ACKNOWLEDGMENTS

Parts of this article were adapted from a previous version of this encyclopedia paper by P. Sullo and L.E. Ostrander from the Rensselaer Polytechnic Institute. The author also wishes to thank anonymous reviewer for their invaluable comments.

BIBLIOGRAPHY

1. Gill JL, Design and Analysis of Experiments in the Animal and Medical Sciences. Ames (IA): Iowa State University Press; 1978.
2. Spiegel MR, Stephens LJ, Schaum's Outlines in Statistics. 3 ed. New York: McGraw Hill; 1999.
3. Bonferroni CE. Sulle medie multiple di potenze, Bollettino dell'Unione Matematica Italiana, 5 third series, 1950. p 267–270.
4. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat* 1979;6, 65–70.
5. Hoppe F. Multiple Comparisons, Selection, and Applications in Biometry: A Festschrift in Honor of Charles W. Dunnett. New York: Marcel Dekker; 1992.
6. Cox DR. Planning of Experiments. New York: John Wiley & Sons; 1992.
7. Norman R, Draper NR, Smith H. Applied Regression Analysis. 2nd ed. New York: John Wiley & Sons; 1998.
8. Lorenzen TJ, Anderson VL. Design of Experiments: A No-Name Approach. New York: Dekker; 1993.
9. Box GEP, Hunter WG, Hunter JS. Statistics for Experimenters. New York: John Wiley & Sons; 1978.
10. Cochran WG, Cox GM. Experimental Designs. New York: John Wiley & Sons; 1992.
11. Hicks CR, Turner KV. Fundamental Concepts in the Design of Experiments. New York: Oxford University Press; 1999.
12. Winer BJ, Brown DR, Michels KM. Statistical Principles in Experimental Design. 3rd ed. New York: McGraw-Hill; 1991.
13. Neter J, Wasserman W, Applied Linear Statistical Models. 4th ed. New York: McGraw-Hill/Irwin; 1996.
14. Conover WJ. Practical Nonparametric Statistics Methods. 3rd ed. New York: John Wiley & Sons; 1998.
15. Hollander M, Wolfe DA. Nonparametric Statistical Methods. 3 ed. New York: John Wiley & Sons; 1999.
16. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York: Chapman and Hall; 1994.
17. Blair R, Karniski W. An alternative method for significance testing of waveform difference potentials. *Psychophysiology*, 1993. p 518–524.
18. Howell DC. Resampling Statistics: Randomization and the Bootstrap. 2005. Available at <http://www.uvm.edu/~dhowell/StatPages/Resampling/Resampling.html>.
19. Park H -M, Jung H -Y, Lee T -W, Lee S -Y. On subband-based blind signal separation for noisy speech recognition. *Elect Lett* 1999;35:2011–2012.
20. Duann JR, Jung TP, Makeig S, Sejnowski TJ. fMRLAB: An ICA Toolbox for fMRI Data Analysis. Presented at Human Brain Mapping, Sendai, Japan, 2002.
21. Delorme A, Makeig S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 2004;134:9–21.
22. Anderson TW. An Introduction to Multivariate Statistical Analysis. New York: John Wiley & Sons; 2003.
23. Gnanadesikan R. Methods for Statistical Data Analysis of Multivariate Observations. New York: John Wiley & Sons; 1997.
24. Stone J. Independent Component Analysis : A Tutorial Introduction. Cambridge (MA): MIT Press; 2004.
25. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data. 2nd ed. New York: John Wiley & Sons; 2002.
26. Lawless JF. Statistical Models and Methods for Lifetime Data. New York: John Wiley & Sons; 2002.
27. Miller R. Survival Analysis. New York: John Wiley & Sons; 1998.
28. Pocock SJ. Clinical Trials: A Practical Approach. New York: John Wiley & Sons; 1984.
29. Shapiro SH. Clinical Trials. New York: Dekker; 2004.
30. Hays W, Statistics. 5th ed. New York: Wadsworth Publishing; 1994.
31. Afifi A, Clark VA, May S. Computer-Aided Multivariate Analysis. 4th ed. New York: Chapman & Hall/CRC; 2004.
32. Snedecor GW, Cochran WG, Statistical Methods. 8 ed. Ames (IA): Iowa State University Press; 1989.
33. Lowry R, Concepts and Applications of Inferential Statistics. 1999. Available at <http://faculty.vassar.edu/lowry/webtext.html>.
34. Wasson J. Statistics in Educational Research—An Internet Based Course. Available at <http://www.mnstate.edu/wasson/ed602.htm>.
35. Cann A. Maths from Scratch for Biologists. John Wiley & Sons; 2002.
36. Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. *Science* 1977;198:679.
37. Chalmers TC, Celano P, Sacks HS, Jr. JS. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983;309:1358.
38. Amsden R, Butler H, Amsden D. SPC Simplified: Practical Steps to Quality. Quality Resources, 1998.

See also BIOINFORMATICS; COMPUTER-ASSISTED DETECTION AND DIAGNOSIS; QUALITY-OF-LIFE MEASURES, CLINICAL SIGNIFICANCE OF.

STANDARDS FOR MEDICAL DEVICES. See CODES AND REGULATIONS: MEDICAL DEVICES.

STEREOTACTIC RADIOSURGERY. See RADIOSURGERY, STEREOTACTIC.

STEREOTACTIC SURGERY

ANTONIO A.F. DE SALLES
ALESSANDRA GORGULHO
UCLA Medical School
Los Angeles, California

INTRODUCTION

Stereotactic surgery evolved from the need of neuroscientists and neurosurgeons to approach areas deep in the brain with minimal disruption of its structure. In (~1000 AD, the Incas already had some idea of which regions of the brain could be operated on without causing detectable functional deficits (1). They concentrated their trepanations in the right frontal area knowing that lesions in this region of the brain would be safe (2). The scientific literature, however, registers the first minimally invasive attempt to approach the noneloquent areas of the brain in the late 1800s. Zernov, a Russian anatomist, described the first device used to localize the sensory-motor areas of the brain (3,4) This device allowed the surgeon to use noneloquent areas as approaches to targets in the depth of the brain. He designed a frame that was attached to the patient's head and supported a hemisphere with a drawing of the brain gyri. This drawing guided the surgeon where to perform the craniotomy avoiding eloquent areas (4). This device was crude, however, and needed to be replaced by a precision instrument capable of being applied to patients' individual anatomy and not on a generic drawing.

Cartesian coordinates, developed by Renee Descartes (5), were called upon to guide the neurosurgeons in their endeavor. Initially developed for use in laboratory animals, the first stereotactic frame based on Cartesian spatial localization was designed by an electrophysiologist and a neurosurgeon. The Cartesian system provides precise localization of a point in space by the distance that the point is located from each of the three planes. Clarke and Horsley applied a stereotactic frame in animals to guide electrodes into the depth of the cerebellum to study neuronal function (6,7) (Fig. 1). This apparatus inspired the Canadian bioengineer Aubrey Mussen to develop the first true stereotactic device for humans (8,9). This device was never applied to a patient, since Mussen could not convince any neurosurgeon in Canada to use it. It is believed to have been created 1918, as it was found several years later wrapped in a newspaper of that year. It is currently in exhibit at the Montreal Neurological Institute (10).

All early localization of brain structures was based on cranial landmarks. Unfortunately, these bony reference points frequently failed to guide the surgeon to the precise area of the brain to be approached. Only when angiography (11) and ventriculography (12,13) became available, stereotactic surgery based on intracranial reference points

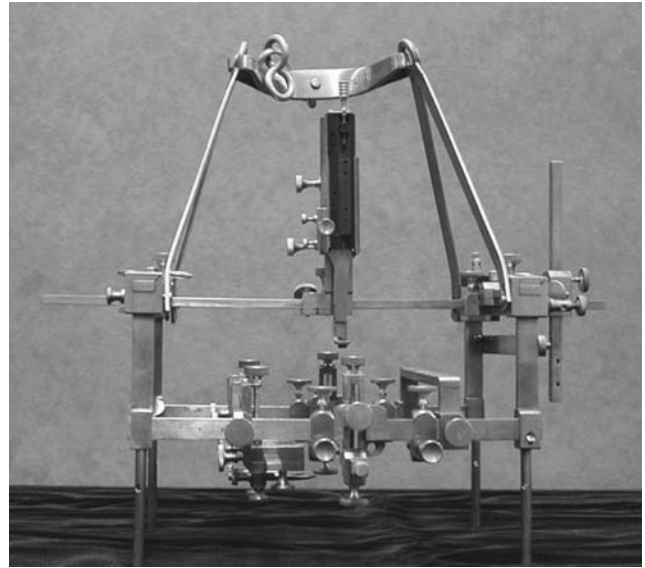


Figure 1. A. Horsley–Clark stereotactic device: Horsley, a neurosurgeon and Clark, a physiologist, developed the first stereotactic apparatus. It was used in experimental studies in animals.

became reliable. It took another 17 years for the first human stereotactic surgery based on Cartesian coordinates to be performed. Spiegel and Wycis performed the first stereotactic surgery in 1947, in Philadelphia (14). It was a thalamotomy to treat psychiatric disease (15).

Functional neurosurgery, mainly movement disorders, was the major field for stereotactic surgery at the beginning. During the 1930s and 1940s, multiple central nervous system procedures involving the motor cortex to the cervical spine were used in an attempt to relieve parkinsonian symptoms. Despite tremor improvement, none of the procedures resulted in rigidity and akinesia amelioration, and usually resulted in paralysis of the affected limb. In 1940, Meyers published his patients series in whom either tremor and rigidity were improved (16–18). Meyers performed a craniotomy to section the pallidofugal fibers as they emerged the medial Globus Pallidus. In 1947, Spiegel and Wycis developed the stereotactic apparatus, therefore avoiding a craniotomy. It was with Irving Cooper, in 1952, that the surgical treatment of Parkinson's Disease gained popularity by the inadvertent lesion of the anterior choroid artery during a pyramidotomy (19). Cooper et al. provoked a pallido-thalamic infarction due to the lesion and observed an effective relief of the parkinsonian tremor. Thereafter, more precise targets were defined (16–24) for the treatment of Parkinson's disease (PD), essential tremor, dystonia and others. Initial results were encouraging, but recurrence and morbidity, specially in bilateral procedures, were not irrelevant. After the introduction of levodopa therapy in 1968, the indication of surgical procedures declined. Nevertheless, levodopa therapy proved not to be as effective after a medium period between 5 and 10 years (22,25). Over this period of time, tolerance to levodopa occurs and a dose increase is necessary to maintain the prior therapeutic effects. The extra amount of medication leads to a complication known as dyskinesia. The patient

presents involuntary movements that may be mild or severely disabling during the peak period of the dose. Since medication was highly effective only for a period of time, attention was directed once more toward surgical treatment. In the 1980s, stereotactic procedures for movement disorders again became routine (26,27).

The applications of the stereotactic techniques have grown since that time (28–30). New devices were developed to facilitate surgery and increase precision. Frames are still widely used in clinical practice; however, guidance techniques independent of skull fixation are improving and may completely replace the use of the stereotactic frame (31–35). This article discusses the evolution of the stereotactic instrumentation during the last century and describes the directions of the stereotactic technique at the beginning of the millennium.

PRINCIPLES OF STEREOTACTIC SURGERY

The stereotactic frame establishes the stereotactic space. It is described mathematically as a cube with X , Y , and Z coordinates corresponding to lateral, anteroposterior (AP) and vertical measurements, respectively. When the head is placed inside the stereotactic device, precise coordinates can be assigned to any location within the brain. Initially, orthogonal approaches were used, which were adequate because only orthogonal images were available, as represented by plain X rays. The AP projection offered the X and Z values and the lateral projection offered the Y , but also Z values (10).

The early stereotactic devices allowed only orthogonal approaches to the brain in relation to the applied stereotactic frame. This is a straightforward method in which the probe is perpendicular to a square base fixed in the skull. Other mathematical approaches were used over the years: the burr hole mounted system (Ward and McKinney), the interlocking arcs system (Brown–Robbers–Wells, BRW), the phantom-based system (Riechert–Munding), and the arc centered system (Leksell). In the burr hole mounted system, the depth of the probe is dependent on the angle of the burr hole mounted apparatus. A minor variance of angulation can lead to a major error in the deep as well as anteroposterior and lateral position of the probe. Therefore, this system is not used nowadays. The interlocking arcs system requires the adjustment of individual arcs to define the trajectory making calculations very complex. Despite that, the first computer-based computed tomography (CT) stereotactic was performed with the interlocking arcs (BRW) system. The arc centered system, the most currently used nowadays, was developed by Lars Leksell in 1949 (36). The trajectory of the probe is perpendicular to the arc (vertical axis) and the quadrant (horizontal axis). A spatial spherical shape is defined by the arc-quadrant. When the probe reaches maximum depth, it will always be in the center or at a focal point of that sphere, independent of the entry point. The arc centered became the instrument of choice since the end of the twenty century (37).

Historically, neurosurgeons have been dependent on neuroimaging to perform their craft. Before that, neuro-

surgery was dependent on symptomatic clues to approach the brain. Only *postmortem* correlation studies served as references for the surgeon to operate in the living brain. Ventriculography was the first great step in the development of stereotactic surgery (12,13). It provided the indispensable brain landmarks for the neurosurgeon to start electrophysiological brain mapping. Functional and anatomic atlases of the brain were developed to guide the stereotactic operations. To date, atlases developed during the 1950s and 1960s are still used in stereotactic surgery (38). Initially, the foramen of Monroe and the pineal gland were the landmarks used for internal guidance of the stereotactic surgeon. These landmarks were well seen on air ventriculography, which was the first contrast material medium used in neurosurgery to delineate the internal structures in a plain X-ray film. Soon after, the positive iodine contrast became available for neurosurgery. The positive contrast injected into the ventricular system provided exquisite delineation of the anatomy of the third ventricle. The anterior commissure (AC) and posterior commissure (PC) could be promptly defined. They became the landmarks of choice for stereotactic surgeons. The main atlases of the brain were developed based on these two landmarks (38,39) (Fig. 2).

The Cartesian planes were largely based on the initial intercommissure plane. This imaginary plane hinged on the AC and PC line being parallel to the skull base. Two other planes perpendicular to this plan and to each other composed the necessary three planes to determine the Cartesian system. The coronal plane passes through the midcommissural point and is perpendicular to the midsagittal plan. The mathematical challenge during stereotactic surgery consists of transforming the numbers generated

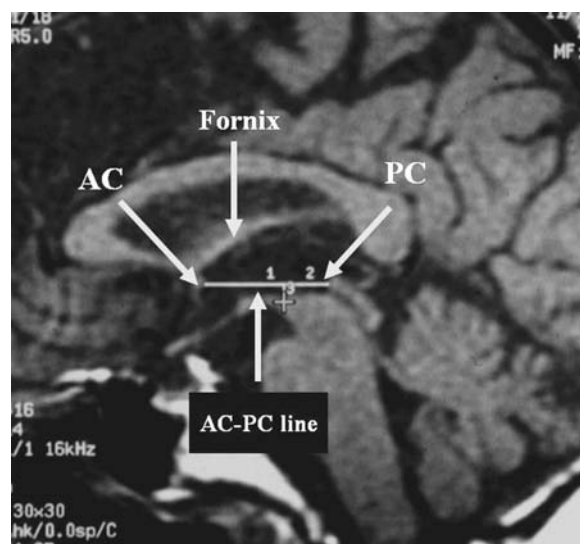


Figure 2. Sagittal T1 weighted magnetic resonance imaging (MRI) scan showing the AC, which is defined at the anterior wall of the third ventricle. The PC is placed at the posterior wall of the third ventricle, directly above the beginning of the Aqueduct. The AC–PC line is of the utmost importance in functional neurosurgery. The coordinates established in the major stereotactic atlases were defined using this line as reference.

in the internal Cartesian system of the brain to one of the stereotactic devices. This simple mathematical transformation was done in the operating room by stereotactic surgeons, now this is automatically generated by computer software.

The indirect visualization of the brain by shadows of the ventricular system obtained with plain X rays of the skull was sufficient for functional localization of sites in the brain. These indirect visualizations and calculations allowed stereotactic surgeons to accumulate a wealth of knowledge of the brain's electrophysiology. Standardized atlases allowed a degree of stereotactic accuracy during functional procedures. Pathological anatomy secondary to space occupying lesions did not require the use of standardized atlases. This knowledge was limited to directing surgeons in the treatment of brain tumors and other morphologic diseases of the brain. Visualization of brain lesions while inside of the stereotactic frame was necessary for reaching these lesions for biopsy and possible therapy. This soon became possible with the angiography since indirect targeting based on deformation of the vascular anatomy made the localization reliable. However, due to the vascular nature of the lesions visualized, biopsy would be too risky. Stereotactic biopsy only flourished with the advent of CT and direct soft tissue visualization. During the angiography period, arteriovenous malformations (AVMs) were treated with radiation and tumors were treated with implants of isotopes guided by stereotactic surgery.

STEREOTAXIS BASED ON CT

Targeting based on tomographic images resulted in a remarkable increase of the use of stereotactic techniques in many centers. It also expanded its applicability to different fields. Focused radiotherapy became possible because CT allowed reliable targeting when compared to early angiography and pneumoencephalography–ventriculography generated targets.

The *X* and *Y* coordinates were acquired from one axial CT slice. Determining the *Z* coordinate by CT was a challenge. The initial idea was to correlate the vertical displacement of the CT table from a reference point to the target slice. This parameter was not reliable and a new form of calculating vertical displacement was elaborated. The fiducial markers were made possible by obtaining an accurate three-dimensional (3D) target (37). The fiducials (Fig. 3) are placed perpendicular to the image acquisition plane, either axial, sagittal, and coronal. The fiduciary system is composed of vertical outer bars and a diagonal internal one. According to the slice obtained, the distance between the diagonal line to the vertical one is going to be specific and the precise 3D localization of a point inside the stereotactic space was easily obtained. This same arrangement is applied for MRI image guidance.

STEREOTAXIS BASED ON MRI

Magnetic resonance imaging became a major diagnostic tool for many reasons: multiplanar capability, high spatial

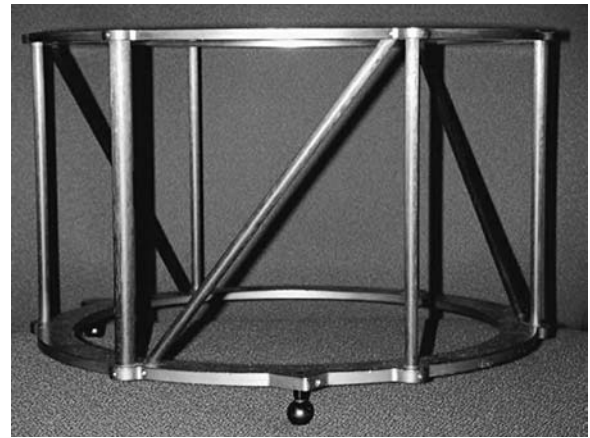


Figure 3. The CRW CT scan localizer (Radionics, Burlington, MA): The fiduciary system is composed of vertical outer bars and a diagonal internal one. Fiducial markers allow a precise 3D targeting.

resolution, excellent soft tissue contrast, absence of ionizing radiation, and bony artifact on adjacent soft tissue (40). These characteristics made MRI attractive for framed procedures and also enabled the use of MRI in frameless guided procedures (Fig. 4).

However, there are some disadvantages. The MRI is more prone to geometric distortion than CT. The maximum localization error reported in the literature ranges from 1 to 5 mm (33,40). The variation in the maximum error may be accounted for partly by the use of different stereotactic systems and different reference imaging modalities to which MRI is compared. Actually, the geometrical accuracy of MRI is not significantly different from the accuracy of conventional stereotactic frames (41,42). The spatial accuracy of MRI depends on the linearity and calibration of magnetic field gradients and on magnetic susceptibility artifacts that manifest as spatial distortions. Meticulous



Figure 4. View of the UCLA interventional MRI (Sonata, Siemens, Erlangen, Germany) operating room.

quality control should decrease errors into magnetic field linearity and calibration. High bandwidth signal acquisition reduces spatial distortion due to susceptibility effects. Several researchers have proposed practical correction algorithms to further improve the spatial accuracy of MRI (40,43,44).

Functional and biopsy cases may be planned based only on the MRI scan since there are other means to confirm final target localization. However, targeting based only on the MRI is not accurate enough when performing radiosurgery (SRS) or stereotactic radiotherapy (SRT) (45). In SRS–SRT cases, only the fusion of the CT and MRI offers the reliability required for the delivery of focused radiation to the target selected.

Several stereotactic systems with frames designed for MRI localization are commercially available: the BRW and CRW (Radionics, Burlington, MA), Leksell (Elekta Instruments Inc, Atlanta, GA), Laitinen (Sandstrom Trade & Technology Inc, Ontario, Canada), and others (10).

STEREOTACTIC FRAME-BASED PROCEDURES

Functional Stereotaxis

Functional procedures are performed under MRI guidance at UCLA. Under local anesthesia and sedation, MRI-compatible stereotactic frame (Leksell, Elekta Instruments Inc, Atlanta, GA) is attached to the patient's head. The

stereotactic frame is aligned to the MRI coil. The T1 weighted images 3 mm in thickness are obtained in the axial, coronal, and sagittal planes. The anterior and posterior commissures are identified to define the AC–PC line. The coordinates are obtained from the Shaltenbrandt and Wahren atlas. The *X*, *Y*, and *Z* coordinates can be reproduced in the axial, sagittal, and coronal MRI views by computer manipulations. The same target can be pinpointed on each of these images. The target determination is therefore better than the thickness of the imaging acquisition. Distortions of the MRI scan are not taken into account since the final position of the electrode is checked with microelectrode recording (MER) and macroelectrophysiology.

The patient is taken to the operating room either for lesion or deep brain electrode implant. The localizing microelectrode is introduced through a burr hole placed in the frontal region. Mapping with MER is obtained and the trajectory is adjusted if necessary. Parameters of electrical stimulation are manipulated aiming to improve symptoms until the point side effects are detectable. The final location of the definitive electrode or lesion site is therefore established (Table 1). The patient undergoes an intraoperative MRI to rule out bleeding and ascertain proper position of the target. Fusion of preoperative with intraoperative imaging is used to compare the accuracy of the surgery (10). This same approach is being used for cell transplantation, growth factors injection, and gene therapy (Table 2).

Table 1. Functional Stereotactic Procedures

Procedure	Indication
<i>Behavior</i>	
Amygdalotomy	Violence, aggressiveness
Anterior capsulotomy	Obsessive compulsive disorder (OCD)
Cingulotomy	Anxiety, depression, OCD
Posteriormedial Hypothalamotomy	Aggressiveness
Subcaudate tractotomy	Anxiety with depression
<i>Pain</i>	
Cingulotomy	Chronic pain emotionally charged
Dorsomedian thalamotomy	Chronic pain emotionally charged
Pulvinotomy	Intractable pain
Mesencephalotomy	Intractable pain
Periaqueductal stimulation	Intractable pain
Periventricular stimulation	Intractable pain
Cortical stimulation	Intractable pain
<i>Movement Disorders</i>	
Ventrolateral thalamotomy	Parkinson's disease, tremor, dystonia
Pallidotomy	Parkinson's disease
Campotomy (Forel's fields)	Parkinson's disease, athetosis, myoclonus
Zona incerta	Parkinson's disease, tremor, torticollis
Dentatotomy	Spasticity
Striatum fetal tissue transplant	Parkinson's disease
<i>Epilepsy</i>	
Amygdalofornicotomy/anterior Commissurotomy	Temporal lobe seizures
Pallidoamygdalotomy or Centromedian lesion	Salaam seizures
Deep electrode	Seizure focus determination

Table 2

Morphologic	Functional
<i>Diagnostic</i>	<i>Diagnostic</i>
Stereotactic biopsy	Deep electrode for seizure focus determination
<i>Therapeutic</i>	<i>Therapeutic</i>
Radiosurgery	DBS ¹ for pain control, movement and behavior disorders
Frameless neuronavigation	Ablative lesion for pain, movement and behavioral disorders
Hyperthermia	Tissue transplantation for movement disorders
Stereotactic craniotomy	Gene therapy delivery
Brachytherapy	Injection of active chemicals (growth factor)

^a Deep brain stimulator = DBS.

Stereotactic Biopsy

The determination of the biopsy target follows the same targeting procedure described above, however, electrophysiology is not used for confirmation. Instead, the histology obtained with a frozen section confirms the adequacy of the target.

Once the patient has the burr hole placed, the needle biopsy is introduced through the driver attached to the frame arc. The planning is established based on the three plan views: axial, coronal, and sagittal. Samples of tissue are collected and frozen pathology is performed. After obtaining histological diagnosis, the operative wound is closed and the patient is submitted to a postoperative MRI. Usually an air bubble is observed at the site of the biopsy, confirming the target (28). Pre- and postoperative images are fused. Minimally invasive approaches to vital structures, such as brainstem, became possible only after the development of high definition imaging methods (Fig. 5).

Radiosurgery

Radiosurgery is likely the most frequent reason for stereotactic frame application nowadays. Under local anesthesia, a CRW (Radionics, Burlington, MA) or SRS frame (BrainLab, Heimstetten, Germany) is attached to the patient's head. The patient is submitted to a CT scan and the CT-framed image is fused to the preoperative MRI. The planning starts by the drawing of the lesion, that is, in reality, the target determination. After conclusion of the planning, the patient is attached to a Novalis (BrainLab, Heimstetten, Germany) couch. Focused radiation is delivered with high precision achieved with frame-based spatial localization. At the end of the treatment, the frame is removed. In SRT cases, since the placement of a frame for 26–30 times in a patient is impracticable, a facial mask is manufactured. The frame and the fiducials are applied to the mask. The patient undergoes CT, which is fused with the MRI previously obtained. The reproducibility of accuracy with the facial mask, on a daily basis, has already been demonstrated to be ~2 mm (Solberg, personal communication). New radiosurgery devices, as Novalis and Cyberknife, allow the use of this frameless technique for radiosurgery of extracranial targets.

Stereotactic Craniotomy

This technique was popularized in the 1980s (46). Precise placement of the craniotomy based on computerized pre-surgical planning, guided retractors, and the use of microscopy were the major advancements in this area. Kelly et al. developed a very elaborate stereotactic system with integration of several imaging techniques. The purpose of this system, named COMPASS, was to guide the craniotomy and removal of deep-seated tumors under image and microscope guidance. Inspired by this idea, other similar equipments were developed. However, all these techniques lack the flexibility that neurosurgery requires. The frame frequently interferes with the craniotomy site and the limits defined for the boundaries of the lesion are not the same once the dura-mater is opened. Nowadays, neuronavigation has replaced stereotactic craniotomy.

Frameless Stereotactic Technique

The advent of image-guided neurosurgery and frameless stereotactic localization, also known as Neuronavigation, has advanced a new concept of stereotaxis (47–49). The development of multiple systems that utilize unique imaging and guidance technologies has established neuronavigation as a commonly used tool in many areas of neurosurgery, such as microsurgery for tumors (19,50–54), vascular lesions (51,52,55,56), biopsies (51,52,58–61) and epilepsy surgery (51,62,63). Moreover, modern neuronavigation techniques have begun to replace traditional image-guided tools, such as fluoroscopy and X rays (64–68). The success of neuronavigation hinges on the practical nature of this technique. It completely replaced the frame-based

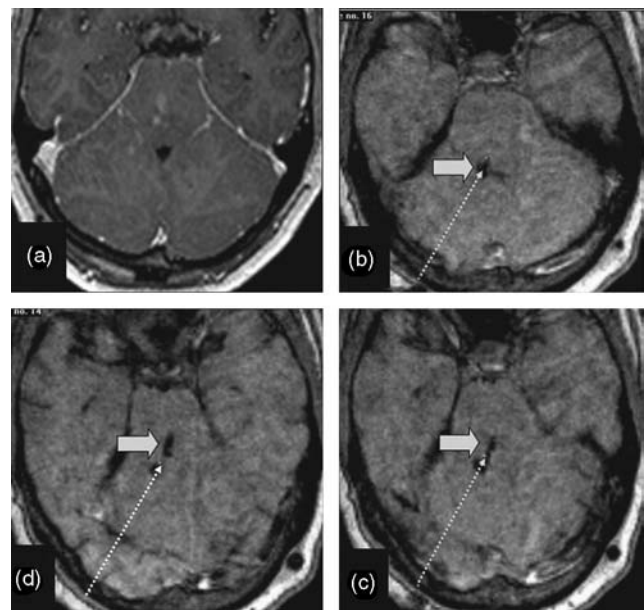


Figure 5. (A) Preoperative MRI scan showing a lesion in the brainstem, anterior to the fourth ventricle (B, C, and D). The pre- and postoperative images were fused. The site of the biopsy (arrow) and the planned trajectory of the needle (dotted arrow) can be noticed.

stereotactic craniotomy approach, which was not well accepted by the general neurosurgeon.

New approaches to lesions situated in critical areas have been developed. For example, frameless guided biopsy of lesions inside the cavernous sinus through the *foramen ovale* is now possible (28). Not long ago, procedures like this would require a craniotomy to be performed. This demonstrates how neurosurgery has become sophisticated to the point of reaching multiple brain areas through natural pathways that could not be accessed before due to lack of precise guidance.

Neuronavigation has set the stage for the application of multiple and complementary imaging techniques to aid the surgical approach. For example, the fusion of magnetoencephalography (MEG), positron emission tomography (PET), and functional MRI (fMRI) has been reported in the literature (31,69,70). Additionally, new MRI modalities, such as diffusion tensor imaging (DTI), also fused to other functional images, have made possible the recognition of important white matter structures, further increasing the safety of surgery in the depth of the brain. The orientation of the eloquent tracts in the white matter can be mapped using the fact that only diffusion anisotropy along the principal direction of the magnetic field gradient is visible (31,70).

The frameless neuronavigation system is composed of four elements: (1) registration of the image to the real anatomy of the patient, (2) interactive localization device (ILD), (3) computer interface, (4) integration of virtual and real-time data (71). Fiducial markers are the most reliable method of registration. There are two types of fiducials: mobile and rigid. Soft tissue fixation (mobile fiducials) are less accurate. Rigid fiducial markers require a minor surgical procedure for placement. On the other hand, the improvement in precision is worthy. A physical pointer can be used to identify points or surfaces to be registered with ILD.

Two types of ILD can be used: linked or nonlinked. These terms refer to the presence of an unbroken physical connection between the patient and the device. An example of the linked type is the robotic arm, while one example of unlinked is triangulation, which can be sound or infrared (IR) light based. Although of historical importance, the localizing articulating arm has been replaced by the triangulation techniques, especially by the IR reflecting devices (28,52,72–74). Most recently, radio frequency (rf) emitters and receivers were an option to replace IR triangulation devices to eliminate line of sight problems in the operating room. Localization techniques using a magnetic field created to encompass the surgical area continue in development. Its advantage is localization independent of direct linear relation of the light source and probe. It allows the utilization of curvilinear probes or even catheters.

A new issue with the real-time data acquisition is the consistency of the brain. Current registration techniques assume the brain to be a solid and nondeforming structure. However, the brain is more dynamic and moves according to its environment. Preoperative image guidance, which assumes solid brain consistency, is not necessarily accurate at all times. Brain moves after dura opening due to release of cerebral spinal fluid (CSF). The magnitude of brain shift

is variable depending on age, brain atrophy, and positioning for the procedure (50,75). This dynamic brain shifts can be overcome by real-time imaging during surgery.

Interventional MRI

New open designs of MR scanners with in-room image monitors allow MRI-guided interventional procedures. (Fig. 4) The first step toward practical interventional MRI of the brain is the development of instruments that function satisfactorily and safely in the strong magnetic field of clinical MRI scanners. Some concerns derive from this assertive.

Traditionally, surgical instruments are made of stainless steel material. Susceptibility artifacts usually develop at the extremities of the instrument. At the moment, considerable effort has been converged to test different biomaterials in the MRI magnetic field. The goal is to define the most suitable material: minimal artifact and high resistance. Less artifact is produced by ceramics, zirconium seems to produce the smallest artifact among the ceramics (76). On the other hand, ceramics present lower retention force and lower flexibility when compared to metallic materials. Susceptibility artifacts are also dependent on the pulse sequence applied. Spin echo sequences are less sensitive to time-independent local magnetic field variations while gradient spin-echo sequences are more susceptible to time-dependent and time-independent local field changes.

The strength of the magnetic field also interferes with the degree of distortion. It is possible to minimize significantly the amount of distortion using a short echo-time, thinner slices, small field of vision (FOV) and higher read-out gradients. Any material used in manufacturing the instruments will produce some amount of artifacts.

The actual location of the instrument, for example, an electrode inside the artifact, is another important issue. This becomes important to evaluate the final placement of the electrode according the planned target comparatively. It is known that artifacts develop parallel to the direction of the main magnetic field applied. The precise position of the electrode inside the distorted image was reported to be in the middle of the artifact (77).

Another concern is related to electromagnetic interference between the MRI scanner and interventional electronics. For example, the rf generator may emit electromagnetic radiation that interferes with image acquisition and generates artifacts. At UCLA, the use of MER in the MRI is routine, either the 0.2 (70) or the 1.5 T (Siemens, Erlangen, Germany), with special attention to work in the fringes of the strong magnetic field. Interferences can be safely avoided in so far as attention is centered in the strategic location of the devices inside the MRI room.

To approximate real-time imaging, either image acquisition or image reconstruction has to be of very high speed. Currently, millimetric resolution in the 0.2 T and submillimetric in the 1.5 T can be achieved with the parameters reported at Table 3.

Interventional MRI scanners offer the possibility of image acquisition in advance for planning the procedure, intraoperatively for compensation of brain shift that occurs

Table 3. Imaging Acquisition Parameters^a

Magnetic Field Strength	Low Field, 0.2 T		High Field, 1.5 T		
Pulse sequences	T2W FL2D	3 DF	T1W MPRAGE	T2W TSE	FL 2D
Region	WB	WB	WB	BG	WB
TR, ms	1500	56	2050	2800	800
TS, ms	60	25	4.4	84	15
FOV, mm	230	280	280	280	280
Matrix	224 × 256	256 × 256	256 × 256	256 × 256	168 × 256
Voxel Size, mm	1.0 × 0.9 × 6.0	1.1 × 1.1 × 3.0	1.1 × 1.1 × 1.0	1.1 × 1.1 × 2.0	1.5 × 1.1 × 7.0
Scan Time, min:s	11:16	7:42	17:31	2:47	4:28

^aT1W = T1 weighed. TSE = turbo spin echo. T2W = T2 weighed. WB = whole brain. FL 2D = flair 2D. BG: basal ganglia. 3 DF = 3D flash. FOV = field of view. MPRAGE = magnetization prepared rapid acquisition gradient echo.

after dura opening (50,78) and postoperatively for early detection of complications and confirmation of target localization. The reliability of intraoperative image acquisition is superior to the method based only on the preoperative data setting for targeting.

APPLICATIONS

Stereotactic surgery is widely applied for morphological and functional procedures in the brain. Functional and morphologic stereotaxis can be subdivided into diagnostic and therapeutic procedures. A brief summary of these applications is presented in Tables 1 and 2.

FUTURE DIRECTIONS

Advances in imaging techniques are characterized by faster acquisition sequences, more precise definition of eloquent structures, and minimization of artifacts. It is feasible to predict a more sophisticated integration among imaging, virtual reality computer, and robotics in a near future. Possibly new surgical tools as thermal ablation, cryoablation, and chemoablation will become routine, not only for the brain, but for the whole body. The MRI would also be able to provide sensitive monitoring of temperature changes and tissue injury with high temporal resolution. This would be an advantage once minimal injury inside or surrounding an eloquent area may lead to a major deficit (79), obviating the need of large surgical access to remove tumors.

The increasing amount of publications (57,80–86) discussing preliminary results using interventional MRI technique, exploring types of biomaterials (76,87–92), and new sequences of image acquisition (31,70) clearly show that interventional MRI is in its early stages of development. Whether frameless procedures under intraoperative MRI guidance will replace frame stereotaxis in the future is still an open question. Studies already claim that the same accuracy obtained with framed technique is achieved (33,40,41,87).

Frameless stereotactic concepts of registration of images to patients' anatomy will continue to be the basis of surgical guidance, even with real-time imaging. Technological advances as interventional MRI carry issues like high costs. Nevertheless, due to its minimally invasive nature compared with other open surgical approaches,

image guidance married to interventional MRI may emerge as a means of actually lowering the overall cost of medical care.

BIBLIOGRAPHY

1. Bakay RA et al. Delayed stereotactic transplantation technique in non-human primates. *Prog Brain Res* 1988;78:463–471.
2. Marino RJ, Gonzales-Portillo M. Preconquest Peruvian neurosurgeons: a study of Inca and pre-Columbian trephination and the art of medicine in ancient Peru. *Neurosurgery* 2000;47:940–950.
3. DN Z. L'encéphalomètre. *Rev Gen Clin Ther* 1980;19:302.
4. Lichterman BL. Roots and routes of Russian neurosurgery (from surgical neurology towards neurological surgery). *J Hist Neurosci* 1998;7:125–135.
5. RD. *Discours de la Methode*. Paris: Vrin; 1992.
6. Clarke RH, Horsley V. On a method of investigating the deep ganglia and tracts of the central nervous system (cerebellum). *Br Med J* 1906;1799–1802:1799–1800.
7. Horsley VC R. The structure and functions of the cerebellum examined by a new method. *Brain* 1908;31:45–124.
8. Olivier A, Bertrand G, Picard C. Discovery of the first human stereotactic instrument. *Appl Neurophysiol* 1983;46:84–91.
9. Picard C, Olivier A, Bertrand G. The first human stereotactic apparatus. The contribution of Aubrey Mussen to the field of stereotaxis. *J Neurosurg* 1983;59:673–676.
10. De Salles AA. Stereotactic applications. In: De Salles AA, Goetsch S, editors. *Stereotactic Surgery and Radiosurgery*. Madison (WI): Medical Physics Publishing; 1993.
11. Moniz E. L'encéphalographie arterielle, son importance dans la localisation des tumeurs cerebrales. *Rev Neurol* 1927;2:72–90.
12. Dandy WE. Localization of brain tumors by cerebral pneumography. *Am J Roentgenol* 1923;10:610–612.
13. Dandy WE. Ventriculography following the injection of air into the cerebral ventricles. *Ann Surg* 1918;68:5–11.
14. Spiegel EA WH, Marks M, Lee ASJ. Stereotactic apparatus for operations of the human brain. *Science* 1947;106:349–350.
15. Spiegel EA WH, Baird HW. Studies in stereoecephalotomy. I. Topical relationships of subcortical structures to the posterior commissure. *Confin Neurolog* 1952;12:9–133.
16. Benabid AL, et al. Chronic electrical stimulation of the ventralis intermedius nucleus of the thalamus as a treatment of movement disorders. *J Neurosurg* 1996;84:203–214.
17. R. M. Surgical interruption of the pallidofugal fibers: its effect on the syndrome of paralysis agitans and technical considerations in its application. *NY State J Med* 1942;42:317–325.

18. Vitek JL, et al. Microelectrode-guided pallidotomy: technical approach and its application in medically intractable Parkinson's disease. *J Neurosurg* 1998;88:1027-1043.
19. Kelly PJ. Computer-assisted stereotaxis: new approaches for the management of intracranial intra-axial tumors. *Neurology* 1986;36:535-541.
20. Hirai T, Miyazaki M, Nakajima H, Shibasaki T, Ohye C. The correlation between tremor characteristics and the predicted volume of effective lesions in stereotaxic nucleus ventralis intermedia thalamotomy. *Brain* 1983;106 (Pt. 4):1001-1018.
21. Laitinen LV. Pallidotomy for Parkinson's disease. *Neurosurg Clin N Am* 1995;6:105-112.
22. Matsumoto K, Shichijo F, Fukami T. Long-term follow-up review of cases of Parkinson's disease after unilateral or bilateral thalamotomy. *J Neurosurg* 1984;60:1033-1044.
23. Nagaseki Y, et al. Long-term follow-up results of selective VIM-thalamotomy. *J Neurosurg* 1986;65:296-302.
24. Pollak P, et al. Long-term effects of chronic stimulation of the ventral intermediate thalamic nucleus in different types of tremor. *Adv Neurol* 1993;60:408-413.
25. Burchiel KJ. Thalamotomy for movement disorders. *Neurosurg Clin N Am* 1995;6:55-71.
26. Lozano A, et al. Methods for microelectrode-guided posteroventral pallidotomy. *J Neurosurg* 1996;84:194-202.
27. Tronnier VM, Fogel W, Kronenburger M, Steinvorth S. Pallidal stimulation: An alternative to pallidotomy? *J Neurosurg* 1997;87:700-705.
28. Frighetto L, et al. Image-guided frameless stereotactic biopsy sampling of parasellar lesions. Technical note. *J Neurosurg* 2003;98:920-925.
29. Grunert P, et al. Frame-based and frameless stereotaxy in the localization of cavernous angiomas. *Neurosurg Rev* 2003;26:53-61.
30. Walker DG, Ohaegbulam C, Black PM. Frameless stereotaxy as an alternative to fluoroscopy for transsphenoidal surgery: Use of the InstaTrak-3000 and a novel headset. *Clin Neurosci* 2002;9:294-297.
31. Kamada K, et al. Visualization of the eloquent motor system by integration of MEG, functional, and anisotropic diffusion-weighted MRI in functional neuronavigation. *Surg Neurol* 2003;59:352-361. discussion 361-352.
32. Roessler K, et al. Frameless stereotactic lesion contour-guided surgery using a computer-navigated microscope. *Surg Neurol* 1998;49:282-288. discussion 288-289.
33. Samset E, Hirschberg H. Image-guided stereotaxy in the interventional MRI. *Minim Invasive Neurosurg* 2003;46:5-10.
34. Samset E, Hirschberg H. Neuronavigation in intraoperative MI. *Comput Aided Surg* 1999;4:200-207.
35. Vitaz TW, Hushek SG, Shields CB, Moriarty TM. Interventional MRI-guided frameless stereotaxy in pediatric patients. *Stereotact Funct Neurosurg* 2002;79:182-190.
36. L. L. A stereotaxic apparatus for intracerebral surgery. *Acta Chir Scand* 1949;99:229-233.
37. PL G. Principles of stereotaxis and instruments In: SJ DSAG editors. *Stereotactic Surgery and Radiosurgery*. Madison, (WI): Medical Physics Publishing, 1993.
38. Talairach JP. *Co-Planar Stereotaxic Atlas of the Human Brain*. New York: Thieme; 1988.
39. Shaltenbrand GWW. *Atlas of Stereotaxy for the Human Brain*. Chicago: Thieme, 1977.
40. De Salles AA, Gronemeyer D, Seibel R, Lufkin R. Instrumentation for Interventional MRI of the Brain. In: De Salles A, Lufkin R., editors. *Minimally Invasive Therapy of the Brain*. New York: Thieme; 1997.
41. Carter DA, Parsai EI, Ayyangar KM. Accuracy of magnetic resonance imaging stereotactic coordinates with the cosman-roberts-wells frame. *Stereotact Funct Neurosurg* 1999;72:35-46.
42. Galloway RL Jr, Maciunas RJ, Latimer JW. The accuracies of four stereotactic frame systems: An independent assessment. *Biomed Instrum Technol* 1991;25:457-460.
43. Bakker CJ, Moerland MA, Bhagwandien R, Beersma R. Analysis of machine-dependent and object-induced geometric distortion in 2DFT MR imaging. *Magn Reson Imaging* 1992;10: 597-608.
44. Fitzpatrick J, et al. A technique for improving accuracy in positron and intensity within images acquired in the presence of field inhomogeneity. New York: SMRM; 1990.
45. Solberg TD MP, DeMarco J, De Salles AAF, Selch MT. Technical aspects of LINAC radiosurgery for functional disorders. *J Radiosurgery* 1998;1:115-127.
46. Kelly PJ, Alker GJ Jr, Kall BA, Goerss S. Method of computed tomography-based stereotactic biopsy with arteriographic control. *Neurosurgery* 1984;14:172-177.
47. Kelly PJ, Alker GJ Jr, Goerss S. Computer-assisted stereotactic microsurgery for the treatment of intracranial neoplasms. *Neurosurgery* 1982;10:324-331.
48. Roberts DW, et al. A frameless stereotaxic integration of computerized tomographic imaging and the operating microscope. *J Neurosurg* 1986;65:545-549.
49. Watanabe E, et al. Three-dimensional digitizer (neuronavigator): new equipment for computed tomography-guided stereotaxic surgery. *Surg Neurol* 1987;27:543-547.
50. Barnett GH, Kormos DW, Steiner CP, Weisenberger J. Use of a frameless, armless stereotactic wand for brain tumor localization with two-dimensional and three-dimensional neuroimaging. *Neurosurgery* 1993;33:674-678.
51. Gofinos JG, Fitzpatrick BC, Smith LR, Spetzler RF. Clinical use of a frameless stereotactic arm: Results of 325 cases. *J Neurosurg* 1995;83:197-205.
52. Gumprecht HK, Widenka DC, Lumenta CB. BrainLab VectorVision Neuronavigation system: Technology and clinical experiences in 131 cases. *Neurosurgery* 1999;44:97-104. discussion 104-105.
53. Reinhardt HF, Zweifel HJ. Interactive sonar-operated device for stereotactic and open surgery. *Stereotact Funct Neurosurg* 1990; 54-55. 393-397.
54. Suess O, et al. Intracranial image-guided neurosurgery: Experience with a new electromagnetic navigation system. *Acta Neurochir (Wien)* 2001;143:927-934.
55. Reinhardt HF, Horstmann GA, Gratzl O. [Microsurgical removal of deep vascular malformations using sonar- stereometry]. *Ultraschall Med* 1991;12:80-83.
56. Zamorano L, et al. Interactive image-guided surgical resection of intracranial arteriovenous malformations. *Comput Aided Surg* 1998;3:57-63.
57. Barnett GH, Miller DW, Weisenberger J. Frameless stereotaxy with scalp-applied fiducial markers for brain biopsy procedures: experience in 218 cases. *J Neurosurg* 1999;91:569-576.
58. Germano IM, Queenan JV. Clinical experience with intracranial brain needle biopsy using frameless surgical navigation. *Comput Aided Surg* 1998;3:33-39.
59. Grunert P, et al. Stereotactic biopsies guided by an optical navigation system: Technique and clinical experience. *Minim Invasive Neurosurg* 2002;45:11-15.
60. Sawin PD, Hitchon PW, Follett KA, Torner JC. Computed imaging-assisted stereotactic brain biopsy: A risk analysis of 225 consecutive cases. *Surg Neurolog* 1998;49:640-649.
61. Ulm AJ, Bova FJ, Friedman WA. Stereotactic biopsy aided by a computer graphics workstation: Experience with 200 consecutive cases. *Surg Neurol* 2001;56:366-371. discussion 371-362.

62. Olivier A, Germano IM, Cukiert A, Peters T. Frameless stereotaxy for surgery of the epilepsies: Preliminary experience. Technical note. *J Neurosurg* 1994;81:629–633.
63. Wurm G, et al. Advanced surgical approach for selective amygdalohippocampectomy through neuronavigation. *Neurosurgery* 2000;46:1377–1382. discussion 1382–1373.
64. Choi WW, Green BA, Levi AD. Computer-assisted fluoroscopic targeting system for pedicle screw insertion. *Neurosurgery* 2000;47:872–878.
65. Dresel SH, et al. Meckel cave lesions: Percutaneous fine-needle-aspiration biopsy cytology. *Radiology* 1991;179:579–582.
66. Foley KT, Simon DA, Rampersaud YR. Virtual fluoroscopy: computer-assisted fluoroscopic navigation. *Spine* 2001; 26:347–351.
67. Sheporaitis LA, et al. Intracranial meningioma. *AJNR Am J Neuroradiol* 1992;13:29–37.
68. Welch WC, Subach BR, Pollack IF, Jacobs GB. Frameless stereotactic guidance for surgery of the upper cervical spine. *Neurosurgery* 1997;40:958–963. discussion 963–954.
69. Reithmeier T, et al. Neuronavigation combined with electrophysiological monitoring for surgery of lesions in eloquent brain areas in 42 cases: A retrospective comparison of the neurological outcome and the quality of resection with a control group with similar lesions. *Minim Invasive Neurosurg* 2003;46:65–71.
70. Tummala RP, Chu RM, Liu H, T T, Hall WA. Application of diffusion tensor imaging to magnetic-resonance-guided brain tumor resection. *Pediatr Neurosurg* 2003;39:39–43.
71. Maciunas RJ. Approaches to Frame-Based and Frameless Stereotaxis. New York: Thieme; 1997.
72. Bohinski RJ, et al. Glioma resection in a shared-resource magnetic resonance operating room after optimal image-guided frameless stereotactic resection. *Neurosurgery* 2001; 48:731–742. discussion 742–734.
73. Germano IM, Villalobos H, Silvers A, Post KD. Clinical use of the optical digitizer for intracranial neuronavigation. *Neurosurgery* 1999;45:261–269. discussion 269–270.
74. Schroeder HW, Wagner W, Tschiltshcke W, Gaab MR. Frameless neuronavigation in intracranial endoscopic neurosurgery. *J Neurosurg* 2001;94:72–79.
75. Nabavi A et al. [Neuronavigation. Computer-assisted surgery in neurosurgery]. *Radiologe* 1995;35:573–577.
76. Matsuura H, et al. Quantification of susceptibility artifacts produced on high-field magnetic resonance images by various biomaterials used for neurosurgical implants Technical note. *J Neurosurg* 2002;97:1472–1475.
77. Yelnik J, et al. Localization of stimulating electrodes in patients with Parkinson disease by using a three-dimensional atlas-magnetic resonance imaging coregistration method. *J Neurosurg* 2003;99:89–99.
78. Nabavi A, et al. Serial intraoperative magnetic resonance imaging of brain shift. *Neurosurgery* 2001;48:787–797. discussion 797–788.
79. Farahani K, et al. Effect of field strength on susceptibility artifacts in magnetic resonance imaging. *Comput Med Imaging Graph* 1990;14:409–413.
80. Bernays RL, et al. Histological yield, complications, and technological considerations in 114 consecutive frameless stereotactic biopsy procedures aided by open intraoperative magnetic resonance imaging. *J Neurosurg* 2002;97:354–362.
81. Black PM, et al. Development and implementation of intraoperative magnetic resonance imaging and its neurosurgical applications. *Neurosurgery* 1997;41:831–842. discussion 842–835.
82. Hall WA, et al. Brain biopsy using high-field strength interventional magnetic resonance imaging. *Neurosurgery* 1999;44:807–813. discussion 813–804.
83. Bradford R, Thomas DG, Bydder GM. MRI-directed stereotactic biopsy of cerebral lesions. *Acta Neurochir. (Suppl.)* (Wien) 1987;39:25–27.
84. Dorward NL, Paleologos TS, Alberti O, Thomas DG. The advantages of frameless stereotactic biopsy over frame-based biopsy. *Br J Neurosurg* 2002;16:110–118.
85. Fahlbusch R, Ganslandt O, Nimsky C. Intraoperative imaging with open magnetic resonance imaging and neuronavigation. *Childs Nerv Syst* 2000;16:829–831.
86. Moriarty TM, et al. Frameless stereotactic neurosurgery using intraoperative magnetic resonance imaging: stereotactic brain biopsy. *Neurosurgery* 2000;47:1138–1145. discussion 1145–1136.
87. Koyama T, Handa J. Porous hydroxyapatite ceramics for use in neurosurgical practice. *Surg Neurol* 1986;25:71–73.
88. Nitatori T, Hanaoka H, Hachiya J, Yokoyama K. MRI artifacts of metallic stents derived from imaging sequencing and the ferromagnetic nature of materials. *Radiat Med* 1999;17:329–334.
89. Port JD, Pomper MG. Quantification and minimization of magnetic susceptibility artifacts on GRE images. *J Comput Assist Tomogr* 2000;24:958–964.
90. Shellock FG, Shellock VJ. Ceramic surgical instruments: Ex vivo evaluation of compatibility with MR imaging at 1.5 T. *J Magn Reson Imaging* 1996;6:954–956.
91. Shellock FG, Shellock VJ. Spetzler titanium aneurysm clips: compatibility at MR imaging. *Radiology* 1998;206: 838–841.
92. Tominaga T, et al. Magnetic resonance imaging of titanium anterior cervical spine plating systems. *Neurosurgery* 1995; 36:951–955.
93. Dorward NL, et al. Accuracy of true frameless stereotaxy: In vivo measurement and laboratory phantom studies. Technical note. *J Neurosurg* 1999;90:160–168.

See also HYPERTHERMIA, INTERSTITIAL; RADIOSURGERY, STEREOTACTIC; TISSUE ABLATION.

STERILIZATION. See CONTRACEPTIVE DEVICES.

STERILIZATION OF BIOLOGIC SCAFFOLD MATERIALS

DONALD O. FREYTES
STEPHEN F. BADYLAK
University of Pittsburgh
Pittsburgh, Pennsylvania

INTRODUCTION

Biologic materials such as collagen, tissue grafts, and extracellular matrix (ECM) derived scaffolds are used for tissue and organ repair and are typically regulated as devices by the U.S. Food and Drug Administration (FDA). As such, these materials must be properly processed and sterilized prior to clinical use. Sterilization of biologic materials involves unique considerations (e.g., the

shrink temperature of collagen, changes in the quaternary ultrastructure of the matrix materials, potential inactivation of any bioactive components, and effects upon surface chemistry and architecture). The purpose of this article is to discuss currently used methods of sterilization for biomaterials with emphasis upon biologic materials, the mechanism by which these sterilization methods destroy or neutralize microbes of interest, and the potential effects of the sterilization methods upon the structure and function of the naturally occurring biologic materials and their inherent bioactive constituents.

STERILIZATION METHODS

The FDA recognizes several effective methods of sterilization for medical devices. These methods include dry heat, moist heat (autoclave), ethylene oxide, ionizing radiation, and liquid chemical sterilants. Each of these methods involves a different mechanism of action for killing microbes, and therefore, the physical and chemical effects of these methods upon naturally occurring molecules will differ. A glossary of relevant terms for this article and the topic of sterilization is provided at the end of the article. A list of current international standards for each of the four sterilization methods described herein can be found in Table 1.

Sterilization Versus Disinfection

There is a clear distinction between the terms sterilization and disinfection. The FDA (1997) defines disinfection as the destruction of pathogenic and other types of microorganisms by thermal or chemical means (1). Sterilization is defined as a process intended to remove or destroy all viable forms of microbial life, including bacterial spores, in order to achieve an acceptable level of sterilization (1). Stated differently, sterilization implies the inactivation and removal of all forms of life. The term terminal sterilization refers to the last sterilization step performed prior to use or commercial distribution of a device.

Bioburden and Sterility Assurance Level

Before sterilizing a device, a certain amount of microbial debris remains (including bacterial wall remnants: pyrogens and endotoxins) on and within each device as a

Table 1. Summary of Sterilization Standards^a

Sterilization Method	Standard(s) ^b
Heat Sterilization	ISO 20857
Steam Sterilization	ISO 11134, ISO/DIS 17665-1, ISO/DIS 17665-2
Ethylene Oxide	ISO/DIS 11135-1, ISO/DIS 11135-2, ISO 10993-7 (Residuals)
Radiation	ISO/DIS 11137-1, ISO/DIS 11137-2, ISO/DIS 11137-3
Others	ISO 10993 (1-18), ASTM E1766-95, F2347-03, F2103-1, F2064-00

^aSee <http://www.iso.org>; <http://www.astm.org>.

^bISO – International Organization for Standardization; ASTM – American Society for Testing and Materials; DIS – Draft International Standard.

result of production and presterilization processing steps. This microbial debris is referred to as the *bioburden*. The bioburden that exists prior to sterilization is directly proportional to the difficulty of sterilization of a medical device. It is almost impossible to remove all organisms from some materials, and therefore acceptable levels of bioburden have been identified for which devices can be considered as sterile. These acceptable levels are referred to as the sterility assurance level (SAL), which represents the number of microorganisms that would be tolerable, or conversely, the probability that the device is nonsterile (e.g., 10^{-3} means there is a 1 in 10^{-3} chance that an organism survived the sterilization process). The specific SAL for each device will depend on the intended clinical application for the device and the standards established by regulatory agencies. The FDA recommends that implantable devices have a SAL of 10^{-6} while devices contacting intact skin may have a SAL of 10^{-3} (1,2).

Validation of Sterilization Methods

Validation studies of the chosen sterilization method must be performed to ensure proper reduction of endotoxins and pyrogens and appropriate SAL levels, while minimizing exposure to the sterilant. During these validation studies, the microorganisms are quantified by conventional techniques (1,3) following the sterilization process. Various parameters of the sterilization procedure are evaluated (e.g., time, heat, gas concentration), and then the devices are tested for sterility. A classic example of this procedure is the determination of the time needed to sterilize a device by dry heat at a fixed temperature. Briefly, devices are placed at the chosen temperature and samples are periodically removed at different time intervals (e.g., every 5 or 10 min) depending on the overall duration of the test. The number of surviving organisms is quantified at each collection time, the logarithm of the surviving organisms computed, and the values plotted against exposure time. From this curve, the exposure time to achieve the target SAL can be extrapolated as shown in Fig. 1. The same concept can be applied to gas and chemical exposure times.

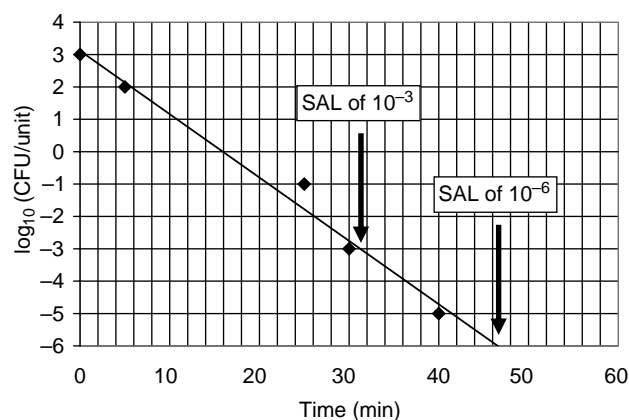


Figure 1. Example of data obtained from a fractional sterilization run to determine the time required to achieve the desired SAL at a fixed temperature. *Note:* The data may not have a linear form and will depend on the microbial load and the type of microbes present.

The determination of the presence of pyrogens and endotoxins levels in a device is another aspect of the validation process. Pyrogens are fever inducing agents most commonly associated with Gram-negative bacteria, but can be produced by most microorganisms. Endotoxins are lipopolysaccharides from the cell wall of Gram-negative bacteria capable of inducing certain inflammatory responses. The most common test used to determine the levels of bacterial endotoxins is the limulus amoebocyte lysate (LAL) test (1,3). Pyrogen and endotoxin removal can be achieved via the use of acids, alkaline hydrolysis, hydrogen peroxide, dry heat destruction, and filtration (4).

HEAT STERILIZATION

Heat sterilization can be either applied in a dry or a moist form. Dry heat sterilization is a commonly used method for the sterilization of metallic instruments, powders, and petroleum products (5). Moist heat sterilization is commonly used for surgical instruments and equipment used for cell culture. Moist heat sterilization has also been used for bone grafts to destroy human immunodeficiency (HIV) and hepatitis viruses (6). Except for the sterilization of bone grafts, dry or moist heat is rarely used for biologic materials. Each form of heat sterilization will be discussed separately.

Dry Heat Sterilization

Mechanism of Action. Dry heat sterilization makes use of heated air to inactivate and/or destroy microorganisms. The mechanism of microbial death involves denaturation and coagulation of nucleic acids and proteins. A reduction in the water content of bacterial spores is also considered to be a mechanism by which dry heat sterilization is effective. Although oxidation is possible, it is less widely accepted as a mechanism by which dry heat sterilization exerts its effects (5).

The application of dry heat sterilization utilizes a dry oven-like environment. The air inside the chamber is heated and allowed to equilibrate to a constant temperature (from 120 to 170 °C depending on the duration). The duration of the procedure is a factor of the applied temperature and is determined during the validation process as described earlier. The time and temperature may also vary depending on the nature of the material being sterilized. Table 2 provides guidelines for the sterilization of medical devices using dry heat.

Table 2. Temperatures and Suggested Sterilization Time for a Typical Medical Device^a

Temperature	Sterilization Time
170 °C (338 °F)	60 min (1 h)
160 °C (320 °F)	120 min (2 h)
150 °C (302 °F)	150 min (2.5 h)
140 °C (284 °F)	180 min (3 h)
121 °C (250 °F)	Overnight

^aTimes suggested by Perkins (5).

Advantages And Disadvantages. The major advantages of dry heat sterilization are low cost and the compatibility with anhydrous oils, powders, and materials not affected by high temperatures. By definition, the moisture content of dry heat sterilization is low. Disadvantages of dry heat sterilization include a relatively long sterilization time and the use of elevated temperatures. Heat diffusion will depend on the ability of the dry heat to diffuse throughout the chamber and the type of material–device being sterilized. The killing rate may be slow due to delayed heat conduction of some materials in which the amount of heat delivered to the microorganisms is limited. Dry heat sterilization cannot be used for most liquids, heat labile substances (i.e., proteins), and heat sensitive materials (1,5).

Physical Effect upon Materials. High temperature can alter bulk properties of polymers and composites and can melt polymers and materials that have a melting temperature below or near the temperature being used for sterilization. The melting temperature of most linear polymers is below the temperatures commonly used in dry heat sterilization (7). Even when the melting temperature is above the temperatures employed, oxidation may still occur in polymers (e.g., nylon) (7). However, polytetrafluoroethylene (PTFE) and silicon rubbers may be effectively sterilized using dry heat (7). High temperatures can adversely affect structural proteins (e.g., collagen and elastin) and modify the mechanical properties of biologic materials (8–10). The high temperature used for dry heat sterilization may also denature bioactive molecules (e.g., growth factors and bioactive peptides present in tissue grafts and naturally occurring biomaterials).

Moist Heat Sterilization (Autoclave)

Mechanism of Action. The mechanism by which moist heat sterilization kills microbes is similar to the mechanism described for dry heat sterilization. Protein and nucleic acid coagulation and denaturation occur quickly once a critical temperature is reached. For the same reasons that protein and nucleic acid destruction inhibit the ability of bacteria to replicate or continue metabolic processes, this method is generally unacceptable for biologic materials that are composed of naturally occurring proteins (5). The high temperatures used with moist heat sterilization are effective for the inactivation of viruses and bacterial spores.

Moist heat (steam) sterilization or autoclaving is conducted under pressurized conditions with saturated steam usually at 121–125 °C. The process typically lasts from 15 to 30 min to ensure that all surfaces are exposed to the moist heat.

Advantages And Disadvantages. The advantages of moist heat sterilization are its efficacy, the relatively low temperature requirements (compared to dry heat sterilization), speed, process simplicity, and the lack of toxic residues when compared to methods such as ethylene oxide (ETO) sterilization (discussed in the next section). Since moist heat sterilization is performed in a pressurized

environment, it can also be used to sterilize liquids. The disadvantages of moist heat sterilization include the presence of water, the use of elevated temperatures, and the possible deposition of impurities present in the steam. Moist heat sterilization is obviously not suitable for heat labile materials or synthetic and natural polymers that are readily degraded by hydrolysis.

Physical Effect upon Materials. Moist heat sterilization can change bulk properties of polymers by hydrolysis and the hydrolytic byproducts may result in the formation of contaminants. For example, exposure of methyl diisocyanate based polyurethanes to prolonged steam sterilization results in the formation of methylene dianiline (due to hydrolysis), which leads to decreased lung function when used in lung perfusion devices (11). Polymers that are prone to hydrolytic degradation include poly(vinyl chlorides) (PVC), polyacetals, polyethylenes, and polyamides (7). For example, the mechanical properties of PVC can be adversely affected by repeated steam sterilization due to rearrangement of macromolecular chains (12). A separate example of the potential harmful effects of moist heat sterilization is the formation of oligomer crystals on the surface of Dacron grafts as a result of steam sterilization that can cause hemolysis when this material is used as a vascular graft (13). With respect to biologic materials, steam sterilization can have adverse effects upon the mechanical properties of tissue grafts as in the case of bone allografts (14). For obvious reasons, the denaturing effects of steam sterilization upon protein structures caused by the elevated temperatures makes this method generally unsuitable for most biologic materials.

In summary, both dry and moist heat forms of sterilization, although effective for nonbiologic materials, are typically unsuitable for the sterilization of biologic materials, due to adverse effects upon the structure and function of protein and non-protein constituents.

ETHYLENE OXIDE STERILIZATION

Ethylene oxide sterilization is a commonly used method for the sterilization of heat-sensitive materials, medical equipment, and biologic materials including those composed of ECM, such as TissueMend (fetal bovine skin from TEI Biosciences), and OaSis (small intestinal submucosa/SIS extracellular matrix from Cook Biotech, Inc.). Sterilization via ETO has also been explored for demineralized bone, tendons, dura mater, and fascia lata (15).

Mechanism of Action. Ethylene oxide is an unstable ring structure capable of reacting via alkylation with functional groups found in nucleic acids and proteins (1,16,17). Examples of reactions between functional groups and ETO are listed in Fig. 2. Sterilization via ETO exposure begins with the placement of the target device into a pressurized sterilization chamber. The humidity within the chamber is controlled by the introduction of moisture (40–90% humidity) and the temperature is maintained between 40 and 50 °C. The ETO is then introduced into the chamber at concentrations ranging from 600 to 1200 mg·L⁻¹ for a sufficient time (typically 2–48 h) to achieve the desired SAL. Following sterilization, room air is used to flush the vessel for removal of residual ETO and its toxic byproducts. Longer flushing time minimizes the presence of toxic byproducts but increases the overall sterilization time.

Advantages and Disadvantages. The advantages of using ETO sterilization include the relatively low temperature requirements (compared to heat sterilization), and the high degree of penetration of the ETO gas into the target device. Temperature and moisture sensitive materials are more readily sterilized via ETO than by heat sterilization methods. Perhaps the greatest disadvantage of ETO sterilization is the toxicity and carcinogenicity of the residual byproducts: ETO, ETC, and ETG (see Fig. 3). For samples weighing >100 mg, acceptable levels of ETO, ETC, and

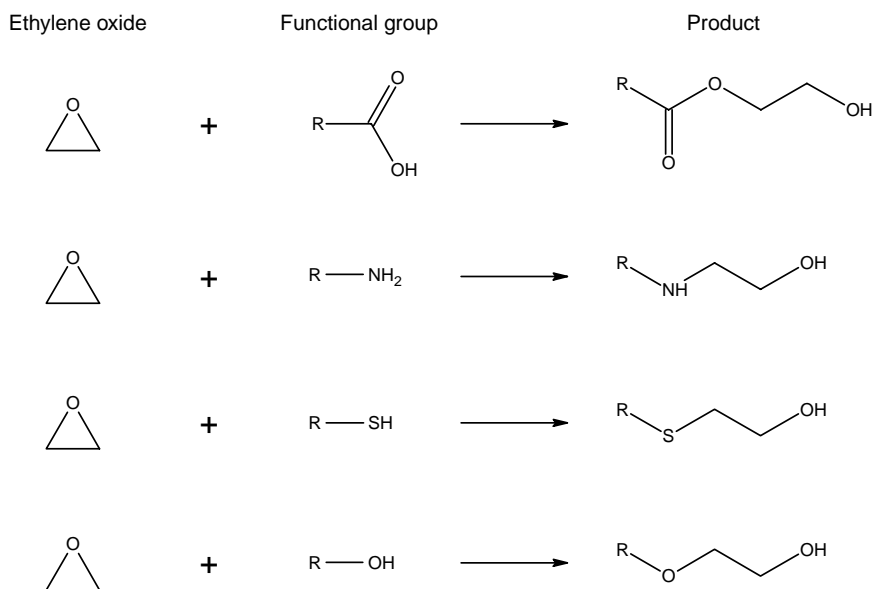


Figure 2. Examples of possible reaction between ETO and common functional groups in biological molecules (16,17). For a more complete explanation of the reactions please refer to (16,19).

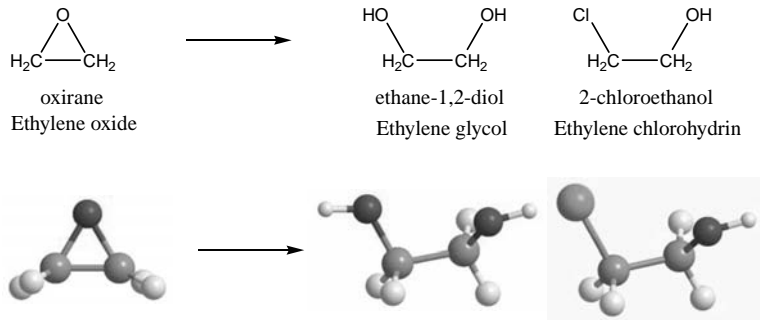


Figure 3. Structure of ETO and its byproducts: Ethylene glycol (ETG) and ethylene chlorohydrin (ETC). The ETO is flammable and highly explosive and has been associated with potential mutagenic, teratogenic, or carcinogenic properties. Both ETG and ETC are toxic as well (18).

ETG are 2500, 2500, and 5000 ppm, respectively (1,17). In addition to the alkylation reaction referred to above, ETO may also react with functional groups on the surface of biomaterials and with constituent proteins causing changes in bioactivity and bulk properties of the device.

Physical Effect upon Materials. Materials coated with proteins (e.g., albumin and heparin) may lose the benefit of these coatings as a result of ETO interactions with these surface molecules (20). Protein cross-links, alkylation reactions, and other chemical changes that occur as a result of ETO interactions may also have adverse effects upon the bioactivity and the mechanical properties of biologic materials (e.g., allografts) (15,21–23).

In summary, ETO sterilization with proper aeration and/or removal of toxic by products may be safely used to sterilize biologic materials. However, changes in bioactivity and structural properties can still occur using this method of sterilization.

IONIZING RADIATION STERILIZATION

Sterilization via radiation is the most commonly used technique to sterilize biologic materials at the present time. Examples of such devices include Restore™ (SIS-ECM used for orthopedic applications, DePuy Orthopedics), CuffPatch™ (carbodiimide crosslinked SIS-ECM, Arthro-tech), and Permacol™ (crosslinked porcine dermis, Tissue Science Laboratories). Ionizing radiation is the preferred method of sterilization for other biologic materials such as heart valves, skin, fascia, dura mater, bone, and tendon grafts (24). There are different types of irradiation, including gamma irradiation and electron beam irradiation, that are approved by the FDA and commonly used by medical device manufacturers. These methods will be discussed separately.

Gamma Irradiation

Mechanism of Action. Gamma radiation arises from the decay of cobalt (^{60}Co) or cesium isotopes (^{137}Cs). Gamma irradiation exists in the form of photons generated from the transition of an atomic nucleus from an excited state to a ground state. The resultant high energy particles induce ionization by transforming an uncharged atom to a charged atom with the subsequent release of an electron that in turn collides with other atoms. The resulting discharge of

secondary electrons creates oxidizing free radicals that damage proteins and deoxyribonucleic acid (DNA) molecules by dimerization of bases and altering the sugar phosphate backbone. These changes reduce microbe's capacity to replicate or continue necessary metabolic functions (25).

Gamma sterilization is conducted by placing the target device in front of a radiation source, usually cobalt (^{60}Co) or cesium (^{137}Cs), that is usually directed by a window in a lead shielded container (Fig. 4). The dose of gamma irradiation is adjusted by varying the distance between the source and the device or by varying the exposure time. The dose or radiation absorption is commonly expressed in terms of “radiation absorbed dose” (rad) or grays (Gy). The former represents the absorption of 10^{-5} joules (J) per gram (g) ($\text{J}\cdot\text{g}^{-1}$) while the later represents the absorption of $1 \text{ J}\cdot\text{kg}^{-1}$. Hence, $1 \text{ kGy} = 0.1 \text{ Mrads}$. Typical sterilizing doses range between 6 and 25 kGy.

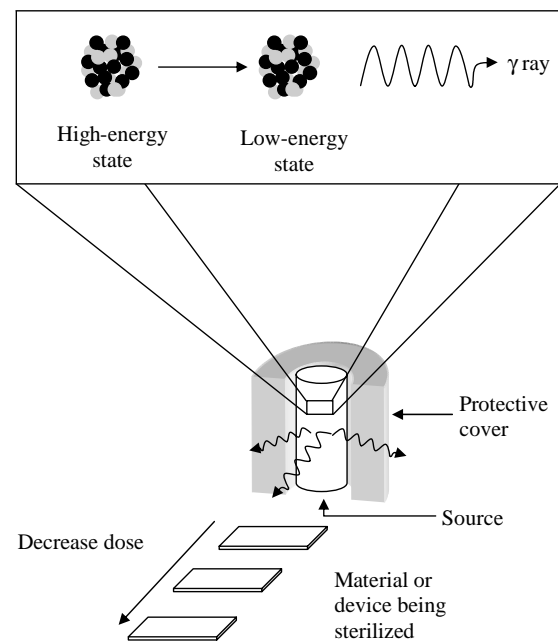


Figure 4. Schematic representation of a gamma source and how devices are typically sterilized. Dosage may be adjusted by calibrating the distance between the source and the device. Other forms may be employed depending on the facilities and the intended application.

Advantages And Disadvantages. The advantages of gamma irradiation include its compatibility with many materials, the minimal amount of toxic residues, and the low temperature requirements. Disadvantages of gamma irradiation include its cost, complexity (e.g., requires highly specialized facilities), and incompatibility with some polymers due to degradation (e.g., polymethylmethacrylamide and cellulose derivatives) or cross-linking (e.g., polyethylene and polystyrene) of the polymer (7). Gamma irradiation also may adversely affect proteins in a dosage dependent manner by the introduction of free radicals and has been associated with reduction in strength of grafts and collagenous biomaterials (26–30).

Physical Effect Upon Materials. Ionizing radiation affects the backbone of synthetic and natural polymers and may cause cross-linking, oxidation, or chain scission (7,24,25). These changes may affect degradation kinetics and material properties (e.g., tensile strength, elastic modulus, elongation, and the color of the material). In general, polymeric materials, including biologic polymers, can be safely sterilized via gamma irradiation. Safe levels of gamma irradiation include a range of 2.5 Mrads (collagen-based materials, ECM scaffolds, polypropylene, polytetrafluoroethylene) to 1000 Mrads (polystyrene and epoxy resins) (7). The lower doses of gamma irradiation (e.g., 2.5 Mrads) effectively sterilize biologic materials with minimal adverse effects to the physical properties of the material.

e-Beam Irradiation

Mechanism of Action. e-Beam irradiation involves the generation of a beam of electrons from a linear accelerator. The beam of electrons can be manipulated to achieve the desired dose by varying power and the exposure time. The sterilizing effect of e-beam irradiation on polymeric materials is similar to that described for gamma irradiation except for less penetrability of the beam into target materials. The main effect of e-beam irradiation comes from the ionizing electrons that result in the destruction of nucleic acids and proteins required for cellular processes. One primary difference between gamma radiation and e-beam is that the electron beam can be focused on the target

material and turned off when the sterilization process is completed.

Advantages And Disadvantages. The advantages of e-beam irradiation compared to gamma irradiation include the fast sterilization time, less safety concerns (i.e., the source can be turned off), greater power, and the source does not deplete as occurs with the source of gamma irradiation. The disadvantages of e-beam irradiation include its cost and less penetrability than gamma irradiation. As in the case with gamma irradiation, electron beam irradiation has also been associated with a decrease in strength of collagenous materials (27–30).

Physical Effect Upon Materials. The overall effect of e-beam irradiation is similar to those described for gamma irradiation. Free radicals are created following exposure to the electron beam and if oxygen is present, peroxy free radicals can be formed that increase the rate of chain scission (25). Polylactic-co-glycolic acid (PLGA) exposed to e-beam shows a decrease in the molecular weight (M_w) and thermal properties mainly due to chain scission and crosslinking (31,32). e-Beam has also been shown to reduce the degradation rate of PVC and polypropylene (PP) when compared to gamma sterilization, mainly due to faster free radical termination reactions (25).

The effects of e-beam irradiation upon naturally occurring, biologic scaffold materials is dose dependent. In the lower dose range (~ 2.5 Mrad), there are minimal changes upon the physical and biologic properties of the irradiated materials (24). However, changes in strength (33,34) and degradability (35) of collagenous materials and the creation of cross-links may still occur as a result of e-beam exposure (33,34).

Table 3 summarizes some of the naturally occurring biomaterials currently available and the method used to terminally sterilize the device. Table 4 summarizes the advantages, disadvantages, and the typical doses for the sterilization methods discussed.

ALTERNATIVE METHODS OF STERILIZATION

Alternative approved methods of sterilization that are recognized by the FDA include liquid chemical sterilants,

Table 3. Example of Currently Marketed Biologic Devices and the Form of Terminal Sterilization Employed

Product	Company	Source	Sterilization
CuffPatch™	Arthrotech, Inc.	Crosslinked porcine small intestine	Gamma
Pelvicol™	Bard, Inc.	Crosslinked porcine dermal collagen	Gamma
Bard® Dermal Glograft	Bard, Inc.	Cadaveric dermis minus epidermal layer	Gamma
FasLata® Allograft	Bard, Inc.	Cadaveric fascia lata	Gamma
OaSis®	Cook Biotech, Inc.	Porcine small intestine	ETO
Stratasis®	Cook Bioetch, Inc.	Porcine small intestine	ETO
Surgisis®	Cook Biotech, Inc.	Porcine small intestine	ETO
Restore™	Depuy Orthopedics, Inc.	Porcine small intestine	Gamma
TissueMend®	TEI Bioscience, Inc.	Fetal bovine skin	ETO
Permacol™	Tissue Science Laboratories, Inc.	Porcine dermis	Gamma
Tutopatch®	Tutogen Medical, Inc.	Bovine pericardium	Gamma
Tutoplast®	Tutogen Medical, Inc.	Human fascia lata	Gamma

Table 4. Summary of the Most Commonly Used Sterilization Methods and Their Advantages, Disadvantages, and Typical Doses as Discussed in this Article

Sterilization Method	Advantages	Disadvantages	Typical Dose
Dry heat	Inexpensive; Ease of use	Thermal damage to proteins	Dry air at 120–170 °C (250–338 °F) for 1–18 h
Moist heat	Faster than dry heat and less heat requirement; Inactivates some viruses	Thermal damage to proteins	Saturated steam at 121–125 °C for 15–30 min (pressure may be adjusted)
Ethylene oxide	Lower temperatures than heat sterilization and less moisture	Toxic by-products; Potential reactions with functional groups	ETO is added at 600–1200 mg·L ⁻¹ for 2–48 h (40–90% humidity at 40–50 °C)
Ionizing radiation	Minimal thermal damage; Good penetration	Cost; Changes in material properties	6–24 kGy

high intensity light, ultraviolet light, vapor systems (combination of hydrogen peroxide and peracetic acid), exposure to chlorine dioxide, and filtration methods (510 k Sterility Review Guidance K90-1). In addition, low temperature gas plasma and machine-generated X rays can be used for sterilization purposes.

Chemicals are often used to reduce the bioburden, disinfect, and sterilize biologic materials. Table 5 lists a few of the most commonly used chemicals for the disinfection and/or sterilization of grafts and naturally occurring biomaterials. Glutaraldehyde is currently used to sterilize and remove the antigenicity of many porcine-derived products (e.g., heart valves and pericardium). Formaldehyde and a combination of formaldehyde and low temperature steam

is another form of sterilization used for medical devices. However, this method is seldom used for biologic materials due to its high toxicity.

Extracellular matrix based bioscaffolds (e.g., Restore, DePuy Orthopaedics) and OaSis[®] (Cook Biotech, Inc.)TM undergo presterilization disinfection with a peracetic acid and ethanol treatment followed by water and phosphate buffered saline washes to remove any chemical residues. Table 6 summarizes a typical disinfection method used for naturally occurring biomaterials prior to sterilization. Collagen-based products (e.g., Contigen, Bard, Inc.) rely on sterile processing techniques, acidic environments, chemical cross-linking (e.g., glutaraldehyde), and sterile filtration (e.g., via 0.22 μ filters) to achieve the desired sterility.

Table 5. Common Chemical Sterilants and Disinfectants Used for Biologic Materials

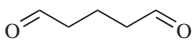
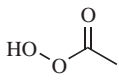

Chemical	Uses	References
 Glutaraldehyde	Used for sterilization and cross-linking of tissues (e.g., heart valves, pericardium, tendons, and collagen based biomaterials).	(1,36–39)
 Peracetic acid	Used to disinfect ECM derived bioscaffolds and disinfect and sterilize medical devices. It is usually combined with hydrogen peroxide and ethanol.	(1,36,40–47)
HO—OH Hydrogen peroxide	Used to sterilize grafts and medical devices. It is usually combined with steam sterilization or with other sterilants such as peracetic acid.	(1,23,48–52)
 Ethanol	Used to disinfect ECM derived bioscaffolds and grafts. It is usually combined with peracetic acid and it is also used to disinfect or sterilize instruments.	(1,36,39,45,53–55)

Table 6. Example of the Disinfection Process Used to Reduce the Bioburden of Naturally Occurring Biomaterials^a

Step	Solution	Time
Peracetic acid disinfection	0.1% (v/v) peracetic acid 4% (v/v) ethanol, and 95.9% (v/v) sterile water	2 h
Phosphate buffer saline wash	Sterile 1X PBS pH 7.4	30 min
Water wash	Sterile water	30 min

^aSee Ref. 56.

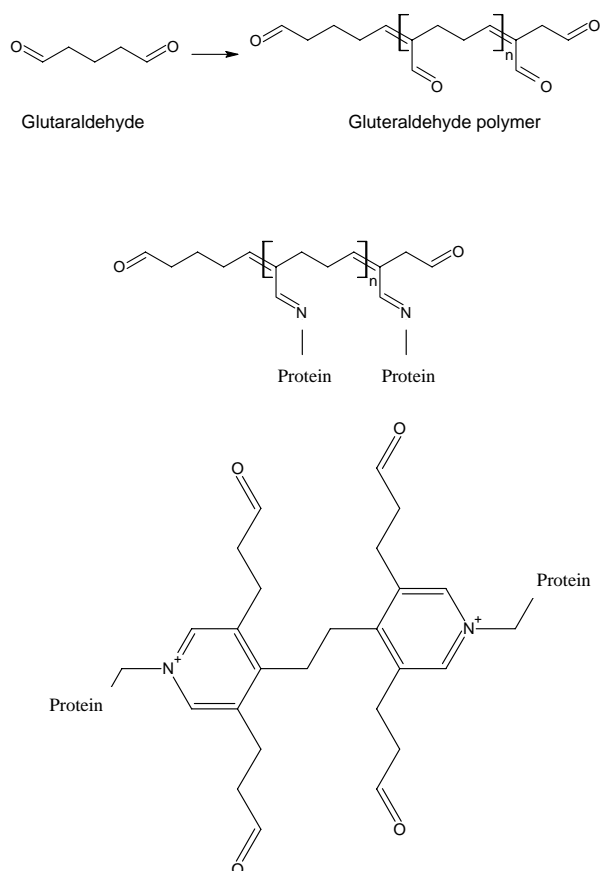


Figure 5. Reaction of glutaraldehyde polymers with amino groups and a pyridinium cross-link resulting from protein exposure to glutaraldehyde. (Redrawn from Ref. 1).

High levels of peracetic acid and combinations of peracetic acid (PAA), ethanol, isopropyl alcohol, hydrochloric acid, antibiotics, hydrogen peroxide, and detergents have been used for sterilization of demineralized bone grafts (48). A combination of steam sterilization and hydrogen peroxide may also be used for sterilizing allogeneic bone, but risks of denaturing collagen molecules still exists. Peracetic acid and hydrogen peroxide have been used for the disinfection and sterilization of acellular dermal matrices (24). The negative effects of these alternative sterilization methods upon biological materials include the molecular cross-linking that is caused by the chemical sterilant (see Fig. 5), collagen and protein denaturation, and the adverse effects of acids (e.g., the disruption of sulfhydryl and sulfur bonds by PAA) upon material properties at the concentrations needed for sterilization.

SUMMARY

There are several options for the sterilization of biologic materials and each option has its advantages and disadvantages. The majority of biologic materials classified as medical devices are sterilized by either gamma or e-beam

irradiation or by ethylene oxide. These methods tend to have dose-dependent effects upon the structure and function of biologic materials (e.g., strength of irradiated materials), but doses that effectively sterilize the material can be used with minimal effects upon the structure and function of currently marketed biologic devices.

GLOSSARY

- | | |
|----------------------------|---|
| Allograft | Organ or tissue graft obtained from the same species. |
| Bioburden | The microbiologic load or the number of contaminating organisms in the product before sterilization. |
| Disinfection | The destruction of pathogenic and nonpathogenic microorganisms by thermal or chemical means. The FDA (1997) definitions as given in Ref. 1. |
| Endotoxins | Toxic components of the outer membrane of Gram-negative bacteria that cause fever. |
| Extracellular matrix (ECM) | An insoluble network of polysaccharides and structural and non-structural proteins secreted by host cells and present in all tissues and organs. |
| Inactivation | Removal or inactivation of the activity of microorganisms by killing or inhibiting reproductive or enzymatic activity. The FDA (1997) definitions as given in Ref. 1. |
| Medical device | A device used for diagnosis, cure, treatment, or prevention of a disease or condition. A device that affects the structure and function of the body, does not achieve intended use through chemical reaction, and it is not metabolized. The FDA (1997) definitions as given in Ref. 1. |
| Pyrogens | Fever inducing agents (e.g., signals from inflammatory cells, chemicals, endotoxins, cell remnants from Gram-positive bacteria, and fungi). |
| Spore | The dormant state of an organism, typically a bacterium or fungus, which exhibits a lack of biosynthetic activity and reduced respiratory activity, and has resistance to heat, radiation, desiccation, and various chemical agents. The FDA (1997) definitions as given in Ref. 1. |

Sterility assurance level (SAL)	An indicator that no greater than the predetermined number of microorganisms exists in a product usually represented as the number of colony forming units per device.
Sterilization	A process intended to remove or destroy all viable forms of microbial life, including bacterial spores.
Xenograft	Organ or tissue graft obtained from a different species.

BIBLIOGRAPHY

- Block SS. Disinfection, Sterilization, and Preservation. 5th ed. Philadelphia (PA): Lippincott Williams & Wilkins; 2001. p 1481.
- Ratner BD. Biomaterials science: An introduction to materials in medicine. 2nd ed. 1996, San Diego: Academic Press; 1996. p 484.
- Baird RM, Hodges NA, Denyer SP. Handbook of Microbial Quality Control. Pharmaceuticals and Medical Devices. New York: Taylor & Francis; 2000.
- Halls NA. Achieving Sterility in Medical and Pharmaceutical Products. Vol. 64. New York: Marcel Dekker, Inc.; 1994.
- Perkins JJ. Principles and Methods of Sterilization in Health Sciences. 2nd ed. Springfield (IL): Thomas; 1969. p 560.
- Pruss A, et al. Comparison of the efficacy of virus inactivation methods in allogeneic avital bone tissue transplants. Cell Tissue Bank 2001;2(4):201–215.
- Bronzino JD. The Biomedical Engineering Handbook. Boca Raton (FL): CRC Press: IEEE Press; 1995. p 591.
- Chen SS, Humphrey JD. Heat-induced changes in the mechanics of a collagenous tissue: Pseudoelastic behavior at 37 degrees C. J Biomech 1998;31(3):211–216.
- Jun JH, et al. Effect of thermal damage and biaxial loading on the optical properties of a collagenous tissue. J Biomech Eng 2003;125(4):540–548.
- Harris JL, Humphrey JD. Kinetics of thermal damage to a collagenous membrane under biaxial isotonic loading. IEEE Trans Biomed Eng 2004;51(2):371–379.
- Mazzu AL, Smith CP. Determination of extractable methylene dianiline in thermoplastic polyurethanes by HPLC. J Biomed Mater Res 1984;18(8):961–968.
- Habermann V, Waitzova D. On the safety evaluation of extracts from synthetic polymers used in medicine. Arch Toxicol (Suppl) 1985. 8:458–460.
- Berger K, Sauvage LR. Late fiber deterioration in Dacron arterial grafts. Ann Surg 1981;193(4):477–491.
- Speirs AD, et al. Biomechanical properties of sterilized human auditory ossicles. J Biomech 1999;32(5):485–491.
- Prolo DJ, Pedrotti PW, White DH. Ethylene oxide sterilization of bone, dura mater, and fascia lata for human transplantation. Neurosurgery 1980;6(5):529–539.
- Fraenkel-Conrat H. The action of 1,2-epoxides on proteins. J Biol Chem 1944;154:227–249.
- Gad SC, ebrary Inc. Safety Evaluation of Medical Devices. 2nd ed. New York: Marcel Dekker; 2002. p 558.
- Star EG. [The toxic effect of ethylene chlorohydrin and ethylene glycol on experimental animals and human cell cultures (author's transl)]. Bakteriell Mikrobiol Hyg [B] 1980;171(1): 25–32.
- Parker RE, Isaacs NS. Mechanisms of epoxide reactions. Chem Rev 1959;59:737–797.
- Guidoin R, et al. A compound arterial prosthesis: the importance of the sterilization procedure on the healing and stability of albuminated polyester grafts. Biomaterials 1985; 6(2):122–128.
- Kearney JN, et al. Evaluation of ethylene oxide sterilization of tissue implants. J Hosp Infect 1989;13(1):71–80.
- Kudryk VL, et al. Toxic effect of ethylene-oxide-sterilized freeze-dried bone allograft on human gingival fibroblasts. J Biomed Mater Res 1992;26(11):1477–1488.
- Thoren K, Aspenberg P. Ethylene oxide sterilization impairs allograft incorporation in a conduction chamber. Clin Orthop Relat Res 1995;318:259–264.
- Gaughran ERL, Goudie AJ, Johnson and Johnson Inc. Sterilization of medical products by ionizing radiation: International conference. Vienna, Austria, April 25–28, 1977. Sterilization by Ionizing Radiation. Vol. 2. Montreal: Multiscience Publication; 1978. p 408.
- Woo L, Purohit KS. Advancements and opportunities in sterilisation. Med Device Technol 2002;13(2):12–17.
- Olde Damink LH, et al. Influence of ethylene oxide gas treatment on the in vitro degradation behavior of dermal sheep collagen. J Biomed Mater Res 1995;29(2): 149–155.
- De Deyne P, Haut RC. Some effects of gamma irradiation on patellar tendon allografts. Connect Tissue Res 1991;27(1): 51–62.
- Roe SC, et al. The effect of gamma irradiation on a xenograft tendon bioprosthesis. Clin Mater 1992;9(3–4):149–154.
- Yahia LH, Drouin G, Zukor D. The irradiation effect on the initial mechanical properties of meniscal grafts. Biomed Mater Eng 1993;3(4):211–221.
- Godette GA, Kopta JA, Egle DM. Biomechanical effects of gamma irradiation on fresh frozen allografts in vivo. Orthopedics 1996;19(8):649–653.
- Loo JS, Ooi CP, Boey FY. Degradation of poly(lactide-co-glycolide) (PLGA) and poly(L-lactide) (PLLA) by electron beam radiation. Biomaterials 2005;26(12):1359–1367.
- Loo SC, Ooi CP, Boey YC. Influence of electron-beam radiation on the hydrolytic degradation behaviour of poly(lactide-co-glycolide) (PLGA). Biomaterials 2005;26(18):3809–3817.
- Chuck RS, et al. Biomechanical characterization of human amniotic membrane preparations for ocular surface reconstruction. Ophthalmic Res 2004;36(6):341–348.
- Fujisato T, et al. Cross-linking of amniotic membranes. J Biomater Sci Polym Ed 1999;10(11):1171–1181.
- Grimes M, Pembroke JT, McGloughlin T. The effect of choice of sterilisation method on the biocompatibility and biodegradability of SIS (small intestinal submucosa). Biomed Mater Eng 2005;15(1–2):65–71.
- Sprossig M, et al. [Sterilization of biologic heart valve prostheses with glutardialdehyde]. Z Exp Chir 1973;6(4):248–251.
- Wallace RB. Tissue valves. Am J Cardiol 1975;35(6):866–871.
- Munting E, et al. Effect of sterilization on osteoinduction. Comparison of five methods in demineralized rat bone. Acta Orthop Scand 1988;59(1):34–38.
- Sung HW, Hsu HL, Hsu CS. Effects of various chemical sterilization methods on the crosslinking and enzymatic

- degradation characteristics of an epoxy-fixed biological tissue. *J Biomed Mater Res* 1997;37(3):376–383.
40. Mucke H, Wenzel KP. [Preparation of heart valves for grafting after sterilization with peracetic acid]. *Z Exp Chir* 1973;6(4):252–255.
 41. Sprossig M, et al. [Sterilization of heart valve transplants with peracetic acid]. *Helv Chir Acta* 1973;40(3):357–362.
 42. Wutzler P, et al. [Combined cleaning and cold sterilization procedure for cleaning rooms in virological establishments]. *Z Med Labortech* 1975;16(5):253–259.
 43. Wenzel KP. [Final sterilization of formaldehyde-preserved bioprostheses using peracetic acid]. *Z Exp Chir* 1982; 15(4):261–263.
 44. Lomas RJ, et al. Assessment of the biological properties of human split skin allografts disinfected with peracetic acid and preserved in glycerol. *Burns* 2003;29(6):515–525.
 45. Pruss A, et al. Peracetic acid-ethanol treatment of allogeneic avital bone tissue transplants—a reliable sterilization method. *Ann Transplant* 2003;8(2):34–42.
 46. Huang Q, et al. Use of peracetic acid to sterilize human donor skin for production of acellular dermal matrices for clinical use. *Wound Repair Regen* 2004;12(3):276–287.
 47. Lomas RJ, et al. Effects of a peracetic acid disinfection protocol on the biocompatibility and biomechanical properties of human patellar tendon allografts. *Cell Tissue Bank* 2004; 5(3):149–160.
 48. Glowacki J. A review of osteoinductive testing methods and sterilization processes for demineralized bone. *Cell Tissue Bank* 2005;6(1):3–12.
 49. Brown SA, et al. Effects of different disinfection and sterilization methods on tensile strength of materials used for single-use devices. *Biomed Instrum Technol* 2002;36(1):23–27.
 50. Lambert RJ, Johnston MD, Simons EA. A kinetic study of the effect of hydrogen peroxide and peracetic acid against *Staphylococcus aureus* and *Pseudomonas aeruginosa* using the bioscreen disinfection method. *J Appl Microbiol* 1999;87(5):782–786.
 51. Rutala WA. Disinfection and sterilization of patient-care items. *Infect Control Hosp Epidemiol* 1996;17(6):377–384.
 52. Rutala WA, Weber DJ. Disinfection of endoscopes: review of new chemical sterilants used for high-level disinfection. *Infect Control Hosp Epidemiol* 1999;20(1):69–76.
 53. Pruss A, et al. Validation of the sterilization procedure of allogeneic avital bone transplants using peracetic acid-ethanol. *Biologicals* 2001;29(2):59–66.
 54. Hodde J, Hiles M. Virus safety of a porcine-derived medical device: evaluation of a viral inactivation method. *Biotechnol Bioeng* 2002;79(2):211–216.
 55. Scheffler SU, et al. Biomechanical comparison of human bone-patellar tendon-bone grafts after sterilization with peracetic acid ethanol. *Cell Tissue Bank* 2005;6(2):109–115.
 56. Freytes DO, et al. Biaxial strength of multilaminated extracellular matrix scaffolds. *Biomaterials* 2004;25(12):2353–2361.

See also BIOMATERIALS FOR DENTISTRY; BIOMATERIALS: TISSUE ENGINEERING AND SCAFFOLDS; ENGINEERED TISSUE.

STETHOSCOPE. See LUNG SOUNDS.

STOMACH, ELECTRICAL ACTIVITY. See ELECTROGASTROGRAM.

STRAIN GAGES

PAUL C. DECHOW
QIAN WANG
A & M University Health
Science Center
Dallas, Texas

INTRODUCTION

Strain gages are devices that allow measurement of the change in the dimensions, displacement, or deformation of an object. These devices have been used extensively in a wide variety of engineering applications over the past century and, in recent decades, their use in biomedical applications, such as in the production of transducers for biomedical monitoring and research, has increased tremendously. These transducers measure a variety of parameters, including strain, displacement, pressure, acceleration, force, and temperature. The chief types of strain gages are mechanical, optical, acoustical, and electrical. There are several types of electrical gages including capacitance, inductance, semiconductor, and resistance strain gages. The electrical resistance strain gages are the most widely used types of gages in engineering and biomedical applications today and receive the most extensive treatment in this article. Before discussing the various types of strain gages in more detail, it is useful to have a working understating of what strain and stress are and their relationship to each other.

STRAIN AND STRESS

Strain is a dimensionless unit, which is defined as the ratio of the change in unit length over the original length,

$$\epsilon = \Delta L/L \quad (1)$$

where ϵ is strain, L is the original length of the object along an axis, and ΔL is the change in length along that axis. Measurements of strain are usually given in microstrain ($\mu\epsilon$) where $1.0 \mu\epsilon = 1 \times 10^{-6} \epsilon$. By convention, if an object is shortened in length, the strain is compressive and takes a negative value. If an object is lengthened, the strain is tensile and takes a positive value.

Although strain gages directly measure a change in electrical resistance that is proportional to a change in dimension, they can also be used to measure force or pressure if some of the properties of the test material are known. For many metals and other solids, a constant can be used to define a linear relationship between deformation or strain (ϵ) and stress (σ), where stress is defined

as force per unit area. This relationship, also known as Hooke's law, can be expressed as

$$\sigma = E\varepsilon \quad (2)$$

Where stress (σ) is directly proportional to strain (ε). The constant E is called the modulus of elasticity, the elastic modulus, or Young's modulus. Since strain is a dimensionless quantity and stress is defined in units of force per area, such as pascals (newtons per meter squared), E is also given in units of force per area.

Strain gages usually measure strain on a small portion of the surface area of a structure. A series of strain gages affixed to a structure allows determination of the way in which the structure is being deformed; the deformation can result from either axial, shearing, twisting, or bending loads. If a structure is loaded so that it is deformed in a predictable way, then strain gages can be affixed to the structure, calibrated, and used to measure the magnitude of the loading. An example of this kind of transducer would be a load cell, which can be used to measure force when load axially.

TYPES OF STRAIN GAGE

Depending on the material and the test situation, the parameters of strain measurement vary widely and require selection of a strain gage appropriate for the particular problem. Considerations in the selection of a gage include factors, such as knowledge of the required accuracy and stability of the gage, the maximum deformation of the material, the duration of the test, patterns and amount of loading of the gage, and the location and situation of gage installation. The following discussion describes some features and limitations of strain gages that are currently in use. Due to their widespread use, greater consideration is given to electrical resistance strain gages.

Mechanical Strain Gages

Mechanical strain gages have been in use longer than other types. However, due to their large size and relative inaccuracy, their use in engineering applications today is limited. These gages are most commonly used to measure strain in industrial applications where it is appropriate to summate strain over a range of several inches. Mechanical strain gages typically consist of a system of two points or knife-edges that can be securely attached to a structure. Then a series of compound levers within the gage magnify the displacement allowing a reading to be taken. The limited range of accuracy of these gages, with measurements restricted to magnifications of up to 2000 ($500 \mu\varepsilon$), contrast with the greater accuracy of electrical resistance strain gages (see below). Yet such gages are most useful where the size of the gage is not an issue and the ability to take a reading from a simple vernier or dial scale, without any associated electric instrumentation, is at a premium.

An additional type of mechanical strain gage, the electromechanical strain gage or extensometer, provides greater accuracy than other types of mechanical strain gages and is used in a variety of current industrial applications and in materials testing.

Optical Strain Gages

Optical techniques have been used in a variety of ways to measure strain. The simplest optical strain gages are similar to mechanical strain gages except that light rays are substituted for mechanical levers in magnifying the displacement. This change serves to decrease the size and inertia of the gage while making it appropriate for use at low frequencies in dynamic applications. The use of lasers as collimated light sources has led to the development of several optical strain gages, including those based on principles of diffraction. The diffraction strain gage, like some of the mechanical gages, has two blade-like edges that are bonded or welded to the test specimen. This gives the gage the advantage of automatic temperature compensation, if the blades are constructed of the same material as the test specimen, making the gage suitable for some use in extreme temperatures.

Another interesting optical gage is the interferometric strain gage. This gage measures strain by examining interference patterns caused by directing a light source, such as a helium-neon laser, at two V-grooves placed on the surface of a specimen. This method is most useful in test situations where it is preferable not to actually attach a strain gage to the test specimen, thus eliminating problems of bonding, inertia of the gage, and temperature compensation. A similar noncontacting device is the infrared (IR) extensometer, which is used in materials testing when deformations are large, such as with elastomeric and highly extensible materials, where strain gages or contracting extensometers cannot be readily attached. Resolutions of up to $5 \mu\text{m}$ can be obtained with these devices.

Electrical Strain Gages: Capacitance and Inductance Types

These two types of electrical strain gages are not as widely used as the semiconductor and resistance types, yet find uses in specialized applications and in the production of transducers. The capacitance strain gage uses a parallel plate capacitor where the positional relationship of the two plates varies with the capacitance. For example, increasing the distance between the two plates will effect such a change.

There are several types of inductance strain gages. One notable example is the linear variable differential transformer (LVDT). This device is useful for measuring displacements that range from several micrometers to several centimeters, thus making it useful in situations that require measurements of larger displacements than can be measured by electrical resistance strain gages. The device can also be modified with auxiliary mechanisms to measure velocity, acceleration, force, pressure, or flow rate. The LVDT consists of a freely moving iron core surrounded by a primary and two identical secondary coils. The device can be designed such that the voltage output is a linear function of the displacement of the core within the coils over a specific range.

Electrical Strain Gages: Semiconductor Type

Semiconductor strain gages are widely used gages, especially in the production of load cells, miniaturized

transducers, and other applications requiring measurement of very small strains. These gages provide a large signal output relative to strain, as indicated by their large gage factors (GF for a definition see below), which can be as high as 200. By contrast, GFs for electrical resistance strain gage are usually two or less. Unfortunately, semiconductor strain gages also suffer from an extreme sensitivity to temperature. In addition, the piezoresistive effect varies with strain giving these gages ranges of nonlinearity. Semiconductor gages are typically constructed from silicon or germanium. Because of the high receptivity of these materials, the gages consist of a small filament of a single crystal of the semiconductor material mounted on some type of carrier. The receptivity and strain-measuring properties of the crystal can be varied by altering the amount of impurities in the crystal. The fatigue life of semiconductor gages for cyclic strains is less than that of electrical resistance-type gages. These gages are then commonly employed in fields with low strains where frequent loading does not lead to gage failure.

Electrical Strain Gages: Resistance Type

The electrical resistance strain gages are the most widely used gages in engineering and biomedical fields (Fig. 1). The principles behind these gages were first discovered by Lord Kelvin in 1856 when he noted (1) that the resistance of a metal wire increases with increasing strain and decreases with decreasing strain and (2) that different materials have different sensitivities to strain.

Electrical resistance strain gages are also sensitive to temperature changes, but not to the extent of semiconductor gages. Most gages have a several hundred degree range under which they function best. However, within that range, temperatures must be monitored carefully in order to compensate for apparent strain caused by shifts in temperature. However, gages are available that have self-temperature compensation. In these gages, the

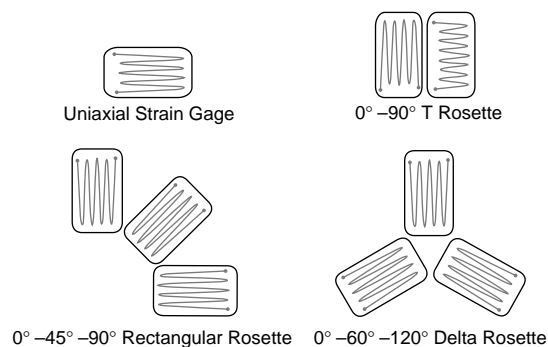


Figure 1. Four basic gage patterns (From eFunda Inc., with permission). Top left: single-element gage. This gage can be used to measure either maximum or minimum strains. Top right: T-style gage with two perpendicular elements. This gage can be used to measure maximum, minimum, and shear strains if the gage is aligned with the principle strains. Bottom left: rectangular rosette strain gage. This gage, as well as the delta rosette strain gage (bottom right), can be used to measure maximum, minimum, and shear strains and need not be aligned with the principal strains.

thermal expansion coefficient of the gage is matched to that of the test material.

Electrical resistance strain gages are available on a variety of backing materials ranging from various metal foils to polyamide and epoxy. The backing materials affect the performance of the gage and are varied to achieve the desired temperature range, extensibility, and fatigue characteristics for the gage. Many electrical resistance strain gages fail only when strain levels exceed 20,000 $\mu\epsilon$ (2%) and one million cycles, making them useful for many static and dynamic applications.

An important characteristic of an electrical resistance strain gage is the gage factor (GF). This factor can be defined as follows:

$$GF = (\Delta R/R)/\epsilon \quad (3)$$

$\Delta R/R$ is the change in resistance divided by the resistance, which is then divided by the strain (ϵ) in order to calculate GF. Thus, GF is a dimensionless constant for any given gage and a measure of sensitivity. Its value is affected by the pattern of the foil, the size of the gage, the geometry of the gage, and the temperature at which measurements are made. Commercially available gages come with calculated GFs throughout their temperature range as well as values for their resistance. A reworking of equation 3 can be used then to calculate ϵ if $\Delta R/R$ is measured. The parameter $\Delta R/R$ can be easily measured through the use of a Wheatstone bridge circuit or potentiometer, as discussed below. Typical, GFs of electrical resistance strain gages are ~ 2 , indicating the low sensitivity of these gages compared to semiconductor strain gages, which have high GFs. As a result, signals from electrical resistance strain gages require extensive conditioning to amplify the signal to a level acceptable to most recording devices.

A rosette strain gage is a combination of three electrical resistance strain gages configured adjacent to each other or stacked on top of one another on a single backing (Fig. 1). Rosettes can be used to calculate the direction and magnitude of the principal strains (minimum and maximum strain), and shear strain on a test material. They are available in two configurations: (1) a delta configuration with the principal axes of the three gages oriented at 60° apart and (2) a rectangular configuration with the principal axes of the three gages oriented at 45° apart. Another gage configuration, the T configuration (Fig. 1), consists of two gages at right angles to each other. This gage can also be used to calculate minimum strain, maximum strain, and shear strain if the directions of the minimum and maximum strains are already known and the elements of the gage are oriented along these axes on the test material.

A wide variety of insulation materials and bonding substances are available for use with electrical resistance strain gages. These include appropriate substances for constructing preparations for various conditions of temperature and moisture for static and dynamic applications.

Electrical Strain Gages: Elastic Resistance Type

Elastic resistance strain gages are specialized types of electrical resistance strain gages that are used in

cardiovascular and respiratory dimensional and plethysmographic (volume-measuring) determinations. The gages are made of a narrow silicone-rubber tubes, usually with an inside diameter of ~ 0.5 mm, and they range in length from 3 to 25 cm. The gages are filled with mercury or with an electrolyte or conductive paste and the ends are sealed with copper, silver, or platinum electrodes. When the gage is stretched, the diameter of the tube decreases and the length increases, increasing the resistance. These gages allow measurement of large-dimensional changes. They are accurate in the 10,000–100,000 $\mu\epsilon$ range. Strains as high as 300,000 $\mu\epsilon$ (30%) can be measured with distortion as small as 4%.

Wireless Strain Gage: Telemetry

Strain gages can be wireless, if coupled with telemetry transmitters and receivers. The elimination of the need for trailing wires from experimental objects makes possible data gathering from inconvenient or unsafe monitoring locations and it may reduce some of the noise of electrical interference (4,5).

SIGNAL PREPARATION AND AMPLIFICATION

Bridges

Two electric circuits, the potentiometer and the Wheatstone bridge, are commonly used to convert $\Delta R/R$ to a voltage that can be measured and used to determine ϵ . The Wheatstone bridge is the more widely used of the two types of circuits. There are some other variations; for further information, see the *Reading List*. A new circuit, the Anderson loop circuit, believed to outperform the Wheatstone bridge, shows much future promise.

Potentiometer. Figure 2 illustrates a potentiometer circuit (3). This circuit consists of a voltage source (E_i), two resistors (R_1 and R_2), and an output voltage (E_o). A strain gage can take the place of either or both of these resistors and standard circuit equations can be used to compute the changes in E_o and ΔE_o and thus $\Delta R/R$. These equations are not given here; for further information, see the *Reading List*.

A major limitation of the potentiometer circuit is that the output voltage E_o is usually quite large compared to

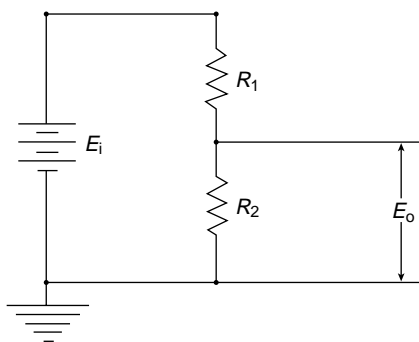


Figure 2. Diagram for a potentiometer circuit (3). Abbreviation: E_i -voltage source, R_1 and R_2 -resistors, E_o -output voltage.

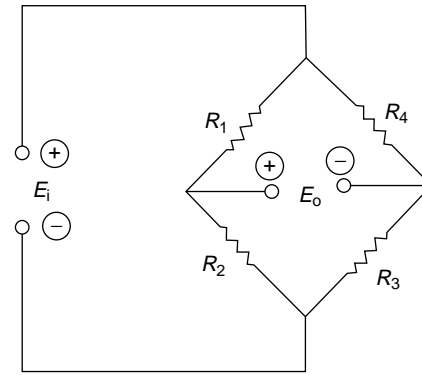


Figure 3. Diagram for a Wheatstone bridge circuit (3). Abbreviation: R_3 and R_4 -resistors, others as in Fig. 2.

ΔE_o , making it difficult to obtain an accurate measurement of ΔE_o . This limits the usefulness of the circuit for static loading applications. However, a filter can be employed to block out E_o while allowing voltage pulses of ΔE_o to be measured. Such a filter enables the circuit to be useful in dynamic applications.

The Wheatstone Bridge. Figure 3 illustrates a Wheatstone bridge circuit (1). Note that the circuit contains four resistors. Strain gages can be used in the place of any combination of these resistors depending on the desired measurement. The advantage of this circuit over the potentiometer is that the bridge can be balanced so that the output voltage (E_o) is zero. The ratio of ΔE_o to the input voltage (E_i) can then be used to calculate the change in resistance in the strain gage ($\Delta R/R$) and the strain (ϵ). This feature makes the Wheatstone bridge circuit useful for the measurement of both static and dynamic strains. The circuit equations used to calculate strain differ depending on the configuration of gages in the Wheatstone bridge. For additional information on these equations, see the *Reading List*.

A limitation of the Wheatstone bridge circuit is that it is nonlinear when ΔR is $> 1\%$ (2). This is generally not a problem within the usual range of resistance changes of electrical resistance strain gages. However, large strains ($> 3000 \mu\epsilon$) measured with semiconductor gages may require the use of different circuitry (3). An additional advantage of the Wheatstone bridge is that it allows signals to be added or subtracted in multiple gage installations. In fact, sensitivity of a transducer can be quadrupled by using four active gages instead of one in the application. This can be accomplished, for example, by placing the two strain gages associated with the negative arms of the Wheatstone bridge on the compressive surface of the test specimen, and the two strain gages associated with the positive arms if the Wheatstone bridge on the tensile surface. Temperature compensation can also be achieved with Wheatstone bridges, through the use of a dummy gage, without sacrificing sensitivity.

The Anderson Loop Circuit. Figure 4 illustrates the topology of the Anderson loop circuit. This circuit consists

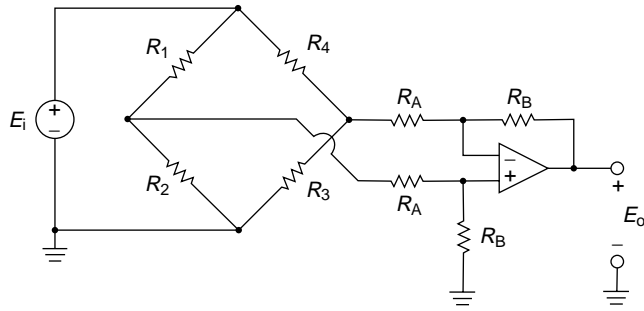


Figure 4. Diagram for a simple differential amplifier circuit for use with a Wheatstone Bridge (Modified from a drawing by O. Poroy (www.engineering.uiowa.edu/~bme080/lab/S05Lab7.pdf)). E_o is amplified by differential gain of sensors R_A and R_B compared to input resistance incurred by the source voltage (E_i).

of one or more sensors (strain gages) and typically one reference element (R_{ref}) connected in a series loop circuit. The same excitation current flows through each circuit element in the loop. The unique feature of this technology is the use of a subtractor function. In order to calculate strain from $\Delta R/R$, this function compares the voltage drop across various circuit elements in the loop, typically to determine the difference between each loop sensor voltage drop and the voltage drop across a reference element. Subtracting the voltage drop across two sensors yields a difference in $\Delta R/R$ between the two sensors.

Compared to the Wheatstone bridge, the Anderson loop circuit does not require four resistive elements. The loop provides a set of linear outputs that are twice that of the typical Wheatstone bridge for the same voltage across (and power dissipation in) each sensor. The loop topology overcomes some limitations normally encountered with the Wheatstone bridge, especially the detrimental effects of varying lead wire and connector resistances. These features open a vista for advanced transducer design. The impact of the Anderson loop circuit on the future of measurement and control is likely to be profound (6). This new technology was invented and developed in projects by NASA, who has placed it in the public domain. For more information, visit the Valid Measurements website, <http://www.v-m.us>.

Amplifiers

There are several good amplifiers and signal conditioners available commercially for use with strain gages (see *Further Reading* for some sources of information on specific amplifiers). An example for a simple differential amplifier circuit for use with a Wheatstone bridge is illustrated in Fig. 5, where the output voltage (E_o) is amplified by the including of sensors R_A and R_B . In general, the better amplifiers should have a variety of features including some of the following:

1. The ability to complete a variety of Wheatstone bridge circuits ranging up to a full bridge configuration with all resistors in the bridge replaced by strain gages. Dummy gages for a less than a full

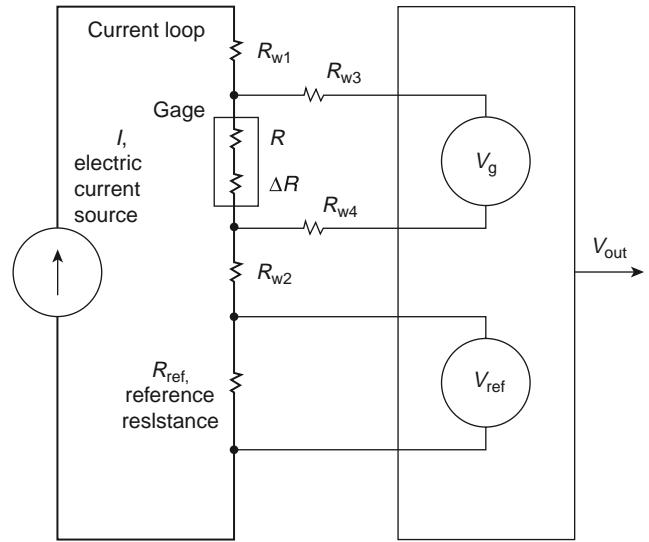


Figure 5. Diagram for an Anderson loop circuit (From Karl F. Anderson/Valid Measurements Inc., with permission)

1. bridge configuration should be available for several of the standard resistances used in strain gages (including 120 and 350 Ω resistors).
2. A range of excitation voltages should be available for the Wheatstone bridge circuits.
3. The amplifier should be able to balance the Wheatstone bridge over a wide range of output voltages, although some computer-linked devices now automate balancing, making this task transparent to the user.
4. Variable gain to allow accurate readings of both large and small strains with a variety of gages.
5. Wide band operation, high input impedance, low noise level, and low temperature coefficient.
6. The ability to condition not only electrical resistance strain gages installed as part of Wheatstone bridge circuits but also potentiometer circuits, and the ability to condition and amplify output from piezoresistive (semiconductor) gages.
7. Additional features, such as digital readout and filters, depending on specific applications.

APPLICATIONS

Strain gages have enabled significant advances in many biological and biomedical fields. They have been used extensively in a wide variety of biomedical and clinical applications, chiefly to measure strain, stress, pressure, force, or displacement in or produced by living structures, or *in vitro* simulations. It is the aim of the following section to document several of these applications in order to give the reader a practical knowledge of the possibilities for using strain gages in basic biological and biomechanical research as well as in clinical sciences and medical diagnosis.

Biomechanical Research and Bone Health

An area of investigation with speculations that extend back into the nineteenth century and before is the study of the relationship between function and loading in bone. Of great mechanical interest are parameters of bone morphology and physiology, such as bone shape and size and the rate and amount of skeletal remodeling. In 1892, Wolff summarized much previous research in his Law of Bone Transformation that stated that every change in the function of a bone results in changes in trabecular orientation and in the external shape of the bone (7). Use of electrical resistance strain gages has led to much new research in this area.

Strain gages were first used in the 1940s to measure bone strain (8,9). Rosette configurations of electrical resistance strain gages have been most useful in these types of studies since they can provide information about changes in both the direction and magnitude of the strains on the skeleton.

A variety of experimental studies have used strain gages to assess how the craniofacial skeleton is loaded during normal oral and masticatory activities. Early studies used skulls with strain gages bonded to them to assess these loads, strain patterns and their links to bone morphology by simulation of muscular and occlusal forces (10). Later advances in strain gage technology and biological experimental techniques have enabled strains to be measured directly from bone *in vivo* in experimental animals (11–13; Fig. 6). For example, studies have used electrical resistance strain gages to assess how the craniofacial skeleton is deformed during normal oral and masticatory activities in primates (14,15). These studies cannot be repeated on humans because of their invasive nature. However, the animal studies have led to speculations and inferences about the biomechanics of mastication in humans, that can be explored through *in vitro* experiments using human cadaver specimens (9,16). Other interesting cranial studies have used strain gages to understand the behavior of open sutures and their effects on the function, growth, and adaptation of craniofacial skeleton (17,18). Moreover, patterns of craniofacial bone fracture have been related to bone loading and strain distribution determined with strain gage techniques (19–21).

Other studies have used strain gages in the postcranial skeleton to study loading patterns in long bones during locomotion (22). These techniques have also been applied to study skeletal loading during human locomotion (23) and to obtain better understanding of the function of prosthetic devices such as hip replacements (24). Other postcranial studies have used strain gages to measure tension in tendons or in tendons and bone simultaneously (25,26). In general, experimental studies have demonstrated a relationship between magnitudes and rates of loadings in bone and skeletal remodeling (22,27), although the precise nature of this relationship is the subject of considerable debate (27–29).

Coupled with techniques of measuring bone material properties and finite element analysis, the application of strain gages in studies of functional morphology contributes greatly in medical and biological fields (30–34). This research provides tremendous knowledge about bone

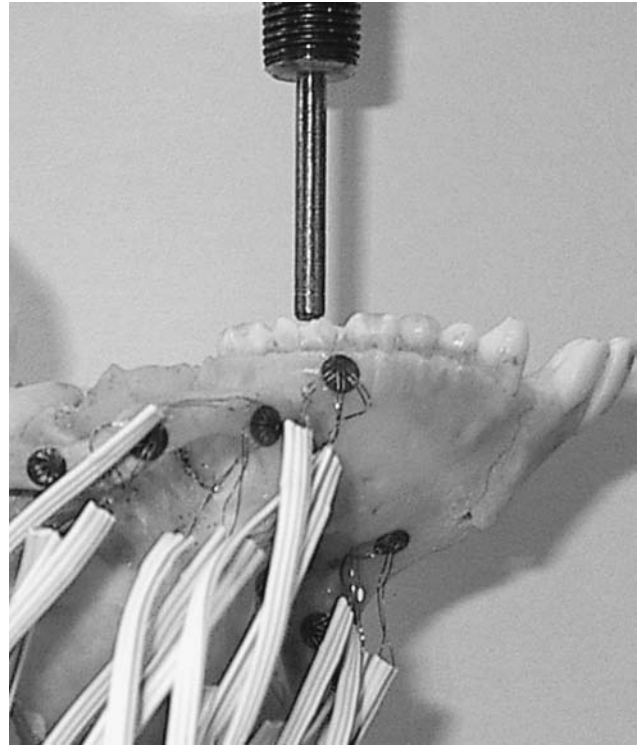


Figure 6. Experimental *In vitro* strain analysis in a skull of *Macaca fascicularis*. A couple of rectangular rosette strain gages were bonded to bone surface of craniofacial skeletons.

morphology, adaptation, and modeling and remodeling during aging, normal function, and under abnormal conditions, and provides rationales for bone health maintenance and bone reconstruction.

Dental and Craniofacial Research

In clinical dentistry and experimental craniofacial biology, measurements of occlusal forces have been used to give an overall assessment of functional capacities of the masticatory system (35–37). Such measurements are made with bite force transducers. These transducers have been used throughout much of the twentieth century and were originally mechanical devices. More recently, they have been constructed with electrical resistance strain gages or semiconductor gages. Figure 7 illustrates one such transducer constructed with electrical resistance strain gages (38). This transducer uses four strain gages in a full bridge configuration for an optimal voltage output. An individual would bite on the distal ends of the two bars to generate a change in resistance in the four gages. This change can then be measured and correlated with a specific biting force. Another feature of this transducer is that the gages are arranged in the circuit such that each upper and lower beam function as differential strain beams. In such beams, only the difference in strain between the two gages on the beam is measured. Thus force can be placed on the beam anywhere distal to the most distal gage and an identical reading will result. For a bite force transducer, this creates an advantage that the bite point does not have

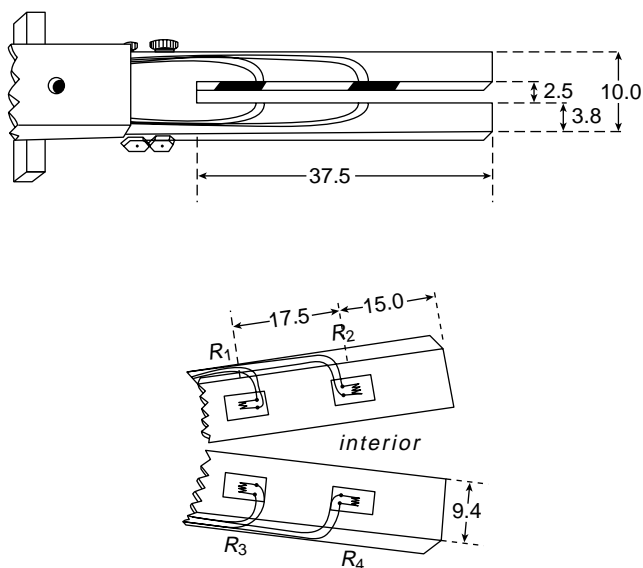


Figure 7. Diagram for a bite force transducer (38). This is an example of a force transducer constructed with four electrical resistance strain gages in a full bridge configuration. Each beam of the device is a differential strain beam. (a) Distal end of the device (missing the handle) illustrating the dimensions. (b) Inner surface of the beams showing the arrangement of the strain gages. Units in millimeters.

to be precisely controlled in order to obtain comparable readings.

Strain gages are used to study teeth and the surrounding alveolar bone during normal masticatory function, and in the presence of orthodontic and functional appliances. For example, strain gages are used to monitor the behavior of a dental prosthetic device (39–41), and its impact on the bone strain environment and eventual bone structure and strength (42). Strain gages are also used in studies of dental materials. A transducer has been constructed with electrical resistance gages to measure dimensional changes in dental amalgam during setting (43). These studies using strain gages provide rationales for orthodontic and prosthodontic treatments.

Physiological Research and Clinical Applications

Strain gages were used to study blood pressure (44), intracranial pressure (45), and cerebrospinal fluid pressure (46). The plethysmograph, made of elastic resistance strain gages, has been used diagnostically for several decades. For example, the venometer uses plethysmography to detect deep vein thrombosis through monitoring changes in certain muscle dimensions while venous outflow is occluded by a cuff placed around body parts (47,48).

A number of transducers constructed with electrical resistance strain gages are available to study the contractile properties of muscle. For example, strain gages were used to design quantitative assessment of neuromuscular deficits (49). Strain gages have also been used for many years to study cardiac muscle physiology *in vitro*. The

Walton–Brodie strain gage arch has been used since the 1950s to measure forces exerted by the heart (50). For example, studies, have used the Walton–Brodie arch along with other strain gage transducers to examine the effect of halothane and enflurane on right ventricular performance (51), antiarrhythmic drugs on the heart rate (52), and biventricular mechanical alternans (53).

Elastic resistance or mercury strain gages have been used in cardiovascular and respiratory, and urological research. These studies have examined, for instance, changes in the volume of the larynx and movements of the rib cage and abdomen during respiration (54), the spontaneous contractions of trachea (55), and the change of bladder pressure with the use of catheters (56). Electrical resistance strain gages are also used to make transducers to study gut motility (57,58).

Strain gages were used in pharmacological studies to test physiological reactions to certain drugs (59), in psychophysiological research to examine how body physiology changes with emotion (60) and pain (61), and in sleep studies to evaluate sleep quality by recording eye movement during sleep (62,63) and to diagnose sleep related syndromes such as sleep apnea (64).

A unique application of electrical resistance strain gages was used to construct a transducer to evaluate erectile function in men by using strain gages to measure or monitor penile circumference and rigidity on various conditions (65,66). For example, a transducer called tonometer, measured force at which the penis buckles, giving a quantitative measurements of penile tumescence (67).

Sports Medicine

Strain gages have had been used in sports medicine, such as for performance evaluation, protection, and healing. For example, electric resistive strain gages have been used to construct instruments to measure hand strength (68), and to measure shoulder strength in football players, which is the most vulnerable body part in this sport (69). Strain gage transducers were used to document elongation of the human anterior cruciate ligament during various activities following a sprain. The transducer was temporarily placed in the ligament during arthroscopic surgery (70). Strain gages have also been used to evaluate sports equipments. For example, experiments were conducted to reveal how shoe gear can affect tibial strains in humans during dynamic loading, such as treadmill walking and free running while wearing various sport shoes (71).

Robotic Medicine

Within the concept of robotic surgery and telemedical care, medical practice is significantly enhanced with robotic systems incorporating tactile sensors made of essential strain gage elements. Examples include a tactile sensor system using a piezoelectric transducer to simulate the properties of the human hand for use as a surgical support instrument for conducting micro- or telesurgery (72), or a palpation probe to evaluate patterns of softness and elasticity of human skin (73).

BIBLIOGRAPHY

1. Dechow PC. Strain gage. In: Webster JG, editor. Encyclopedia of Medical Devices and Instrumentation. New York: Wiley; 1988. p 2715–2721.
2. Shull LC. Basic circuits. In: Hannah RL, Reed SE, editors. Strain Gage Users' Handbook. London: Elsevier; 1992. p 79–132.
3. Dorsey J. Semiconductor strain gages. In: Hannah RL, Reed SE, editors. Strain Gage Users' Handbook. London: Elsevier; 1992. p 365–448.
4. Chadwick EK, Nicol AC, Floyd S, Gray TG. A telemetry-based device to determine the force-displacement behaviour of materials in high impact loading situations. *J Biomech* 2000; 33:361–365.
5. Schatzker J, Sumner-Smith G, Hoare J, McBroom R. A telemetric system for the strain gauge determination of strain in bone *in vivo*. *Arch Orthop Trauma Surg* 1980;96:309–311.
6. Anderson KF. The Anderson Loop: NASA's successor to the Wheatstone bridge. *ISA Transactions* 1997;36:351–356.
7. Wolff J. *Das Gesetz der transformation der knochen*. Berlin; 1892.
8. Gurdjian ES, Lissner HR. Mechanism of head injury as studied by the cathode ray oscilloscope. Preliminary report. *J Neurosurg* 1944;1:393–399.
9. Evans FG. Methods of studying the biomechanical significance of bone form. *Am J Phys Anthropol* 1953;11:413–436.
10. Endo B. Experimental studies on the mechanical significance of the form of the human facial skeleton. *J Fac Sci Univ Tokyo Sect 5* 1966;3:(Pt I).
11. Cochran GVB. A method for direct recording of electromechanical data from skeletal bone in living animals. *J Biomech* 1974;7:563–565.
12. Wright TM, Hayes WC. Strain gage application on compact bone. *J Biomech* 1979;12:471–475.
13. Hylander WL, Bays R. An *in vivo* strain-gauge analysis of the squamosal-dentary joint reaction force during mastication and incisal biting in *Macaca mulatta* and *Macaca fascicularis*. *Arch Oral Biol* 1979;24:689–697.
14. Hylander WL. *In vivo* bone strain in the mandible of *Galago crassicaudatus*. *Am J Phys Anthropol* 1977;46:309–326.
15. Ross CF. *In vivo* function of the craniofacial haft: The interorbital "pillar". *Am J Phys Anthropol* 2001;116:108–139.
16. Daegling DJ, Hylander WL. Biomechanics of torsion in the human mandible. *Am J Phys Anthropol* 1998;105:73–87.
17. Behrents RG, Carlson DS, Abdelnour T. *In vivo* analysis of bone strain about the sagittal suture in *macaca mulatta* during masticatory movements. *J Dent Res* 1978;57:904–908.
18. Herring SW, Rafferty KL. Cranial and facial sutures: functional loading in relation to growth and morphology. In: Davidovitch Z, Mah J, editors. *Biological Mechanisms of Tooth Eruption, Resorption and Replacement by Implants*. Boston: Harvard Society for Advanced Orthodontics; 2000. p 269–276.
19. Ahmad F, et al. Strain gauge biomechanical evaluation of forces in orbital floor fractures. *Br J Plast Surg* 2003;56:3–9.
20. Ekenman I, et al. Local bone deformation at two predominant sites for stress fractures of the tibia: an *in vivo* study. *Foot Ankle Int* 1998;19:479–484.
21. Unnewehr M, et al. Fracture properties of the human mandible. *Int J Legal Med* 2003;117:326–330.
22. Lanyon LE. Analysis of surface bone strain in calcaneus of sheep during normal locomotion—strain analysis of calcaneus. *J Biomech* 1973;6:41–49.
23. Lanyon LE, Magee PT, Baggott DG. Relationship of functional stress and strain to the processes of bone remodeling—experimental study on the sheep radius. *J Biomech* 1979;12:593–600.
24. Lanyon LE, et al. *In vivo* strain-measurements from bone and prosthesis following total hip-replacement—an experimental study in sheep. *J Bone Joint Surg* 1981;63:989–1001.
25. Draganich LF, Reider B, Miller PR. An *in vitro* study of the Muller anterolateral femorotibial ligament tenodesis in the anterior cruciate ligament deficient knee. *Am J Sports Med* 1989;17:357–362.
26. Salmons S. *In vivo* tendon tension and bone strain measurement and correlation. *J Biomech* 1975;8:87.
27. Hart RT. Bone modeling and remodeling: theories and computations. In: Cowin SC, editor. *Bone Mechanics Handbook*. Boca Raton (FL): CRC Press; 2001. 31: p 1–42.
28. Martin RB, Burr DB, Sharkey NA. *Skeletal Tissue Mechanics*. New York: Springer; 1998.
29. Lanyon LE, et al. Osteocytes, strain detection, bone modeling and remodeling. *Calcified Tissue International* 1993;53: S102–S107.
30. Dechow PC, Hylander WL. Elastic properties and masticatory bone stress in the macaque mandible. *Am J Phys Anthropol* 2000;112:553–574.
31. Lertchirakarn V, Palamara JE, Messer HH. Finite element analysis and strain-gauge studies of vertical root fracture. *J Endod* 2003;29:529–534.
32. Rohlmann A, Mossner U, Bergmann G, Kolbel R. Finite element analysis and experimental investigation in a femur with hip endoprosthesis. *J Biomech* 1983;16:727–742.
33. Strait DS, et al. Modeling elastic properties in finite element analysis: How much precision is needed to produce an accurate model? *Anat Rec* 2005;283A:275–287.
34. Wang Q, et al. *In vitro* strain of monkey facial sutures. *Am J Phys Anthropol* 2005;126:S40–223.
35. Dechow PC, Carlson DS. Occlusal force after mandibular advancement in adult rhesus monkeys. *J Oral Maxillofac Surg* 1986;44:887–893.
36. Dechow PC, Carlson DS. Occlusal force and craniofacial biomechanics during growth in rhesus monkeys. *Am J Phys Anthropol* 1990;83:219–237.
37. Throckmorton GS, Ellis 3rd E, Buschang PH. Morphologic and biomechanical correlates with maximum bite forces in orthognathic surgery patients. *J Oral Maxillofac Surg* 2000;58:515–524.
38. Dechow PC, Carlson DS. A method of bite force measurement in primates. *J Biomech* 1983;16:797–802.
39. Throckmorton GS, Ellis 3rd E, Winkler AJ, Dechow PC. Bone strain following application of a rigid bone plate: An *in vitro* study in human mandibles. *J Oral Maxillofac Surg* 1992;50:1066–1074.
40. Dechow PC, Ellis 3rd E, Throckmorton GS. Structural properties of mandibular bone following application of a bone plate. *J Oral Maxillofac Surg* 1995;53:1044–1051.
41. Watanabe F, et al. Analysis of stress distribution in a screw-retained implant prosthesis. *Int J Oral Maxillofac Implants* 2000;15:209–218.
42. Kim YH, Kim JS, Cho SH. Strain distribution in the proximal human femur. An *in vitro* comparison in the intact femur and after insertion of reference and experimental femoral stems. *J Bone Joint Surg Br* 2001;83:295–301.
43. Lemaitre L, Moors M, Vanpeteghem AP. Method for the measurement of dimensional change of dental amalgam. *J Biomed Mat Res* 1979;13:887–892.
44. Nielsen PE, Rasmussen SM. Indirect measurement of systolic blood pressure by strain gauge technique at finger, ankle and toe in diabetic patients without symptoms of occlusive arterial disease. *Diabetologia* 1973;9:25–29.
45. Fryer TB, Silverberg GD, Corbin SD, Schmidt G. Telemetry of intracranial-pressure. *Biotelemetry* 1978;5:46.

46. Brosnan RJ, et al. Effects of ventilation and isoflurane end-tidal concentration on intracranial and cerebral perfusion pressures in horses. *Am J Vet Res* 2003;64:21–25.
47. Cooperman M, et al. Detection of deep venous thrombosis by impedance plethysmography. *Am J Surg* 1979;137:252–254.
48. Maskell NA, et al. The use of automated strain gauge plethysmography in the diagnosis of deep vein thrombosis. *Br J Radiol* 2002;75:648–665.
49. Andres PL, et al. Quantitative assessment of neuromuscular deficit in ALS. *Neurol Clin* 1987;5:125–141.
50. de V Cotton M. Circulatory changes affecting measurement of heart force in situ with strain gage arches. *Am J Physiol* 1953;174:365–370.
51. Mote PS, Pruett JK, Gramling ZW. Effects of halothane and enflurane on right ventricular performance in hearts of dogs anesthetized with pentobarbital sodium. *Anesthesiology* 1983;58:53–60.
52. Sarel O, Hasin Y, Rogel S. Myocardial conduction time and antiarrhythmic drugs. *J Electrocardiol* 1981;14:261–266.
53. Hasin Y, Sarel O, Rogel S. Electrical and mechanical response in biventricular mechanical alternans. *Arch Int Physiol Biochim* 1979;87:19–28.
54. Cavallo SA, Baken RJ. Prephonatory laryngeal and chest wall dynamics. *J Speech Hearing Res* 1985;28:79–87.
55. Souhrada JF, Dickey DW. Mechanical activities of trachea as measured *in vitro* and *in vivo*. *Respir Physiol* 1976;26:27–40.
56. Flack FC, James ED. Case study using simultaneous bladder pressure and urine loss measurements. *Urol Int* 1975;30:103–108.
57. Johnson CP, et al. Effects of intestinal transplantation on postprandial motility and regulation of intestinal transit. *Surgery* 2001;129:6–14.
58. Pascaud XB, Genton MJ, Bass P. A miniature transducer for recording intestinal motility in unrestrained chronic rats. *Am J Physiol* 1978;235:E532–538.
59. Ueda S, Wada A, Umemura S. Methodological validity and feasibility of the nitric oxide clamp technique for nitric oxide research in human resistant vessels. *Hypertens Res* 2001;27:351–357.
60. Bigelow N, et al. A preliminary report on a study of a correlation between emotional reactions and peripheral blood circulation, using a strain gauge plethysmograph. *Psychiatr Q* 1955;29:193–202.
61. Forgione AG, Barber TX. A strain gauge pain stimulator. *Psychophysiology* 1971;8:102–106.
62. Coakley D, Williams R, Morris J. Minute eye movement during sleep. *Electroencephalogr Clin Neurophysiol* 1979;47:126–131.
63. Mamelak A, Hobson JA. Nightcap: A home-based sleep monitoring system. *Sleep* 1989;12:157–166.
64. Miyazaki S, et al. Using an air-pad sensor for the diagnosis of sleep apnea: A trial study. *Psychiatr Clin Neurosci* 2002;56:315–316.
65. Janssen E, Vissenberg M, Visser S, Everaerd W. An *in vivo* comparison of two circumferential penile strain gauges: The introduction of a new calibration method. *Psychophysiology* 1997;34:717–20.
66. Kiely ME, Thavundayil JX, Lal S. Effect of blood sampling on apomorphine-induced penile tumescence in erectile impotence: A case report. *J Psychiat Neurosci* 1995;20:233–235.
67. Hahn PM, Leder R. Quantification of penile buckling force. *Sleep* 1980;3:95–97.
68. An KN, Chao EY, Askew LJ. Hand strength measurement instruments. *Arch Phys Med Rehabil* 1980;61:366–368.
69. Burnham RS, Bell G, Olenik L, Reid DC. Shoulder abduction strength measurement in football players: Reliability and validity of two field tests. *Clin J Sport Med* 1995;5:90–94.
70. Henning CE, Lynch MA, Glick Jr KR. An *in vivo* strain gage study of elongation of the anterior cruciate ligament. *Am J Sports Med* 1985;13:22–26.
71. Milgrom C, et al. The effect of shoe gear on human tibial strains recorded during dynamic loading: A pilot study. *Foot Ankle Int* 1996;17:667–671.
72. Omata S, Murayama Y, Constantinou CE. Development of a novel surgical support instrument and virtual system incorporating new tactile sensor technology. *Stud Health Technol Inform* 2004;98:288–290.
73. Iida I, Noro K. An analysis of the reduction of elasticity on the ageing of human skin and the recovering effect of a facial massage. *Ergonomics* 1995;38:1921–1931.

Further Reading

These books contain much useful information on strain gages, stress and strain analysis, and instrumentation for strain gages. Internet searches using strain gage or strain gauge as key words also yield a wide variety of information about bridges, strain gages, associated products, and instrumentations, manufacturers, technique supports, and training sessions.

- Dally JW, Riley WF. *Experimental Stress Analysis*. 3rd ed. New York: McGraw-Hill; 1991.
- Hannah RL, Reed SE, editors. *Strain Gage Users' Handbook*. London: Elsevier; 1992.
- Khan AS, Wang X. *Strain Measurements and Stress Analysis*. Upper Saddle River (NJ): Prentice Hall; 2000.
- Kost GJ, editor. *Handbook of Clinical Automation, Robotics, and Optimization*. New York: Wiley; 1996.
- Perez R. *Design of Medical Electronic Devices*. London: Academic Press; 2002.
- Prutchi D, Norris M. *Design and Development of Medical Electronic Instrumentation: A Practical Perspective of the Design, Construction, and Test of Medical Devices*. Hoboken, NJ: Wiley-Interscience; 2004.
- Webster JG, editor. *Medical Instrumentation: Application and Design*. 3rd ed. New York: Wiley; 1998.
- Window AL. *Strain Gauge Technology*. 2nd ed. Burlington (MA): Elsevier Science; 1992.
- Ballantyne GH, Marescaux J, Giulianotti PC, *Primer of Robotic & Telerobotic Surgery*. Hagerstown MD: Lippincott Williams & Wilkins; 2004.

See also BIOMATERIALS, TESTING AND STRUCTURAL PROPERTIES OF; BONE AND TEETH, PROPERTIES OF; LIGAMENT AND TENDON, PROPERTIES OF; REHABILITATION AND MUSCLE TESTING; TOOTH AND JAW, BIOMECHANICS OF.

STRESS TESTING, CARDIOVASCULAR. See EXERCISE STRESS TESTING.

SURFACE PROPERTIES OF BIOMATERIALS. See BIOMATERIALS, SURFACE PROPERTIES OF.

SURGERY, FROZEN. See CRYOSURGERY.

SURGERY, INTRAUTERINE. See INTRAUTERINE SURGICAL TECHNIQUES.

SURGERY, MINIMALLY INVASIVE. See MINIMALLY INVASIVE SURGERY.

SURGERY, STEREOTACTIC. See STEREOTACTIC SURGERY.

SWEAT TEST FOR CYSTIC FIBROSIS. See CYSTIC FIBROSIS SWEAT TEST.

SYSTEMIC HYPERTHERMIA. See HYPERTHERMIA, SYSTEMIC

TACTILE STIMULATION

CHRISTOPHER J. POLETTO
National Institutes of Health

INTRODUCTION

Definitions

In the most general terms, tactile stimulation is the deliberate elicitation of any of a range of sensations perceived through the sense of touch. The stimulation can be delivered by any means and to any part of the body where touch can be felt. Most commonly, tactile stimulation is applied to the skin, usually to the fingertips, abdomen, or forearms, but tactile stimulation is also used on the tongue as a possible communications tool and at the back of the throat during therapy in people with dysphagia. Clinically relevant tactile stimulation can be as simple as a caregiver stroking a baby to encourage healthy development or involve of thousands of sophisticated actuators in arrays working in concert to present a virtual tactile environment.

Tactile perception should be distinguished from kinesthetic and haptic perception. Tactile perceptions include temperature, skin curvature and stretch, vibration, slip, pressure and local contact force. Kinesthesia is the perception of the relative position and movement of our body parts (proprioception), as well as the sensation of muscular effort exerted while touching or manipulating objects. Although M. Dessoir (1892) originally defined haptic (*haptik* in German, modified from the Greek word *haptikos*) to mean “the study of touch and tactile sensations, especially as a means of communication” (Oxford English Dictionary), the term haptic is applied much more broadly now. Haptic perceptions combine tactile and kinesthetic perceptions to provide a sense of environmental or object properties such as shape (1). Many devices referred to as “tactile displays”, especially those intended for virtual reality or telepresence applications, actually combine tactile and kinesthetic feedback, but this article is concerned only with devices that deliver tactile stimulation. The focus will be on the mechanical (vibrotactile and shape displays) and electrical (electrotactile) stimulators that elicit sensations of vibration, pressure, and local contact force, because these types of stimulators are most often used in current tactile stimulation applications.

Applications

Telepresence and telerobotic technologies are emerging as important fields with great potential for use in biomedical applications such as the control of surgical robots. These applications require mechanisms for the feedback of information to the human operator from a set of remote sensors. This feedback has been visual and auditory in nature, but developing technologies are permitting the use of certain forms of haptic feedback such as force reflection to com-

municate information about a manipulated object’s compliance, viscosity, mass, size, and gross shape. Researchers have made less progress toward communicating tactile information such as surface texture, fine shape, roughness, slip, vibration or temperature. These later object properties could potentially be important in helping the user distinguish between morphologically similar objects or, in the case of surgical or medical imaging applications, to assess tissue type or of the degree and extent of tissue damage (2,3). Indeed, in cases where visual information is not available, tactile information may form a user’s only basis for decision making.

Other applications for tactile stimulation include sensory substitution or augmentation systems such as those for the blind or deaf. For example, using a computer poses a particular challenge for the visually impaired. Textual information can be read aloud by the computer, but as graphical computer interfaces become more and more ubiquitous, the blind user is placed at a greater disadvantage. One solution is to represent the graphical information through a tactile display. Further applications are discussed below in the section on stimulation techniques.

TACTILE PHYSIOLOGY

Understanding how tactile stimulation techniques function requires a basic understanding of the physiological mechanisms that mediate our sense of touch. When we touch an object, the distribution of contact forces deforms our skin. The temporal and spatial distribution of the deformation determines which mechanoreceptive nerve terminals are excited. Each mechanoreceptive nerve fiber terminates as one or more dendrites. The dendrites have stretch-gated ion channels that allow ions to flow and depolarize the dendrite when their membranes are stretched or deformed. If the depolarization is sufficient, nearby voltage gated channels initiate an action potential that travels up the nerve to the central nervous system. The particular way that the deformation of the skin leads to bending or stretching of the dendrites varies from one type of mechanoreceptor to another and is due to the structure of the sensory receptor at the end of the fiber. The firing frequencies of the mechanoreceptive nerve fibers are then decoded by the central nervous system to ultimately give rise to our sense of touch.

The skin is made up of three distinct layers: the epidermis, dermis, and subcutaneous fat. Made up of primarily keratinocytes, the epidermis is the outermost layer and serves as a protective interface with the outside world. The outermost portion of the epidermis is the *stratum corneum*, which is made up of dead, flat cells that shed and are replenished about every 2 weeks. The epidermis does not contain any blood vessels and is dependant on the deeper skin layers for its oxygen and nutrients. A very thin membrane, the basement membrane, attaches the

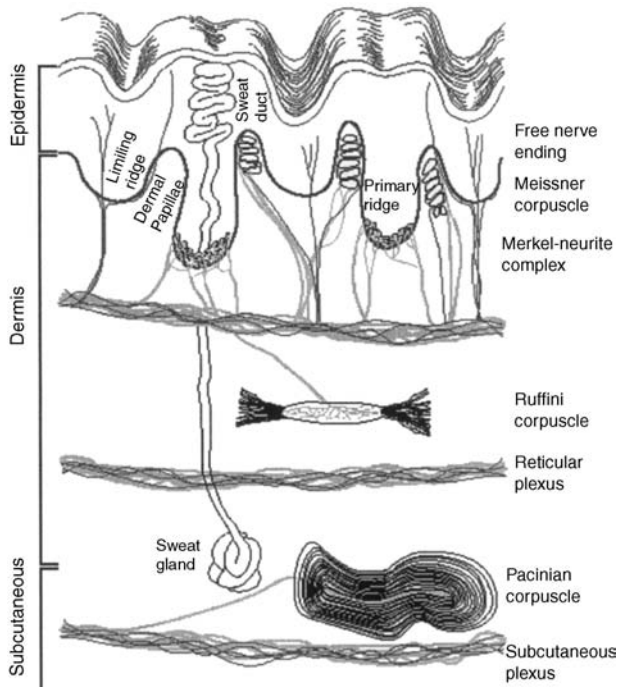


Figure 1. Mechanoreceptors in skin. Mechanoreceptive innervation of human hairless (glabrous) skin. (Diagram provided by Kenneth Johnson.)

epidermis firmly, though not rigidly, to the second layer, the dermis. The dermis contains blood vessels, nerves, hair roots, and sweat glands. Below the dermis lies a layer of fat, the subcutaneous fat, which serves as a cushion between the skin and the underlying muscle and bone. The depth of this layer differs from one person to another.

Four major types of mechanoreceptors are present in human glabrous skin (the hairless skin of the lips, soles, palms, and finger pads) that respond to touch or vibration: Pacinian corpuscles (PC or RAI), Meissner corpuscles (rapidly adapting or RAI), Merkel endings (slowly adapting or SA I), and Ruffini endings (slowly adapting or SA II) (Fig. 1). In hairy skin, the Meissner corpuscles are replaced by hair receptors with similar mechanoreceptive properties. Only a very basic overview of mechanoreceptor morphology. A recent review of the roles that mechanoreceptors play in perception can be found in Ref. 7.

Each type of mechanoreceptor is tuned so that its threshold for excitation is lowest for a particular range of vibration frequencies, although there is considerable overlap between receptor types. The receptors that respond best to higher frequency stimulation are referred to as rapidly adapting, while those that respond best to lower frequencies are called slowly adapting. A sudden but sustained skin indentation initially results in rapid firing of all the affected mechanoreceptors. The rapidly adapting ones soon cease firing (however, they may fire again at stimulus offset), whereas the slowly adapting mechanoreceptors continue firing (albeit at a lower rate) until contact is broken (Fig. 2).

The area in which stimulation leads to a response by a particular mechanoreceptive neuron is known as the

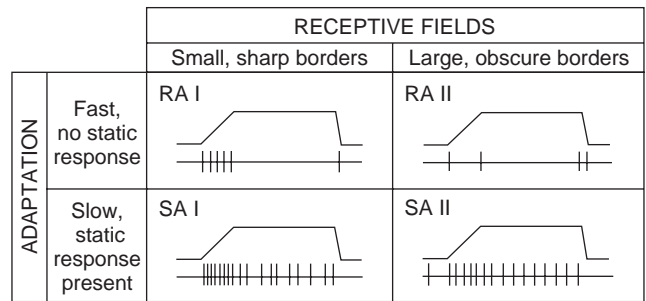


Figure 2. Response of RA and SA fibers. Responses of typical rapidly adapting (RA) and slowly adapting (SA) mechanoreceptors with large or small receptive fields to ramped and sustained indentation. The onset of the indentation stimulates rapid volleys of action potentials in both sets of mechanoreceptors, but the RAs adapt and cease firing after a few milliseconds. The SAs continue firing until indentation offset, which may trigger another burst from the RAs. (Figure redrawn from Ref. 124.)

receptive field of that neuron. The skin tissue acts as a low pass spatial filter so that the deeper a mechanoreceptor lies in the skin, the larger its receptive field. The receptive field also depends on how many receptors a mechanoreceptive neuron innervates and how widely they are dispersed. Mechanoreceptors with small receptive fields are more suitable for fine spatial discrimination than those with large receptive fields. Mechanoreceptors with small receptive fields are referred to as type I mechanoreceptors and those with large receptive fields are called type II.

The frequency tuning of a mechanoreceptive fiber (i.e., whether it is rapidly or slowly adapting) is due to its mechanical structure. For example, the Pacinian corpuscle receptor is shaped like an onion with multiple layers wrapped around the dendrite. When the skin is indented, the onion-shaped receptor is compressed. With sufficient pressure to deform the corpuscle, parts of the neural membrane are stretched and action potentials are stimulated. The Pacinian rapidly adapts to constant pressure because with maintained pressure, the gel-filled layers of the corpuscle begin to slide across each other in a way that relieves the stress on the dendrite and stops the generation of action potentials. A vibratory stimulus, on the other hand, repeatedly deforms the corpuscle, eliciting continued neuronal firing. The Pacinians are sensitive to a wide range of frequencies (50–1000 Hz), but are exquisitely sensitive to vibrations between 250 and 350 Hz, requiring only sub-micron surface displacements for excitation. The Pacinians lie deep in the dermis, so pressure anywhere over a fair amount of skin area can compress it enough to affect the dendrite, providing it with a large receptive field. Isolated stimulation of a particular Pacinian nerve fiber leads to a sensation of diffuse vibration. Even though their deep position renders them unsuitable for fine stimulus localization, the Pacinians aid in the perception of surface texture by detecting the vibrations elicited when the skin is brushed over a rough object (8). Pacinian corpuscles are also thought to play a role in tool use, specifically in our ability to attribute actions to the end of the tools we hold, by sensing the vibrations transmitted through the tool.

The ultrastructure of the Meissner corpuscle is more complicated than that of the Pacinian corpuscle, but Meissners are also rapidly adapting. The Meissners respond best to vibratory stimuli with frequencies between 30 and 70 Hz, but also respond to frequencies as low as 3 Hz. The Meissner corpuscles are located just below the dermal-epidermal boundary and have smaller receptive fields than Pacinians. Meissners are considered the primary mediators of the sensations of light touch and flutter while the Pacinians are considered the primary mediators of high frequency vibration and deep pressure.

The primary types of slowly adapting mechanoreceptors are the Merkel complexes and the Ruffini corpuscles. Merckels have a receptive field diameter of 3–4 mm, and a frequency range of 2–32 Hz (9). The Merkel sensory neurons have nonencapsulated nerve endings (neurites) that form a complex arrangement with disk-shaped Merkel cells in the basal layer of the epidermis. There is conflicting data about just how this structural arrangement accounts for the mechanoreceptive properties of the receptors, but there is general agreement that the Merkel-neurite complexes mediate the perception of steady skin indentation and their population response is thought to account for our perceptions of form and texture (7).

Ruffini corpuscles are relatively large spindle-shaped structures that are rooted in the connective tissue of the dermis. They have been studied less than other mechanoreceptors because we derive much of our understanding about mechanoreceptor physiology from studies performed on monkeys and Ruffini corpuscles have never been observed in neurophysiological studies of monkey hands. Additionally, it is not clear which perceptions should be attributed to them, because microstimulation of Ruffini fibers produces no sensation (10), and elicits no cortical somatosensory evoked potentials (11). There is some evidence, however, that Ruffini endings may participate in tactile sensations as SAI receptors (12). Currently, it is believed that Ruffinis mediate perceptions of directional skin stretch. Directional stretch sensitivity would allow them to participate in the perception of hand and finger position as well the perception of the direction of object motion or force (13,14).

In addition to the mechanoreceptors discussed above, the skin contains other sensory fibers that might be important for tactile stimulation that are outside the scope of this article, but deserve mention. The previously mentioned cutaneous and subcutaneous mechanoreceptors have large, fast, myelinated (A β) fibers, but there are also mechanically sensitive cutaneous free nerve endings with small, slower myelinated (A δ) fibers that respond to strong mechanical stimulation, especially by sharp objects. These mechanoreceptive *nociceptors* (receptors that respond to potentially damaging stimuli) mediate primarily pricking pain, but some also respond to thermal stimuli. Additionally, there are even smaller and slower, unmyelinated (C) fibers present that respond to noxious mechanical and/or thermal stimuli and are primarily responsible for sensations of burning pain. Although almost all sensory C fibers are traditionally considered to mediate pain, there is growing evidence that some mediate touch that is specifically pleasant and emotionally salient, such as from stroking by

another human (15,16). It is unknown what role C-fibers will play in tactile stimulation devices in the future.

TACTILE STIMULATION TECHNIQUES AND APPLICATIONS

The two main approaches used to artificially stimulate the tactile senses are based on exciting the receptors through mechanical perturbation of the skin or electrically exciting the mechanoreceptive nerve fibers that innervate the receptors. In both approaches, the goal is to manipulate the tactile information through variations in the four tactile primitives that code tactile information: intensity, frequency, spatial pattern, and temporal pattern (17). In addition to these classical primitives, there is evidence for a multidimensional electrotactile primitive that could be termed “color” or “quality”, of which frequency may be one dimension (18,19).

Efforts have also been reported to create tactile displays that illicit sensations of warmth or cold using thermal elements such as Peltier cells (20–22). Although interesting and potentially useful, these thermal displays are beyond the scope of this article.

Mechanical

Mechanical tactile stimulators predominately fall into one of two classes: vibrotactile or shape displays (Fig. 3). Vibrotactile displays use contactors that vibrate at a particular, usually fixed, frequency. Shape displays provide small-scale indentation, pressure, or shear profiles to the skin, often using an array of pins that can each be pressed into or withdrawn from the skin.

A useful vibrotactile display can be as simple as the early Tactaid (Audiological Engineering Corporation, Somerville, MA). The Tactaid is a single vibrator that is affixed to the chest, back of the neck, or wrist that transduces acoustic signals into vibration to convey certain speech information (the modern Tactaid VII uses an array of seven vibrators in a line). Single vibrators are also becoming common on force feedback devices such as joysticks and robotic manipulators. They are even becoming popular for home computer devices such as Logitech’s iFeel MouseMan computer mouse.

Relatively simple vibrotactile stimulators also offer clinical benefit for individuals with impaired tactile sensation. For example, Collins et al. (23) recently demonstrated that three small vibrating *tactors* (tactile stimulators), imbedded in gel-based shoe insoles, helped to maintain balance and improve postural sway during quiet standing in healthy young and elderly individuals, as well as patients with stroke, and patients with diabetic neuropathy (24,25). This application is particularly interesting because the stimuli were essentially random noise delivered at intensities *below* sensory threshold.

The subthreshold noise enhances the user’s ability to detect useful sensory signals from the sole of the foot through the principle of *stochastic resonance*. This principle, found to apply in many nonlinear physical and biological systems, describes systems in which the ability to detect a signal in the presence of noise is actually enhanced by adding a critical amount of noise. Collins contends that

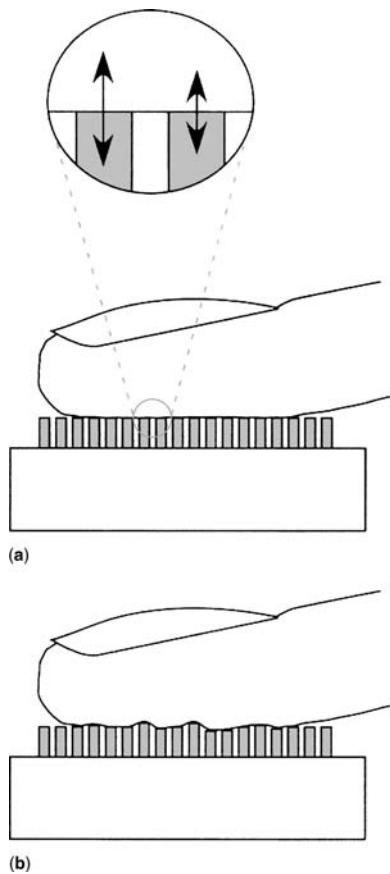


Figure 3. Vibrotactile versus shape displays. Vibrotactile displays (a) use pins that vibrate around a common set point. Larger oscillations are used to produce sensations of higher intensity (top inset). Shape, or static, displays (b) present more naturalistic surface contours and use larger indentations to encode regions of higher intensity.

such noise-based devices could ameliorate age-related impairments in balance control. Similar devices using electrotactile stimulation are offered by Afferent, a company Collins helped found.

A small number of vibrotactile elements can also be useful in teleoperation applications. For example, a single vibrating element encoding manipulator force was found to improve the ease of use of a deep sea remote manipulator (26,27). Similarly, Debus et al. (28) used a cylindrical handle with four embedded vibrating elements to demonstrate that force feedback vibrations can significantly improve a teleoperated force control task by reducing the mean force errors by 35% compared to visual feedback alone.

Vibrotactile displays that utilize dense arrays of tactors can communicate more information than single oscillators. Tactile stimulators for use outside the laboratory were initially developed as sensory substitution aids for the sight and hearing impaired. One of the first tactile arrays used vibrating contacts on the users back to communicate visual information to blind users (29), an effect called "Tactile Television". Another example of an early vibrotactile array for sensory substitution was the Optacon (Optical to TActile CONverter), a portable reading device for the blind (30) that was originally developed in 1966 by John

Linvill and JC Bliss to assist Linvill's vision impaired daughter. The Optacon allowed the visually impaired to read printed material by feeling a tactile version of a visual image. A small hand-held camera scanned pictures or text in any document. The image was presented to the user via an array of vibrating metal rods that stimulated an index finger. The stimulator consisted of an array of 144 (24 rows \times 6 columns) of vibrating metal pins mounted on top of the controller. Each of the vibrating pins was driven by a piezoelectric bimorph reed actuator. The device was a technological success; many blind individuals reported that it provide increased autonomy. Debbie Stein, the First Vice President of the National Federation of the Blind of Illinois, and a long-time Optacon user stated (31) "The Optacon has given blind people a level of autonomy and flexibility unparalleled in history." As technically successful as the compact, lightweight, and highly portable Optacon was, it was not a financial success so Telesensory, Inc. discontinued manufacturing the Optacon in 1996, citing high cost and low demand.

From the time of the first vibrotactile displays, power consumption has been recognized as an important design factor. Higher power consumption requires larger and more expensive electronics, larger and heavier batteries, increased heat production, and less portability overall. Matching the display parameters (e.g., frequency, amplitude, tactor size, tactor density) to the mechanoreceptor characteristics (e.g., frequency response, sensitivity, receptive field) allows reduction of the power requirements, and many theoretical studies have pursued this goal (32–34). For example, the frequency at which the pins in a vibrotactile display oscillate is often selected to maximize the amplitude of the perceived signal (typically 200–300 Hz) while minimizing the power requirements of the display. The standard psychophysical term used to refer to the perceived amplitude of a vibratory stimulus is *Loudness*. For a given amount of actuator power, the loudness of a sinusoidal stimulus is maximized at \sim 250 Hz (35–37), the frequency Pacinian corpuscles are most sensitive to. Accordingly, several vibrotactile stimulators use sinusoidal vibrations at or near this frequency (200 Hz in the case of the Optacon). An average young, healthy user can feel 250 Hz vibrations $<1 \mu\text{m}$ in amplitude (38).

The use of rectangular pulses instead of sinusoidal waveforms can result in lower power requirement, especially if short pulses are used so that multiple pins can share the same driver stages. The electronics required to generate rectangular pulses can also be much simpler, potentially as simple as a single transistor operating as a switch (34), further reducing the size and cost of the display. A recent study that measured power consumption experimentally for electromechanical and piezoelectric vibrotactile actuators for sinusoidal and rectangular pulse waveforms (39) found that power consumption for the piezoelectric reed bimorph transducers was two orders of magnitude lower than for the electromechanical transducers. The piezoelectric transducers were more efficient despite their higher voltage requirements, because they required almost no current. Curiously, within the electromechanical transducers, the least power required to reach threshold was using short (0.7 ms low duty cycle)

rectangular pulses delivered at a rate of 25/s, but for the piezoelectric transducers the most efficient waveform was sinusoidal at 250 Hz. Clearly, the most efficient stimulus waveform depends on the type of actuators chosen for the display.

Shape displays tend to have low temporal resolution (unless constructed of piezoelectric actuators), but potentially very high spatial resolution. Static shape displays are perhaps the original example of tactile displays with the primary example being Braille. Blinded at the age of three, Louis Braille (1809–1852), invented the familiar tactile shape display reading system that bears his name. Individual letters, numbers, or common combinations of letters are represented by combinations of six or eight raised dots.

Ideally, a dynamic (changing) tactile shape display could be developed that would present patterns that would be indistinguishable to the user from direct contact with the original (or virtual) surface. The display must therefore deform the skin in the same way [or at least produce the same strain patterns at the receptor level (40–41)] that the represented surface would. The slow development of these dynamic fingertip tactile displays is due to the numerous unknown factors that characterize the sense of touch.

Designing an optimal dynamic tactile display requires precise knowledge of the biomechanics of the skin, which is both tough and deformable at the same time (42). When scanning a texture, we use contact forces ranging between 0.3 and 4.5 N and a scanning speed between 1 and 25 cm/s with an average of 2 cm/s (43). Theoretically, to have full control of the surface stresses, the ideal tactile display system needs an infinitely dense array of actuators, each with three degrees of freedom and infinite stiffness. Nearly ideal function would also require a display with actuators capable of delivering at least 50 N/cm² pressure, able to indent the skin at least 4 mm (with 10% resolution), an actuator density of 1/mm², and at least 50 Hz refresh bandwidth, requiring a power density of 10 W/cm² (44). In fact, it may even be advantageous to provide a refresh rate that matches the highest frequency response of the RA receptors, in the kilohertz. The optimal display would also be small enough for use with several fingers at once over a large range of positions and orientations. Due to compromises that have to be made between power, bandwidth, manufacturability, cost, size, actuator stiffness, ease of maintenance, and other factors, no tactile display currently comes close to satisfying these requirements.

One goal of some shape displays, such as those for telesurgery or telerobotics, is to communicate information about the compliance of the manipulated (or virtual) object. Human discriminability of softness (compliance) of objects depends on the object having a deformable surface (45–46). For deformable surfaces, the spatial pressure distribution within the contact region depends on object compliance, so that information from cutaneous mechanoreceptors is sufficient for the user to gauge subtle differences in compliance. When the surface is rigid, however, we require kinesthetic information for discrimination, and our ability to gauge softness is much poorer than that for objects with deformable surfaces. It is likely that the spatiotemporal

variation of pressure on the skin (or, equivalently the skin displacement and its derivatives) form the basis for the perception of softness of compressible objects. Consequently, a shape display that aims to communicate compliance information must itself be deformable (compliant) (45). This compliance might be achieved through either passive or active means.

Designers of shape displays have generally taken one of two approaches; (1) control the force the actuators exert on the skin (in the direction perpendicular to the skin), or (2) control the displacement (again, perpendicular) that the actuators impose on the skin. A smaller number have controlled the lateral stresses or strains applied to the skin. Almost all shape displays have been designed for the fingertips. Shape display designs have used solenoids (47), electrostatics (48–49), voice coil actuators (50), shape memory alloys (51–52), pneumatics (44,53), RC servomotors (54), MEMS (55), and even air jets (56). A good summary of the technologies and their relative merits can be found in Ref. (17), with an updated version currently available at http://www.cim.mcgill.ca/~jay/index_files/research_files/actuators.htm.

Electrotactile

The use of mechanical systems to communicate detailed tactile information is limited by the need for dense arrays of end effectors with sufficient mechanical compliance to overcome the stiffness of the skin and deeper tissues. Such stimulators require bulky actuators that are difficult to arrange in the dense arrays required for texture communication. For these reasons, researchers have turned to electrical stimulation of the tactile senses. Electrical stimulation that aims to produce a localized tactile perception is referred to as “electrotactile” or “electrocuteaneous” stimulation, whereas “transcutaneous” stimulation describes the more generalized stimulation of nerve bundles [e.g., transcutaneous electrical nerve stimulators (TENS)]. Electrotactile stimulation has the advantages that dense arrays of electrodes can be fabricated and that the stimulator itself can be remotely located so it does not limit the spacing or rigid presentation of the electrodes. Electrotactile stimulation displays also typically require much less power than vibrotactile devices.

In contrast to mechanical stimulators, which deform the receptor at the end of a mechanoreceptive nerve fiber, electrical stimulators depolarize the membrane of the mechanoreceptive nerve fibers directly by passing current through the skin from one electrode (or group of electrodes) to another. The current passing through the skin produces an electric field that extends some distance into the skin. This electric field creates currents within the nerve fibers that depolarize the neural membrane. If this depolarization is sufficient, the neuron fires an action potential that propagates to the central nervous system. In general, the smaller the stimulating electrodes and the closer they are spaced, the more superficial the electric field will be. For a given current, the strength of the electric field beneath an electrode varies inversely with the surface area of the electrode so stimulation will always occur at the smaller of the two electrodes. The usual practice is to use a large

remote electrode or a large surrounding electrode (or group of electrodes) as the current return and a smaller electrode as the stimulating electrode. The negatively charged electrode is referred to as the *cathode* and the positively charged one is the *anode*. Cathodic stimulation is therefore when the negatively charged electrode is the stimulating electrode, whereas anodic stimulation uses the positively charged electrode for stimulation. Cathodic stimulation is the most conventional and is best for exciting nerve fibers that are passing by the electrode, while anodic stimulation is best for exciting fibers that terminate in the vicinity of the electrode. The stimulating phase of the waveform is usually followed by a charge-recovery phase of opposite polarity and much lower amplitude. The reason for the charge-recovery phase is to avoid electrode corrosion and tissue damage by reversing as much as possible the chemical reactions that occur at the electrodes.

Electrocutaneous and transcutaneous electrical stimulation are used in a wide variety of applications including TENS units for pain control and sensory substitution systems such as speech aids (57,58) for the deaf and visual substitution systems (59,60) for the blind. Electrocutaneous stimulation also has been used to improve the utility of upper extremity (61–68) and lower extremity (69–71) prostheses for amputees. Subdermal electrocutaneous stimulation has been used to provide sensory feedback to users of an upper extremity neuroprosthesis (72). Electrocutaneous arrays on the forehead have been used to help people with advanced cases of Hansen's disease (leprosy) to perform detailed manual tasks with reduced chance of injury (73). An early electro tactile device that is still in widespread use is the Tickle Talker speech perception device for the severely hearing impaired (57). The Tickle Talker stimulates eight tactile sensory nerves as they pass between the second and third knuckles of the fingers with speech-encoding patterns and has been shown to provide benefits in supplementing lipreading or residual hearing for hearing-impaired adults and children (54). Visual information has been successfully presented to blind users through electro tactile arrays located on the abdomen (75,76), back, or tongue (81). The tongue is also the target region for a newly commercialized electro tactile device from Wicab, Inc. The 144-point electro tactile array sits on the roof of the mouth against the tongue surface and communicates patterns for a wide range of proposed sensory substitution applications such as providing crude visual information for the blind (78) or restoring postural stability in patients with vestibular deficits (79).

Although the trunk or tongue may be most appropriate for applications requiring independent use of the hands, the fingertips present a much more natural means to explore a virtual surface or environment. Indeed, active haptic exploration leads to better perception of two-dimensional (2D) shapes than does passive static touch (1). Additionally, both mechanical and electrostatic displays require fixed, flat displays, but electro tactile arrays can be mounted in the fingers of a glove, allowing the user to explore arbitrary tactile surfaces in a more natural manner. As with mechanical displays, arrays of tiny electrodes are needed to utilize the high spatial acuity of the fingertip.

The development of a fingertip, electro tactile interface with good spatial resolution will require techniques to selectively activate nerve fibers to produce well localized, precisely placed sensations. Such techniques will necessarily exploit the dynamical and geometrical characteristics of fingertip mechanoreceptive afferents while accounting for similar characteristics of the nociceptive afferents. One drawback of electro tactile stimulation, however, is that producing well-localized, painless sensations remains challenging (75,80–82). Previously successful attempts to use electrocutaneous stimulation as a communications tool have typically been limited to the use of a small number of large electrodes, a paradigm unsuitable for telepresence and most other fingertip applications. Another drawback of electro tactile stimulation for fingertip displays is that we know little about how surface electrical stimulation excites mechanoreceptors. In particular, there is still much to be learned about how to selectively activate one population of mechanoreceptors over another. Mechanical stimulation excites the mechanoreceptive sensors themselves, while electrical stimulation can potentially excite the mechanoreceptive neurons anywhere along their length. It is therefore very difficult to control both the quality and the location of the electrically induced perception.

There has been some progress toward better control of the perceived quality and location of stimuli presented through small electrodes. Poletto and Van Doren (83) performed a study a few years ago to investigate the effects of pulse width, electrode diameter, and stimulus polarity on mechanoreceptive and nociceptive thresholds, as well as their effects on the perceived location and quality of the induced sensations (83). Touch and pain thresholds were measured for every combination of six pulse widths (0.05, 0.126, 0.315, 0.792, 1.991, and 5 ms), four electrode diameters (1, 2, 4, and 8 mm), and two stimulus polarities (anodic and cathodic). After each threshold measurement, the subject reported the perceived location (local or distal to the electrode) of the induced touch and pain sensations, as well as a qualitative descriptor of the pain sensation. The perceived location results indicated that anodic stimulation through small electrodes can be used to reliably excite fibers terminating near the electrode without concurrent excitation of near-by axons of passage. On the other hand, traditional cathodic stimulation, particularly through large electrodes, is much more likely than anodic to excite axons of passage, resulting in a sensation distal to the electrode (Fig. 4). These results (along with other related trends evident in the threshold data) were echoed in the model simulations only when fiber morphologies similar to those observed microscopically were incorporated. A complete discussion of these experiments, as well as neurophysiological models that were created to explain them, can be found in (83). Similar modeling results and explanations for the experimental data have been found by Kajimoto et al. (84), who used a model of a different form.

Another challenge for electro tactile stimulation is that the difference between the pain thresholds and the sensation thresholds (this difference is known as "dynamic range") is smaller for electrical stimulation through small electrodes than for large. This could potentially limit the maximum intensity that can be presented through small

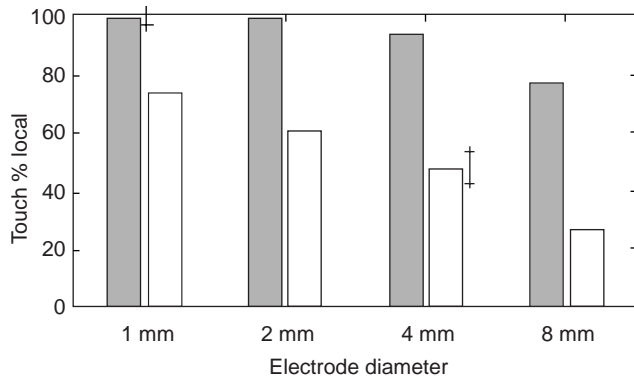


Figure 4. Qualitative results graphs. Electrotactile stimulation can produce sensations that are local or distal to the electrode. The graph shows the percentage of touch sensations that were felt local to (as opposed to distal to) the stimulating electrode, calculated as the percentage of all trials (summed across subject, session, pulse width and repetition, 271–288 total trials per bar) that the subject reported as “local”. Bar color indicates stimulus polarity: black, anodic; white, cathodic. Error bars indicate smallest and largest 95% confidence intervals. [Data taken from (83).]

electrodes and could limit the desirability of using electrotactile displays. One way to increase the dynamic range would be to inactivate the pain fibers by applying a long depolarizing prepulse (DPP) prior to the stimulus pulse (the DPP would only be introduced significantly above sensation threshold, to prevent the concurrent elevation of both thresholds). This approach has been shown to significantly increase the pain thresholds, thereby increasing the dynamic range (83).

Despite the difficulties associated with fingertip electrotactile displays using small electrodes, some groups have been successful using electrotactile stimulation to present simple patterns to the fingertips. Kaczmarek et al. (81) developed a 49-point fingertip-scanned electrotactile display that consists of 0.89 mm diameter, flat-topped stainless steel electrode “pins”, each surrounded by a 2.36 mm diameter air gap insulator arranged in a square 7×7 array

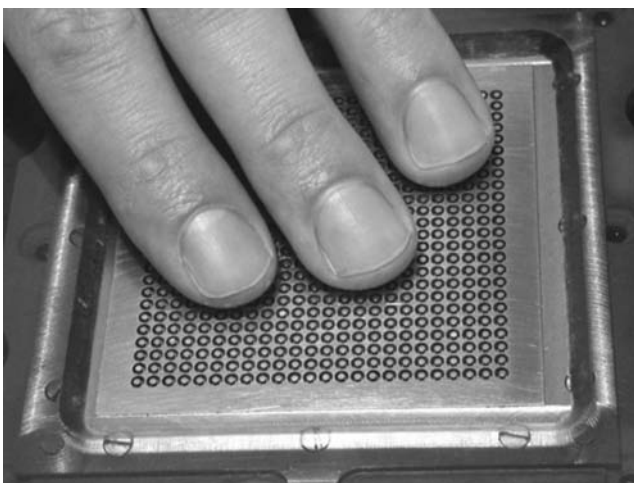


Figure 5. Fingertip electrotactile array. Electrotactile array with 576 electrodes (24×24) that can be scanned with multiple fingers. (Photo provided by Kurt Kaczmarek.)

(see Fig. 5 for a 24×24 array from the same group). They found that, using the highest comfortable current levels, subjects were able to correctly identify several simple geometric patterns $\sim 90\%$ of the time compared to near chance levels when stimulation was set to the lowest subthreshold current levels tested (81).

Other groups have also been working on electrotactile displays for selective stimulation of the various types of mechanoreceptors (56,84,86). Because the mechanoreceptors share the same neurophysiology (A δ nerve fibers), the only way to selectively stimulate them electrically is to take advantage of their differences in geometry and location. This is possible by focusing the current to flow deeply or superficially into the skin by changing the current source distribution at the skin surface. In this way, it is possible to manipulate the electric field applied to each type of mechanoreceptive fiber (83,84). Kajimoto et al. developed a tactile display capable of selectively eliciting sensations of pressure, high frequency vibration and low frequency vibration, which they refer to as “tactile primary colors”, analogous to the three primary colors for vision. This display was mounted on a computer mouse for tactile feedback in virtual reality (Fig. 6) and also on a surface with optical sensors underneath for conversion of light–dark patterns to tactile patterns (Fig. 7). The mouse-mounted application is innovative because it uses a force transducer beneath the fingertip array to allow the user to control the intensity of stimuli. This approach avoids the unpleasant sensation of sudden shocks that have discouraged the use of other electrotactile displays (87).

Electrotactile displays do not need to present pictographic information to be useful. In fact, even stimuli that are too weak to be felt can provide functional benefit. Low (subthreshold) levels of current (zero mean, white noise, 1 kHz bandwidth) delivered through large electrodes on the foot have been shown to enhance tactile sensitivity in the region between the electrodes in older adults (88). The effect is thought to result from the same principle of stochastic resonance as the vibrating insoles described above.



Figure 6. Mouse-mounted tactile display. Mouse-mounted tactile display with 64 electrodes. The stimulating current is controlled by a force sensor located under the electrodes, allowing the user to control stimulus intensity and reducing frequency of painful shocks. (Photo provided by Hiroyuki Kajimoto.)

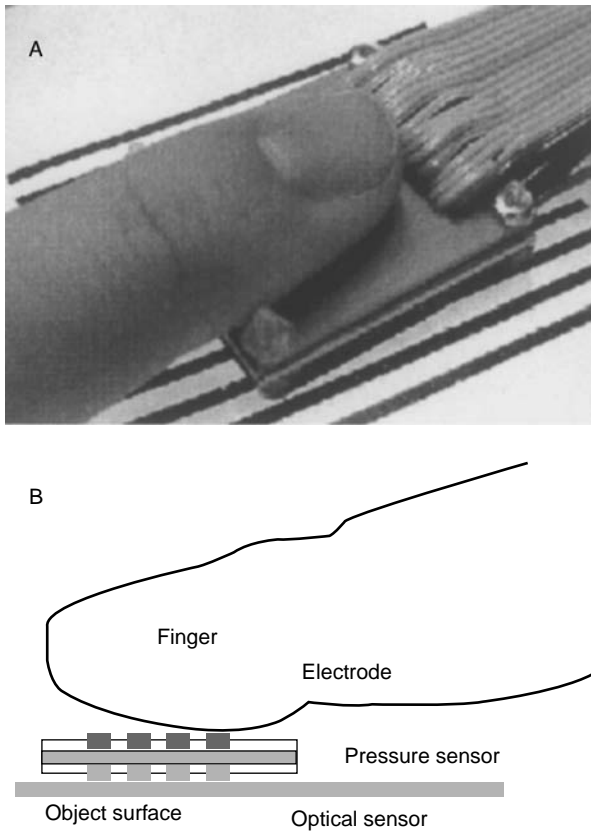


Figure 7. Smart Touch. Prototype of SmartTouch: (a) Visual image (black and white stripes) is captured by optical sensors and displayed through electrical stimulation. (b) Cross-section. (Photo provided by Hiroyuki Kajimoto.)

FUTURE DIRECTIONS

For most of its history, tactile stimulation has been developed and used primarily for sensory substitution or sensory augmentation systems for persons with sensory impairment. Many of these projects have been generally successful in restoring function and greater autonomy to their users, but target audiences have been small and large scale corporate investment has not been forthcoming. Although applications of tactile stimulation intended to help individuals with impairments are seen as important and will likely continue, it is also likely that applications with mass market appeal will drive the greater part of tactile stimulation technology development. Such applications are currently emerging in immersive virtual reality for entertainment and training purposes, teleoperation and telepresence for industry and medicine, and in personal communications.

There appear to be two separate trends in the technical progression of tactile displays. One trend is toward the development of fingertip tactile interfaces with greater sophistication and complexity for comfortably and portably communicating large amounts of data, such as surface texture. Such development will be important for applications where realistic surface perception is required, such as teleoperation, telesurgery, and immersive virtual reality.

The other trend is toward simpler, more intuitive and portable systems to interface with wearable computing or personal communication systems.

Sophisticated array designs for communicating realistic texture sensations will require greater collaboration between material scientists, engineers, physiologists, and psychophysicists. Optimizing the information transfer rates, increasing usability, and improving the naturalism of the displays will require improvements in all areas of display design. Both mechanical and electrical designs could benefit from greater reliability and ease of maintenance, as well as better power supply and control strategies. Mechanical displays require improved actuator technology for lower power consumption, more degrees of freedom, and greater actuator density. Ongoing efforts include adding lateral skin stretch to fingertip mechanical arrays as a means to reproduce novel sensations and provide improved power performance (89). Preliminary studies showed that subjects perceived small tangential displacements as intense as larger normal displacement (90). Novel dimensions such as slip and thermal stimulation could also be combined with mechanical or electrotactile displays (21).

Optimal display geometries beyond the common row and column design will also need to be addressed. Although at least three decades of work have provided us with a reasonable understanding of the physiology and psychophysics of single tactor (or, to a lesser extent, a single electrode) stimulation, mechanical display designs could be improved through better theoretical understanding of the mechanics of skin deformation and mechanoreceptor excitation. Recent modeling efforts provide a good start (91) and further studies in this area need to include the viscoelastic properties for dynamic simulation, the further refinement of relevant stimuli for each type of mechanoreceptor, as well as investigating the role of finger ridges in tactile neural coding.

Much work is also still needed to understand the physical, physiological, psychophysical, and cognitive interactions that exist between tactile stimuli presented at different sites. One example is the so-called rabbit effect, or sensory saltation, in which the perceived stimulus location falls between the locations of two stimulators (92–94). This effect can be elicited through mechanical (95,96) or electrical (97,98) stimulation and may be useful in increasing the effective density of stimulation sites. As well documented as saltation is for stimulators that lie on a common line, its application for arrays is not as straightforward as one might expect. When researchers using an array of nine electrodes arranged in a square on a subject's back attempted to use saltation to produce the sensation of a square being traced, the evoked sensation was actually that of a circle being traced (99). Clearly, when hundreds of stimulation sites are being used simultaneously, unanticipated interactions of this sort could be expected.

Substantial surround-masking effects have also been observed with two dimensional electrotactile arrays (100). Further research will be required to elucidate the impact of surround masking on fingertip array function. The precise application will most likely determine whether surround masking is beneficial (e.g., improving contrast and edge

detection, as in the retina) or detrimental (e.g., blurring the image or suppressing valuable tactile information).

Masking also occurs between two consecutive stimuli (101), and must be taken into account when predicting elicited percepts.

Continued improvements are also necessary in our understanding the way that the complex interactions between the various stimulus parameters influence stimulus perception. For example, the information transfer rate for a single stimulator can be improved through coherent modulation of stimulus frequency and amplitude, but when the frequency and amplitude are modulated independently, the information transfer rate of each channel is reduced due to cognitive processing interactions (103). Similar interactions may exist for all of the dimensions along which stimulus parameters can be modulated.

Many challenges still face designers of fingertip electro-tactile arrays. Physically, such an array must be held on the finger so that firm contact is maintained even during exploratory finger movements. Embedding the array in a custom-made glove should help solve this problem, but perspiration tends to accumulate between adjacent electrodes, providing a low impedance current pathway that circumvents the skin (103–106). Possibly, the material or chemical properties of the glove itself could be used to wick perspiration away or inhibit its secretion. Note that in an array where the electrodes positions are fixed relative to the fingertip, this problem is much less severe than when the finger is actively scanned over a flat array (108). Kajimoto et al. (108) suggested that for some applications, the electrodes could even be printed directly on the skin using conductive ink.

In addition to the physical problems, significant physiological and psychophysical problems of adaptation (habituation) and interaction must be addressed. It is well known that habituation, that is, a change in the perceived magnitude for constant stimulus amplitude, occurs over time with either mechanical (12,109–111) or electrical (68,112,113) stimulation of mechanoreceptors. Compensation for adaptation may be complicated when stimulating through an electro-tactile array since each electrode site will be stimulated at a different rate, and therefore the amount of habituation will vary across the array.

Recent evidence shows that a few hours of fingertip tactile stimulation results in cortical reorganization that improves spatial discrimination but impairs frequency discrimination performance (114). It is not yet known how long the impairment lasts and whether extensive long-term use of vibrotactile or electro-tactile stimulators could lead to learned functional sensory deficits.

The emerging field of wearable computing is giving rise to a new set of applications for tactile stimulation. Wearable computers for industrial and military maintenance applications have been reported for at least a decade and applications for the mass consumer market have been proposed, with several in current development. Tactile displays are considered viable alternatives to visual or auditory displays for wearable computers because they can be unobtrusive, socially discrete, and do not interfere with normal vision or sound. The goal of wearable tactile displays can be conceptually different than the sensory

substitution, virtual reality and teleoperation displays as discussed above. Most of the displays described above were concerned primarily with directly translating real or virtual visual, tactile, or audio information into tactile stimulation. The tactile information displays envisioned for wearable computers, on the other hand, present information that is not directly based on visual, tactile or audio information. Instead, they aim to communicate data on the state of the worn computer, the environment, or the person wearing it. A very simple example is the vibration of a cell phone in vibrate mode. The vibration is intended to alert the user to an incoming call, not to communicate the voice message itself. MacLean and Enriquez suggested that such tactile alerts will become more common and more complex, taking into account optimization along several perceptual dimensions to optimize information transfer (115). They coined the term “haptic icons” to refer to brief synthetic haptic or tactile signals to convey information such as event notification, identity, content or state.

According to Gemperle et al. (116), wearable tactile displays must be lightweight, silent, small, and physically discreet. They must also have low power requirements, be able to be felt through the clothing they are embedded in or worn on, and they must be held tightly enough to the body to maintain reliable contact. Initial designs would also benefit from being flexible enough to be used in multiple applications with little modification.

One wearable tactile display application currently being pursued by several independent groups is an aid for a personal navigation. Tactile navigation aids have been proposed for pilots (117), drivers (118), scuba divers (119), walkers or hikers (116), astronauts (120,121) and the blind (122). The basic principle is that vibrating tactors embedded in the user’s clothing signal either a particular direction (e.g., north, up, the direction to return the vehicle) or the distance to and direction of the next turn or position correction required to reach a destination.

Other applications of wearable tactile displays include providing silent and private alerts, socially subtle transmission of information, providing biofeedback of physiological states such as blood sugar or blood alcohol levels. Future work includes the development of standardized tactile display clothing elements (123), and requires continued research on where to define the spaces on the human body where solid and flexible forms can rest without interfering with fluid human movement. The widespread use of wearable tactile displays depends on consideration not only of function, but of comfort, mobility and social acceptance factors.

BIBLIOGRAPHY

1. Loomis JM, Lederman SJ. Tactual perception. In: Boff KR, Kaufman L, Thomas JP, editors. *Handbook of Perception and Human Performance*: Vol. II, Cognitive Processes and Performance. New York: John Wiley & Sons; 1986. p 31-1 to 31-41.
2. Perez A, et al. What is the value of telerobotic technology in gastrointestinal surgery? *Surg Endosc* 2003;17(5):811–813.
3. Rassweiler J, Binder J, Frede T. Robotic and telesurgery: Will they change our future? *Curr Opin Urol* 2001;11(3):309–320.

4. Halata Z. The mechanoreceptors of the mammalian skin: Ultrastructure and morphological classification. *Adv Embryol Cell Biol* 1975;50:1-77.
5. Chouchkov C. Cutaneous receptors. *Adv Anatomy Embryol Cell Biol* 1978;54(5):1-62.
6. Iggo A, Andres KH. Morphology of cutaneous receptors. *Ann Rev Neurosci* 1982;5:1-31.
7. Johnson KO, Hsiao SS, Yoshioka T. Neural coding and the basic law of psychophysics. *Neuroscientist* 2002;8(2):111-121.
8. Srinivasan MA, Whitehouse JM, LaMotte RH. Tactile detection of slip: Surface microgeometry and peripheral neural codes. *J Neurophysiol* 1990;63(6):1323-1332.
9. Johansson RS, Landstrom U, Lundstrom R. Responses of mechanoreceptive afferent units in the glabrous skin of the human hand to sinusoidal skin displacements. *Brain Res* 1982;244(1):17-25.
10. Ochoa J, Torebjörk. Sensations evoked by intraneural microstimulation of single mechanoreceptor units innervating the human hand. *J Physiol* 1983;342:633-654.
11. Kunesch E, et al. Somatosensory evoked potentials elicited by intraneural microstimulation of afferent nerve fibers. *J Clin Neurophysiol* 1995;12:476-487.
12. Bolanowski SJ, et al. Four channels mediate the mechanical aspects of touch. *JASA* 1988;84(5):1680-1694.
13. Johnson KO. The roles and functions of cutaneous mechanoreceptors. *Curr Opin Neurobiol* 2001;11(4):455-461.
14. Olausson H, Wessberg J, Kakuda N. Tactile directional sensibility: peripheral neural mechanisms in man. *Brain Res* 2000;866(1-2):178-187.
15. Olausson H, et al. Unmyelinated tactile afferents signal touch and project to insular cortex. *Nat Neurosci* 2002;5(9):900-904.
16. Wessberg J, et al. Receptive field properties of unmyelinated tactile afferents in the human skin. *J Neurophysiol* 2003;89(3):1567-1575.
17. Pasquero J. A tactile display using lateral skin stretch. In: *Mechanical Engineering*. Montreal, Canada: McGill University; 2003.
18. Aiello GL. Multidimensional electrocutaneous stimulation. *IEEE Trans Rehabil Eng* 1998;6(1):95-101.
19. Kaczmarek KA, Haase SJ. Pattern identification and perceived stimulus quality as a function of stimulation waveform on a fingertip-scanned electro tactile display. *IEEE Trans Neural Syst Rehabil Eng* 2003;11(1):9-16.
20. Ino S, Shimizu S, Odagawa T, Sato M, Takahashi M, Izumi T, Ifukube T. A Tactile Display for Presenting Quality of Materials by Changing the Temperature of Skin Surface. In: *IEEE International Workshop on Robot and Human Communication*. Tokyo: 1993.
21. Benali-Khoudja M, Hafez M, Alexandre J-M, Kheddar A, Moreau V. VITAL: A VibroTActiLe Interface with Thermal Feedback. In: *2004 IEEE International Conference on Robotics & Automation*; New Orleans, LA: 2004.
22. Caldwell G, Gosney C. Enhanced tactile feedback (teletaction) using a multi-functional sensory system. In: *IEEE International Conference on Robotics and Automation*; Atlanta, GA: 1993.
23. Collins JJ, Imhoff TT, Grigg P. Noise-enhanced tactile sensation. *Nature* 1996;383(6603):770.
24. Liu W, et al. Noise-enhanced vibrotactile sensitivity in older adults, patients with stroke, and patients with diabetic neuropathy. *Arch Phys Med Rehabil* 2002;83(2):171-176.
25. Khaodhiar L, et al. Enhancing sensation in diabetic neuropathic foot with mechanical noise. *Diabetes Care* 2003;26(12):3280-3283.
26. Dennerlein J, Millman P, Howe R. Vibrotactile feedback for industrial telemanipulators. In: *Sixth Annual Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, ASME International Mechanical Engineering Congress and Exposition. Dallas: 1997.
27. Dennerlein J, Shahion E, Howe R. Vibrotactile feedback for an underwater teleoperated robot. In: *International Symposium on Robotics with Applications (ISORA)*. Maui Hawaii: 2000.
28. Debus T, et al. Multichannel vibrotaction display for sensory substitution during teleoperation. In: *Telemanipulator and teleresence technology VIII*. Newton, MA: SPIE; 2001.
29. Collins CC. Tactile television—Mechanical and electrical image projection. *IEEE Trans Man-Mach Sys* 1970;MMS-11:65-71.
30. Linvill JG, Bliss JC. A direct translation reading aid for the blind. *Proceedings of the IEEE* 1966;54(1):40-51.
31. Stein DK. The Optacon: Past, present, and future. In: *The Braille Monitor*; 1998.
32. Saunders FA. Information transmission across the skin: High-resolution tactile sensory aids for the deaf and the blind. *Int J Neurosci* 1983;19(1-4):21-28.
33. Kaczmarek K, et al. A tactile vision-substitution system for the blind: Computer-controlled partial image sequencing. *IEEE Trans BME* 1985;32(8):602-608.
34. Perez CA, Holzmann CA, Jaeschke HE. Two-point vibrotactile discrimination related to parameters of pulse burst stimulus. *Med Biol Eng Comput* 2000;38(1):74-79.
35. Verrillo RT, Fraioli AJ, Smith RL. Sensation magnitude of vibrotactile stimuli. *P&P* 1969;6:366-372.
36. Gescheider GA, et al. Vibrotactile intensity discrimination measured by three methods. *JASA* 1990;87(1):330-338.
37. Summers IR, et al. Vibrotactile and electro tactile perception of time-varying pulse trains. *JASA* 1994;95(3):1548-1558.
38. Johansson RS, Landstrom U, Lundstrom R. Sensitivity to edges of mechanoreceptive afferent units innervating the glabrous skin of the human head. *Brain Res* 1982;244(1):27-35.
39. Perez CA, et al. Power requirements for vibrotactile piezoelectric and electromechanical transducers. *Medical & Biological Engineering & Computing* 2003;41(6):718-726.
40. Hayward V, Cruz-Hernandez M. Tactile display device using distributed lateral skin stretch. In: *Haptic Interfaces for Virtual Environment and Teleoperator Systems Symposium*. Orlando, Florida, USA: ASME IMECE2000; 2000.
41. Phillips JR, Johnson KO. Tactile spatial resolution. III. A continuum mechanics model of skin predicting mechanoreceptor responses to bars, edges, and gratings. *J Neurophysiol* 1981;46(6):1204-1225.
42. Vincent FVJ. *Structural Biomaterials*. Princeton University Press; 1991.
43. Caldwell DG, Sagarakis NT, Giesler C. An integrated tactile/shear feedback array for stimulation of finger mechanoreceptors. In: *IEEE International Conference on Robotics & Automation*; Detroit: Michigan; 1999.
44. Moy G, et al. Human psychophysics for teletaction system design. *Haptics-e* 2000;1(3).
45. Srinivasan MA, LaMotte RH. Tactile discrimination of softness. *J Neurophysiol* 1995;73(1):88-101.
46. LaMotte RH, Srinivasan MA. Neural encoding of shape: responses of cutaneous mechanoreceptors to a wavy surface stroked across the monkey fingerpad. *J Neurophysiol* 1996;76(6):3787-3797.
47. Fischer H, Trapp R. Tactile optical sensor for use in minimal invasive surgery. *Stud Health Technol Inform* 1996;29:623-629.

48. Strong RM, Troxel DE. An electrotactile display. *IEEE Trans Man Machine Systems* 1970;MMS-11(1):72–79.
49. Tang H, Beebe DJ. A microfabricated electrostatic haptic display for persons with visual impairments. *IEEE Trans Rehabil Eng* 1998;6(3):241–248.
50. Pawluk DT, et al. Control and pattern specification for a high density tactile array. In: *Proceedings of the ASME Dynamic Systems and Control Division*; New York, NY: ASME; 1998.
51. Howe R, Kontarinis D, Peine W. Shape memory alloy actuator controller design for tactile displays. In: *Proc of the 34th IEEE Conf. on Decision and Control*; IEEE 1995.
52. Hasser CJ, Weisenberger JM. Preliminary evaluation of a shape-memory alloy tactile feedback display. *DSC: Advances in Robotics, Mechatronics and Haptic Interfaces* 1993;49:73–80.
53. Cohn MB, ML, Fearing RS. Tactile feedback for teleoperation. In: *SPIE Conf. 1833, Telemanipulator Technology*; Boston, MA: SPIE; 1992.
54. Wagner CR, Lederman SJ, Howe RD. A tactile shape display using RC servomotors. In: *The Tenth Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. Orlando, FL: 2002.
55. Ghodssi R, et al. Development of a tangential tactor using a LIGA/MEMS linear microactuator technology. In: *Micro-Electro-Mechanical systems; International Mechanical Engineering Congress and Exposition*. New York, NY: 1998.
56. Asamura N, Yokoyama N, Shinoda H. Selectively stimulating skin receptors for tactile display. *IEEE Computer Graphics and Applications* 1998;98:32–37.
57. Blamey PJ, Clark GM. Psychophysical studies relevant to the design of a digital electrotactile speech processor. *JASA* 1987;82(1):116–125.
58. Cowan RSC, et al. Perception of multiple electrode stimulus patterns: Implications for design of an electrotactile speech processor. *JASA* 1991;89(1):360–369.
59. Bach-y-Rita P. Tactile vision substitution: Past and future. *Intern J Neurosci* 1983.
60. Collins CC. Electrotactile visual prosthesis. In: *Hambrecht TF, Reswick JB, editors. Functional Electrical Stimulation*. New York: Marcel Dekker; 1977. p 189.
61. Rohland TA. Sensory feedback in upper-limb prosthetic systems. *Inter-Clinic Information Bulletin* 1974;13(9):1–8.
62. Schmid H. The importance of information feedback in prostheses for the upper limbs. *Prosthetics and Orthotics International* 1977;1:21–24.
63. Shannon GF. Sensory feedback for artificial limbs. *Medical Progress Technology* 1979;6:73–79.
64. Prior RE, et al. Supplemental sensory feedback for the VA/NU myoelectric hand. Background and preliminary designs. *Bull Prosth Res* 1976;11:171–191.
65. Scott RN, et al. Sensory-feedback system compatible with myoelectric control. *Med Biol Eng Comput* 1980;18:65–69.
66. Scott RH. Feedback in myoelectric prostheses. *Clin Orth Related Res* 1990;256:58–63.
67. Lovely DF, Hudgins BS, Scott RN. Implantable myoelectric control system with sensory feedback. *Med Biol Eng & Comput* 1985;23:87–89.
68. Szeto AYJ, Lyman J. Comparison of codes for sensory feedback using electrocutaneous tracking. *Ann Biomed Eng* 1977;5:367–383.
69. Kawamura J, et al. Sensory feedback systems for the lower-limb prosthesis. *J Osaka Rosai Hospital* 1981;5(2):104–112.
70. Sabolich JA, Ortega GM. Sense of feel for lower-limb amputees: A phase-one study. *J Prosthetics Orthotics* 1994;6:36–41.
71. Schmid HP, Bekey GA. Tactile information processing by human operators in control systems. *IEEE Trans Syst Man Cybern* 1978;8(12):860.
72. VanDoren CL, Riso RR, Milchus K. Sensory feedback for enhancing upper extremity neuromuscular prostheses. *J Neurol Rehab* 1991;5:63–74.
73. Collins CC, Madey JM. Tactile sensory replacement. In: *Proc San Diego Biomed Symp* 1974; 15–26.
74. Cowan R, et al. Improved electrotactile speech processor: Tickle Talker. *Ann Otol Rhinol Laryngol Suppl* 1995;166: 454–456.
75. Saunders FA. Electrocutaneous displays. In: *Cutaneous communication systems and devices*. Monterey, CA: Psychonomic Society; 1974.
76. Tyler M, et al. A New Electrotactile Prosthesis for the Blind. *Unitech Res Inc.*; 1993.
77. Kaczmarek KA, Tyler ME. Effect of electrode geometry and intensity control method on comfort of electrotactile stimulation on the tongue. In: *ASME Dyn Sys Contr Div. Orlando, FL: ASME*; 2000.
78. Sampaio E, Maris S, Bach-y-Rita P. Brain plasticity: ‘visual’ acuity of blind persons via the tongue. *Brain Res* 2001;908(2):204–207.
79. Tyler M, Danilov Y, Bach YRP. Closing an open-loop control system: vestibular substitution through the tongue. *J Integr Neurosci* 2003;2(2):159–164.
80. Saunders FA. Recommended procedures for electrocutaneous displays. In: *Hambrecht TF, Reswick JB, editors. Functional Electrical Stimulation*. New York: Marcel Dekker; 1977. p 303.
81. Kaczmarek KA, et al. Electrotactile and vibrotactile displays for sensory substitution systems. *IEEE Trans Biomed Eng* 1991;BME-38(1):1–16.
82. Poletto CJ, Van Doren CL. Elevating pain thresholds in humans using depolarizing prepulses. *IEEE Trans Biomed Eng* 2002;49(10):1221–1224.
83. Poletto CJ. Fintertip electrocutaneous stimulation through small electrodes. In: *Biomedical Engineering*. Cleveland, OH: Case Western Reserve University; 2001.
84. Kajimoto H, et al. Electrocutaneous display as an interface to a virtual tactile world. In: *Virtual Reality Conference*. Yokohama, Japan: IEEE Computer Society; 2001.
85. Kaczmarek KA, Haase SJ. Pattern identification as a function of stimulation current on a fingertip-scanned electrotactile display. *IEEE Trans Neural Syst Rehabil Eng* 2003;11(3):269–275.
86. Asamura N, Yokoyama N, Shinoda H. A method of selective stimulation to epidermal skin receptors for realistic touch feedback. In: *IEEE Conference on Virtual Reality*. 1999.
87. Kajimoto H, et al. Electro-tactile display with force feedback. *World Multiconference on Systemics, Cybernetics and Informatics*. Orlando, FL: 2001.
88. Dhruv NT, et al. Enhancing tactile sensation in older adults with electrical noise stimulation. *Neuroreport* 2002;13(5):597–600.
89. Pasquero J, et al. Display of virtual Braille dots by lateral skin deformation: A pilot study. In: *Eurohaptics*. Munich, Germany: 2004.
90. Biggs J, Srinivasan MA. Tangential versus normal displacements of skin: Relative effectiveness for producing tactile sensations. In: *10th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. Orlando FL: IEEE Computer Society; 2002.
91. Dandekar K, Raju BI, Srinivasan MA. 3-D finite-element models of human and monkey fingertips to investigate the mechanics of tactile sense. *J Biomech Eng* 2003;125(5):682–691.

92. Geldard FA, Sherrick CE. The cutaneous "rabbit": A perceptual illusion. *Science* 1972;178(57):178-179.
93. Geldard FA, Sherrick CE. The cutaneous saltatory area and its presumed neural basis. *Perception and Psychophys* 1983;33(4):299-304.
94. Geldard FA. Saltation in somesthesia. *Psychological Bulletin* 1982;92(1):136-175.
95. Cholewiak RW. Exploring the condition that generate a good vibrotactile line. In: *Psychonomic Society Meetings*. Los Angeles: 1995.
96. Cholewiak RW, Sherrick CE, Collins AA. Studies of saltation. In: *Princeton Cutaneous Research Project*. Princeton University, Department of Psychology; 1996.
97. Tanie K, et al. Basic study on discriminability of mental location of electrocutaneous phantom sensation. *Transactions of the Society of Instrument and Control Engineers* 1979;15(4):505-512.
98. Tanie K, et al. Information transmission characteristics of two-dimensional electrocutaneous phantom sensation. *Transactions of the Society of Instrument and Control Engineers* 1980;16(5):732-739.
99. Tan HZ, Lim A, Traylor R. A psychophysical study of sensory saltation with an open response paradigm. In: *DSC-ASME Dynamic Systems and Control Division*. ASME; 2000.
100. Szeto AY, Saunders FA. Electrocutaneous stimulation for sensory communication in rehabilitation engineering. *IEEE Trans Biomed Eng* 1982;29(4):300-308.
101. Craig JC. Identification of scanned and static tactile patterns. *Percept Psychophys* 2002;64(1):107-120.
102. Poletto CJ, VanDoren CL. Perceptual interactions between electrocutaneous loudness and pitch. *IEEE Transactions Rehab Eng* 1995;3(4):334-342.
103. Melen RD, Meindl JD. Electrocutaneous stimulation in a reading aid for the blind. *IEEE Trans Biomed Eng* 1971;18(1):1-3.
104. Tyler M, et al. A dynamic multi-point electrohaptic display: The effects of peak voltage, pulse-phase width, number of pulses, and geometric area on the perception of haptic threshold. *IEEE Transactions on Rehabilitation Engineering*.
105. Kaczmarek K. Tactile displays. In: Barfield W, TF III, editors. *Virtual Environments and Advanced Interface Design*. Oxford: Oxford University Press; 1995.
106. Kaczmarek K, Tyler ME, Bach-y-Rita P. Electrohaptic display on the fingertips: Preliminary results. In: *16th Annual International Conference IEEE Eng Med Biol Soc*; Baltimore, MD: IEEE; 1994.
107. Kaczmarek K. Sensory augmentation and substitution. *CRC Handbook of Biomedical Engineering*. 1995.
108. Kajimoto H, et al. SmartTouch: Electric skin to touch the untouchable. *IEEE Comput Graph Appl* 2004;24(1):36-43.
109. Gescheider GA, et al. Vibrotactile forward masking: Psychophysical evidence for a triplex theory of cutaneous mechanoreception. *JASA* 1985;78(2):534-543.
110. Hahn JF. Tactile adaptation. in *The Skin Senses*. 322-326.
111. Verrillo RT, et al. Vibrotactile masking: Effects of one- and two-site stimulation. *P&P* 1982;33(4):379-387.
112. Szeto AYJ, Farrenkopf GR. Optimization of single electrode tactile codes. *Ann Biomed Eng* 1992;20:647-665.
113. Kaczmarek KA. Electrohaptic adaptation on the abdomen: Preliminary results. *IEEE Trans Rehabil Eng* 2000;8(4):499-505.
114. Hodzic A, et al. Improvement and decline in tactile discrimination behavior after cortical plasticity induced by passive tactile coactivation. *Neuroscience* 2004;24(2):442-446.
115. MacLean K, Enriquez M. *Perceptual Design of Haptic Icons*. In: *EuroHaptics 2003*. Dublin, UK: 2003.
116. Gemperle F, Ota N, Siewiorek D. Design of a wearable tactile display. In: *Fifth International Symposium on Wearable Computers*. Zürich, Switzerland: 2001.
117. van Erp JBF, Veltman JA, van Veen HAHC. A tactile cockpit instrument to support altitude control. In: *Human Factors and Ergonomic Society 47th Annual Meeting*. 2003.
118. van Erp JB, Padmos P. Image parameters for driving with indirect viewing systems. *Ergonomic* 2003;46(15):1471-1499.
119. McTrusty T, Walters C. *Swimmer Inshore Navigation System (SINS) Tactile Situation Awareness System (TSAS) Test Report*. Panama City, FL: Coastal Systems Station; 1997.
120. Rochlis JL, Newman DJ. A tactile display for international space station (ISS) extravehicular activity (EVA). *Aviat Space Environ Med* 2000;71(6):571-578.
121. van Erp JBF, van Veen HAHC. A multipurpose tactile vest for astronauts in the International Space Station. In: *Eurohaptics*. Dublin, Ireland: 2003.
122. Ross D, Blasch B. Development of a Wearable Computer Orientation System. *Personal Ubiquitous Computing* 2002;6(1):49-63.
123. Toney A, et al. A shoulder pad insert vibrotactile display. In: *Seventh International Symposium on Wearable Computers*. 2003.
124. Johansson RS, Vallbo ÅB. Tactile sensory coding in the glabrous skin of the human hand. *Trends in Neurosci* 1983;6(1):27-32.

See also FUNCTIONAL ELECTRICAL STIMULATION; HEAT AND COLD, THERAPEUTIC; SKIN, BIOMECHANICS OF; TRANSCUTANEOUS ELECTRICAL NERVE STIMULATION (TENS)

TEETH, PROPERTIES OF. See BONE AND TEETH, PROPERTIES OF.

TELEMETRY. See BIOTELEMETRY.

TELERADIOLOGY

H.K. HUANG
University of Southern
California
Los Angeles, California

INTRODUCTION

Telemedicine and teleradiology have become increasingly important as our country's healthcare delivery system gradually changes from fee-for-service to managed, capitated care. During the past several years, we have seen the trend of primary care physicians joining health maintenance organizations (HMOs). These HMOs purchase smaller hospitals and form hospital groups under the umbrella of HMOs. Also, academic institutions form consortia to compete with other local hospitals and HMOs. This consolidation allows the elimination of duplication and the streamlining of healthcare services among hospitals. As a result, costs are reduced, but at the same time because of the downsizing, the number of experts available for service also decreases. Utilization of telemedicine and

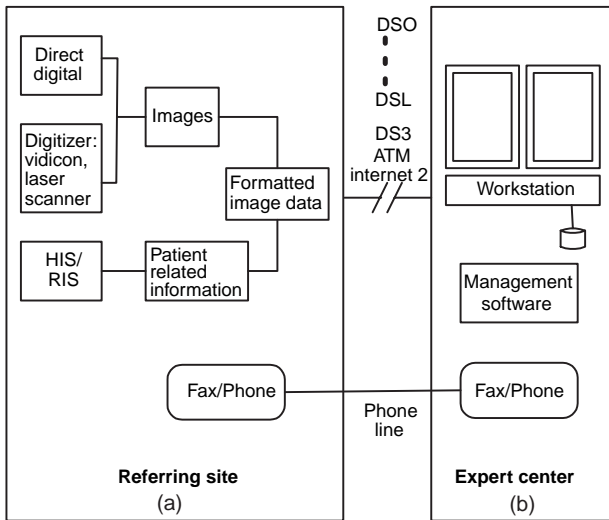


Figure 1. A generic teleradiology setup. (a) Referring site, (b) expert center.

teleradiology is a method to alleviate the diminishing of experts, streamline the diagnosis process, and save health-care costs.

Teleradiology is a subset of telemedicine operation focusing in remote diagnosis of medical images. Teleradiology utilizes computer, display, and telecommunication technologies for radiologists to make remote diagnosis from radiological images generated at distant examination sites. The diagnostic report is sent to the examination site where a primary physician can provide proper treatments to the patient immediately. Figure 1 shows a generic teleradiology set-up illustrating that teleradiology is not a single medical device or an instrument, instead, it is a system integration of various imaging devices using communication technology and system software connecting multiple imaging centers and expert centers together (1-3). Dependent on the required turn around time in obtaining the diagnosis from the examination site, the expert center has three reading modes: teleradiology, teleconsultation, and telemanagement, which are shown in Fig. 2. These reading modes dictate the communication requirements of transmitting the images between the sites. Teleradiology operation can be very simple or extremely complicated. In the simple case, a radiology resident may send an image set from a CT (computed tomography) scanner using low quality teleradiology equipment and slow speed communication technology in the evening to the radiologist's home for a second opinion. This type of teleradiology operation does not require highly sophisticated equipment. A conventional telephone and simple desktop personal computer with modem connection and display software are sufficient to perform the teleradiology operation. This type of application originated in early 1970.

The complicated teleradiology operation can have different models starting from simple to complicated in ascending order as shown in Table 1. The complications occur when the current examination requires historical images for comparison, and when the radiologist needs information from the radiology information system (RIS) to

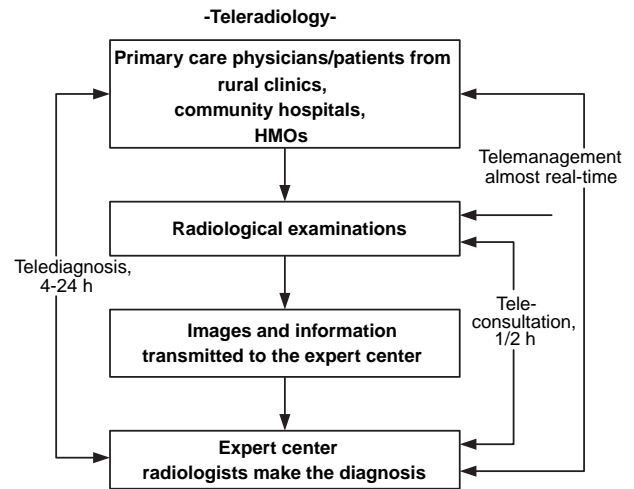


Figure 2. The expert center in teleradiology with three reading modes: telediagnosis, teleconsultation, and telemanagement. These three reading modes dictate the image communication requirements.

make a diagnosis. In addition, complications arise when the images and the corresponding diagnosis report are required to be archived to the patient data file. Teleradiology is relatively simple to operate when neither retrieval nor archive of previous information and images is required. However, when both archive and retrieval are required, the operation becomes extremely complicated.

TELERRADIOLOGY AND PACS

Picture Archiving and Communication System (PACS) is a hospital integrated imaging management system developed in the early 1990s (Fig. 3) (4). The infrastructure of PACS is shown in the upper three rows. Two types of servers in the bottom of the figure are for various PACS applications. When teleradiology service requires patient's historical images as well as related information, technologies used in both teleradiology and PACS become very similar. Table 2 shows technologies used in teleradiology and PACS, the major differences are in image capture, communication, and storage. Some current teleradiology operations still use a film digitizer as the primary method of converting a projection film image-to-digital format, although the trend is moving toward direct digital capture. In PACS, direct digital image capture using Digital Imaging Communication in Medicine (DICOM) standard format is mostly used. In networking, teleradiology uses slower speed wide area networks (WAN) compared with the higher speed local area network (LAN) used in PACS.

Table 1. Four Models of Teleradiology Operation According to Its Complexity

	Historical Images/RIS	Archive
Most simplistic	No	No
Simplistic	Yes	No
Complicated	No	Yes
Most complicated	Yes	Yes

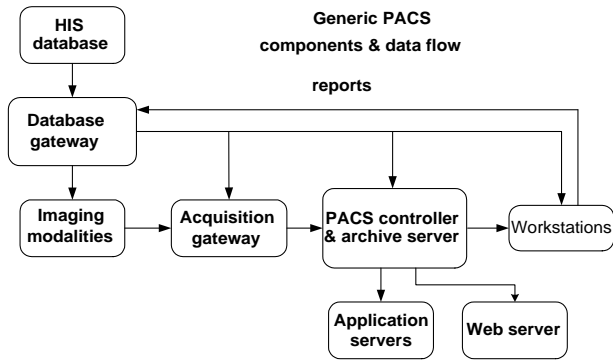


Figure 3. A generic PACS components and data flow. The upper three rows are the infrastructure, the bottom row is PACS applications.

Table 2. Differences in Technology Used between Teleradiology and PACS

Function	Telerad	PACS
Image capture	Digitizer	DICOM
Display technology	Same	Same
Networking	WAN	LAN
Storage	Short	Long
Compression	Yes	May be

In teleradiology, image storage is mostly short term, whereas in PACS it is long term. Teleradiology relies heavily on image compression, whereas PACS may or may not.

PACS and teleradiology use medical images shown in Table 3 for radiologists to make diagnosis. In Table 3, the first and second columns give types and sizes of some common medical images. In clinical applications, one examination is composed of many images of different views and anatomical emphasis, since a single image is generally not sufficient for making the proper diagnosis. In general, a typical examination generates between 10 and 20 MB, although some current CT and MR imaging modalities can generate up to 3000 images per examination. The fourth column shows an average size of one typical exam-

ination in each of these image modalities. The high extreme is in digital mammography, which routinely requires 160 MB. To transmit 160 MB of images through WAN for teleradiology requires a very high bandwidth communication technology.

WHY DO WE NEED TELERADIOLOGY?

The managed care trend in healthcare delivery expedites the formation of teleradiology expert centers. However, even without the healthcare reform, teleradiology is still an extremely important component in radiology practice for the following reasons: First, teleradiology secures images for radiologists to read so that no images will be accidentally lost in transit. Second, teleradiology reduces the reporting cycle time after the image is generated. Third, since radiology is subdivided into many subspecialties, a general radiologist requires a specialist's second opinion on occasion. The availability of teleradiology will facilitate seeking a second opinion. Fourth, teleradiology increases radiologists' income since no images would accidentally be lost and subsequently not reported. The healthcare reform adds two more reasons. (1) It saves healthcare costs since an expert center can serve multiple sites reducing the number of radiologists required. (2) It improves the efficiency and effectiveness of healthcare because the report turn-around time would be reduced and there would be no loss of image (5).

TELERADIOLOGY COMPONENTS

A generic schematic of teleradiology depicted in Fig. 1 shows teleradiology components and their connections. Modalities generating images in teleradiology applications include CT, MR, CR, DR, US, NM, DSA—digital fluorography (DF), and film digitizer. Their respective image and examination file sizes Images are shown in Table 3. These acquisition devices are first generated from the examination site and then sent through communication networks to the expert center if they are already in digital format. Or, if these images are stored on films, then they need to be digitized by a film scanner at the examination site.

Table 3. Data Size of Some Common Medical Images and Examinations

Instrument	One Image, bits	No. of Images/Exam	One Examination
Nuclear medicine (NM)	128 × 128 × 12	30–60	1–2 MB
Magnetic resonance imaging (MRI)	256 × 256 × 12	60–3000	8 MB up
Ultrasound (US) ^a	512 × 512 × 8(24)	20–240	5–60 MB
Digital subt. Angiography (DSA)	512 × 512 × 8	15–40	4–10 MB
Digital microscopy	512 × 512 × 8	1	0.26 MB
Digital color microscopy	512 × 512 × 24	1	0.79 MB
Color light images	512 × 512 × 24	4–20	3–15 MB
Computed tomography (CT)	512 × 512 × 12	40–3000	20 MB up
Computed/digital radiog (CR/DR)	2048 × 2048 × 12	2	16 MB
Digitized X rays	2048 × 2048 × 12	2	16 MB
Digital mammography	4000 × 5000 × 12	4	160 MB

^aDoppler US with 24 bit color images.

Image Capture

In image capture, if the original image data are on film, then either a video frame grabber or a laser film digitizer is used to convert them to digital format. A video frame grabber produces low quality digital images, but is faster and cheaper. On the other hand, laser film digitizers produce very high quality digital data, but take longer and cost more compared to the video frame grabber. During the past several years, direct Digital Imaging and Communication in Medicine (DICOM) standard output images from CR, DR, CT, and MR have been used extensively in teleradiology.

Data Reformatting

After images are captured, it is advantageous to convert these images and related data to industry standards because multiple vendors' equipment can be used in the teleradiology chain. The two common standards used in medical imaging industry are the DICOM (6) for images and Health Level 7 (HL7) (7) for textual data. The DICOM standard includes both the image format as well as the communication protocols based on the standard TCP/IP. Health level 7 is the standard for textual data, it uses the TCP/IP communication protocols.

Image Storage

At the expert center, a local storage device is used before images are displayed. The capacity of this device can range from several hundred megabytes to many gigabytes. A long-term archive, such as a small DLT (digital linear tape) library, is used for teleradiology applications that require historical images and diagnostic reports, related patient information, and current images and diagnosis.

Display Workstation

For an inexpensive teleradiology system, a low cost 512-line single monitor can be used for displaying images. However, high resolution multimonitor display workstations are needed for the primary diagnosis.

Table 4 shows the specifications of a 2000- and a 1600-line workstation used for teleradiology primary readings. These state-of-the-art technology diagnostic workstations, use two monitors with over 2 GB of local storage, and can display images and reports from the local storage in 1–2 s. A 2000-line LCD monitor workstation costs from \$20,000 to 30,000, and a 1,600 line costs from \$15,000 to \$20,000. User-friendly image display software is necessary for easy and convenient use by the radiologist at the workstation.

Table 4. Specifications of High-End 2000 and 1600 Line Workstations for Teleradiology

Two LCD Monitors
1–2 week local storage for current + previous exams
1–2 s display of images and reports from local storage
HL7 and DICOM conformance
Simple image processing functions

Table 5. Transmission Rate of Current Wide Area Network Technology

DS-0	56 kbits/s
DS-1	56 to (24 × 56) kbits/s
DSL	144 kbits/s – 8 Mbits/s
DS-1 (T1)	1.5 Mbits/s
ISDN	56 kbits/s to 1.5 Mbits/s
DS-3 (T3)	28 DS-1 = 45 Mbits/s
ATM (OC-3)	155 Mbits/s and up
Internet-2	100 Mbits/s and up

Communication Networks

An important component in teleradiology is communication networks used for the transmission of images and related data from the acquisition site to the expert center for diagnosis. Since most teleradiology applications are not within the same hospital complex, but through inter-healthcare facilities in metropolitan areas or at longer distances, the communication technology involved requires wide area network (WAN) technology. Wide area network can be wireless or with cables. In wireless WAN, some technologies available are microwave transmission and communication satellites. Wireless WAN has not been used extensively in teleradiology due to its higher cost. Table 5 shows cable technology available in WAN from the low communication rate DS-0 with 56 kb/s, to DSL (Digital Subscriber Line, 144 kb/s to 8 Mb/s, depending on data traffic and the subscription), T-1 and T-3 lines starting from 1.5 Mb/s, to very high broadband communication DS-3 with 45 Mb/s (8). These WAN technologies are available through either a long distance or local telephone carrier, or both. The cost of using WAN is a function of transmission speed and the distance between sites. Thus, within a fixed distance, for a DS-0 line with low transmission rate, the cost is fairly low compared to DS-3, which is much faster, but very expensive. Most of the private lines, for example, T-1 and T-3, are point-to-point and the cost depends on the distance between connections. Table 6 gives an example showing the relative cost of the DSL and the T-1 between University of Southern California and St. John's Health Center ~ 15 miles apart in the Greater Los Angeles Metropolitan Area.

Table 6 demonstrates that the initial investment for the DSL is minimal since the WAN carrier pays for the DSL Modem for the network connection. The lowest

Table 6. WAN Cost Using DSL (144 kB/s–8 MB/s) and T-1 (1.5 MB/s) between USC and St. John's Health Center—20 miles

DSL		T-1	
Up front Investment	Minimal	Up Front Investment	\$ 5000
Modems (2)	None	T1 DSU/CSU ^a	
		WAN interface (2)	
		Router (2)	\$ 4000
Installation (2)	Minimal	T-1 Installation	\$ 1000
Monthly charge:	\$40	T-1 Monthly Charge:	\$600
(the lowest rate)			

^aDSU/CSU: Data service unit/ Channel service unit as of June, 2003.

monthly cost is ~\$40/month. On the other hand, for T-1 service, the up front investment is \$4000 for the two T-1 routers and \$1000 for installation. The monthly cost is \$600. The up-front investment for the T-1 is much higher than DSL, and for longer distances, its monthly charge is expensive. For example, the charge between Los Angeles and Washington, D.C. for a T-1 line could be as high as \$10,000/month. However, T-1 is a point-to-point private line, and it guarantees its 1.5 MB/s specification, and provides communication security. The disadvantages of DSL are (1) it is through shared networks, and hence has no security; (2) its performance depends on the load of the DSL carrier at that moment; and (3) it is not available everywhere. Using T-1 and DSL for teleradiology is very popular. Some larger IT (Information Technology) companies lease several T-1 lines from telephone carriers and sublease portions of them to smaller companies for teleradiology applications.

Another wide area network listed in Table 5, Internet 2 (I2) technology, is ideal for teleradiology application because of its speed of transmission and low cost of operation after the site is connected to the I2 backbone (9). The I2 network is a national infrastructure of high speed communication backbones [> 10 GB/s using gigabit switches and Asynchronous Mode Technology (ATM)] supported by the National Science Foundation (NSF), currently consisting of the vBNS (very high performance Backbone Network Service), the CalREN-2 (California Research and Education Network), and the Abilene. In the global level, vBNS, Abilene and CalREN-2, provide readily available high speed backbones and administrative infrastructure. In the local level, the users have to learn how to connect the hospital and clinic environments to these backbones. Table 7 shows the current performance of the I2 between some sites in the United States. The advantages of using Internet 2 for teleradiology are its high speed and low operational cost once the local site is connected to the backbones. The disadvantages are (1) the local site has to upgrade its conventional Internet infrastructure to be compatible with the high speed I2 performance, which is costly; (2) not enough experts know how to connect from the radiology department to the backbone; and (3) I2 is not yet open for commercial use.

Table 7. Current Internet 2 Performance between Sites in the United States

Test Sites	Response Time (32 Bytes)	Throughput*
CHLA/USC LAN (Childrens Hospital/ U Southern California)	<1 ms	9.8 MBytes/s
CHLA/USC-UCLA	4 ms	2.7 MBytes/s
CHLA/USC-Stanford U	23 ms	900 KBytes/s
CHLA/USC-UCSF	24 ms	700 KBytes/s
CHLA/USC-NLM (National Library of Medicine)	67 ms	320 KBytes/s
CHLA/USC -U Hawaii	76 ms	700 KBytes/s

*units used are MBytes/s and KBytes/s which are different from those used in Table 5.

User Friendliness

User friendliness includes both the connection procedure of the teleradiology equipment at both the examination site and the expert center, and the simplicity of using the display workstation at the expert center.

User friendliness means that the complete teleradiology operation should be as automatic as possible requiring only minimal user intervention. For the image workstation to be user friendly requires three criteria:

1. Automatic image and related data prefetch.
2. Automatic image sequencing and hanging protocol at the monitors.
3. Automatic look-up table, image rotation, and unwanted background removal from the image.

Image and related data prefetch means that all necessary historical images and related data required for comparison by the radiologist should be prefetched from the patient folder at the imaging sites and send to the expert center. When the radiologist is ready to review the case, these prefetched images and related data are already available at the expert center. Automatic image sequencing and hanging protocol at the display workstation means that all these images and related data are sequentially arranged so that at the touch of the mouse, properly arranged images and sequences are immediately displayed on the monitors. Prearranged data minimizes the time required for the searching and organizing of data by the radiologist at the expert center. This translates to an effective and efficient teleradiology operation. The third factor, automatic look-up table, rotation, and background removal, is necessary because images acquired at the distant site might not have the proper look-up table set up for optimal visual display, images might not be generated in the proper orientation, and might have some unwanted white background in the image due to radiographic collimation. All these parameters have an effect on the proper and efficient diagnosis of the images. Figure 4 shows an example of automatic splitting of a CT examination of both the chest and the abdomen into a chest and an abdomen sequence for automatic display using the concept of presentation of grouped procedures (PGP) in IHE (Integrating the Healthcare Enterprise) profile technology (10).

Image Compression

Teleradiology requires image compression because of the slow speed and high cost of using WAN. For lossless image compression, current technology can achieve between 3:1 and 2:1 compression ratios, whereas in lossy compression using cosine transform based MPEG and JPEG hardware or software, 20:1–10:1 compression ratios can be obtained with acceptable image quality. The latest advance in image compression technology is the wavelet transform (11 and JPEG 2000), which has the advantages over cosine transform for higher compression ratio and better image quality, however, hardware wavelet compression is not yet available. Some Web-based teleradiology systems use progressive wavelet image compression techniques. In this

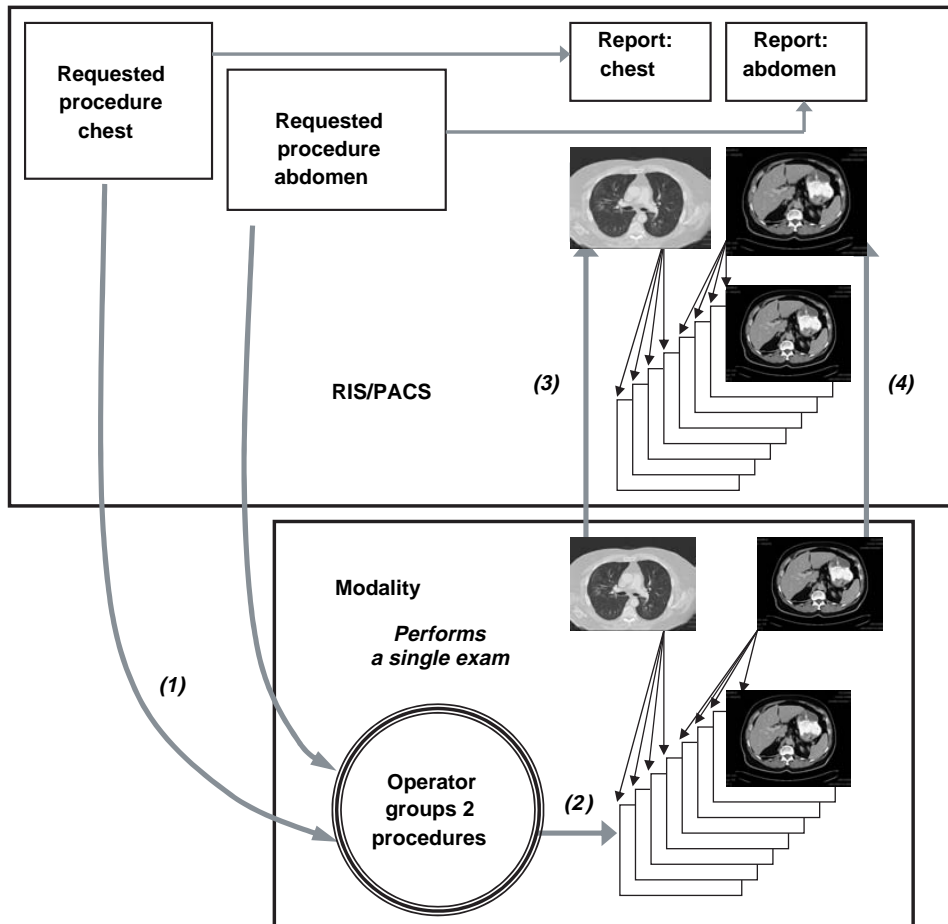


Figure 4. Integrating the Healthcare Enterprise (IHE) Presentation of Grouped Procedures (PGP) Profile. Example shows a single CT acquisition procedure is used to acquire both chest and abdominal scans. The scheduled workflow profile provides information for the PGP to split the acquired images into two subsets, one for the chest and the other for the abdomen (10). Numerals represent the workflow steps in the IHE PGP.

techniques, image reconstruction from the compressed file is in a progressive manner in that a lower resolution image is first reconstructed almost instantaneously and displayed upon request. The user would have the psychological effect that the image is transmitted through the network in real-time. Higher quality images are continuously reconstructed to replace the previous ones until the original image is reconstructed and displayed.

Image Data Privacy, Authenticity, and Integrity

Image transmission in teleradiology is mostly through public networks, for this reason, trust in image data becomes an important issue. Trust in image data is characterized in terms of privacy, authenticity, and integrity of the data. Privacy refers to denial of access to information by unauthorized individuals. Authenticity refers to validating the source of the image. Integrity refers to the assurance that the image has not been modified accidentally or deliberately during the transmission. Privacy and authenticity are the responsibility of the public network provider based on firewall and password technologies, whereas integrity is the responsibility of the end user.

Imaging integrity is mostly done based on the concept of public and private keys digital signature encrypted with mathematical algorithms during the process of image generation. In general, the public and private keys digital signature concept consists of seven steps (12):

1. Private and Public Keys: Set up a method in assigning public and private keys between the examination site and the expert center.
2. Image preprocessing: To segment the object of interest in the image from the background (e.g., the head in a CT image is the object of interest), and extract patient information from the DICOM image header at the examination site while the image is being generated.
3. Image digest: To compute the image digest (digital signature) of the object of interest in the image based on its characteristics using mathematical algorithms.
4. Data encryption: To produce a digital envelope containing the encrypted image digest and the corresponding patient information from the image header.
5. Data embedding: To embed the digital envelope into the background of the image as a further security. The background is used because the embedding would not alter the image quality of the object of interest. In cases where the image has no background, such as a chest radiograph, a more involved lossless embedding technique can be used.
6. The image with the embedded digital envelope is sent to the expert site.

7. The expert center receives Item 6, decrypts the image and the signature. It compares the two digital signatures. One comes with the image, the second is computed from the received image to validate the image integrity.

TELERADIOLOGY OPERATION MODELS

In this section, we discuss four teleradiology operation models that are common in current practice.

Off-Hour Reading

An off-hour reading model is to take care of the off-hour reviewing of the images including evenings, weekends, and holidays when most radiologists are not available at the examination sites. In this set up, image acquisition devices at different examination sites including hospitals and clinics are connected to an off-hour reading center with medium or low grade transmission speed (like the DSL) because the turn around time is not critical except for emergency cases. The connections are mostly direct digital with the DICOM standard. The reading center is equipped with network switches and various types of workstations compatible to the images generated by imaging devices at examination sites. The staffing includes technical personnel taking care of the communication networks and workstations, and radiologists who come in during the evening, weekend, and holiday shifts and perform on-line digital reading. They provide preliminary impression of the exam and transmit it to the examination site instantaneously after reading. The regular radiologists at the examination sites verify the readings and sign off the report the next day. This type of teleradiology set up is low technology but it serves its purpose of solving the shortage of radiologists during off hours.

ASP Model

Application Service Provider (ASP) model is a business venture taking care of the radiological image diagnosis for examination sites where on site radiology interpretations are not available. This model can be for supplying equipment only or for both equipment and radiologists. In the former, an ASP entity sets up a technical center housing network equipment and workstations. It also provides turnkey connectivity for the examination site where images would be transmitted to the center. The examination site can hire its own radiologists to perform reading at the center. In the latter, the center provides both technical support as well as radiologists for reading.

Web-Based Teleradiology

Web-based teleradiology is mostly used by hospital or larger clinics to distribute images to various parts of the hospitals or clinics, or outside of the hospital. A web server is designed where filtered images from PAC systems are either pushed from the PACS server to, or pulled by the Web server. Filtered images mean that the Web server has a predetermined directory to manage the image distribution based on certain criteria like what types of images to

where and to whom, and so on. The clients can view these filtered images from the client workstation through the web server. The clients can be referring physicians who just want to take a look at the images or for radiologists to make a remote diagnosis. Web-based teleradiology is very convenient and low cost to set up because most technologies are readily available, especially within the hospital intranet environment. The drawback is that since Web is a general technology, the viewing capability and conditions are not as good as that in a regular PACS workstation where the set up is geared for radiology diagnosis. In order to have full DCIOM image resolution for visualization and manipulation at the clients, modifications have to be made at the Web Server to receive full 12 bits/pixel data from the PACS server, and additional display software at the clients.

PACS and Teleradiology Combined

Teleradiology can function as a pure teleradiology operation shown in Fig. 5. In this operation, the teleradiology management center serves as the operation manager. It receives images from different imaging centers, 1, . . . , N , keeps a record, but not the images, routes images to different expert centers, 1, . . . , M according to need for reading. Reports comes back to the management center, it records the reading reports, forwards reports to the appropriate imaging centers. The management center is

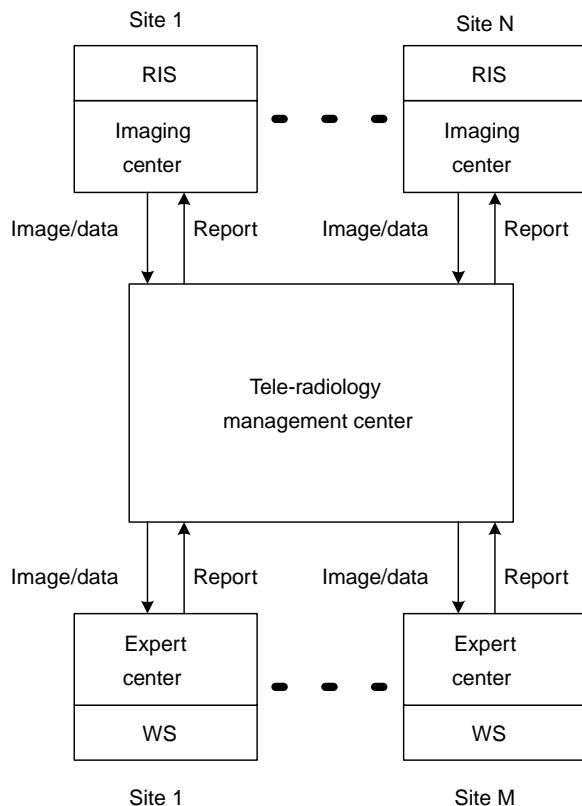


Figure 5. Pure teleradiology model. The management center monitors the operation to direct workflow between imaging centers and expert centers.

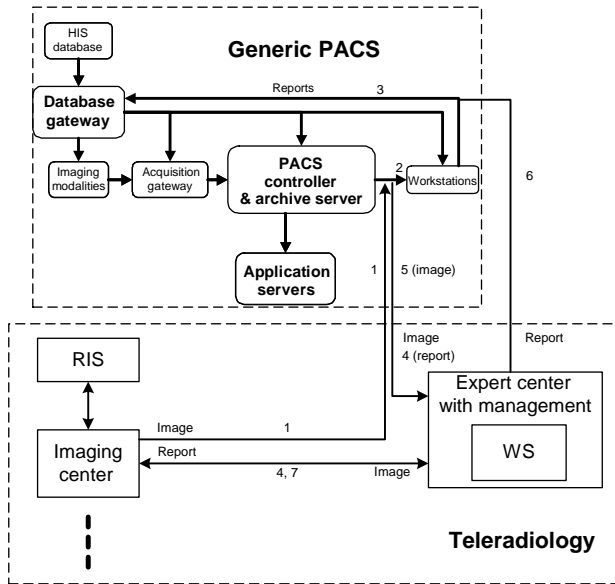


Figure 6. PACS and Radiology Combined Model: The PACS supports either the imaging centers, or PACS and Teleradiology support each other. See text for explanation of workflow steps in numerals.

also responsible for the billing and other administrative functions like image distribution and workload balancing. The networks used for connection between image centers, the management center, and expert centers can be mixed with various performances dependent on the requirements and costs.

Teleradiology can be combined together with PACS as a healthcare enterprise operation, as shown in Fig. 6. The two major components in the combined model are the PACS (see Fig. 3) shown inside the upper dotted rectangle, and the pure teleradiology (see Fig. 5) model shown in the lower dotted rectangle. The workflow of this combined model is as follows:

1. The image center can send images to the expert center for reading as in the pure teleradiology model (7).
2. Radiologists at PACS workstations can read exams from outside imaging centers as well (1).
3. After reading by in-house radiologists from its own workstations (2), reports are sent to the database gateway for its own in-house record (3), or to the expert center from where the report is also sent to the imaging center (4).
4. The PACS can also send exams directly to outside expert center for reading (5). The expert center then returns report to the PACS database gateway (6).

The combined Teleradiology and PACS model is mostly used in a large enterprise level healthcare center with satellite imaging centers, or in back-up radiology coverage between the hospital and imaging centers.

Enterprise Level PACS and Teleradiology Combined with Grid Computing

The PACS and teleradiology combined model described in the last section can be extended to the enterprise level using the grid computing infrastructure. The enterprise level PACS and teleradiology combined model is for very large-scale PAC systems and teleradiology applications. This large-scale model is becoming more and more popular in today's enterprise healthcare delivery system (13).

Grid computing is the integrated use of geographically distributed computers, networks, and storage systems to create a virtual computing system for solving large-scale, data-intensive problems in science, engineering, and commerce (14). A grid is a high performance hardware and software infrastructure providing scalable, dependable, and secure access to the distributed resources. Unlike distributed computing and cluster computing, the individual resources in grid computing maintain administrative autonomy and are allowed system heterogeneity; this aspect of grid computing guarantees scalability and vigor. Therefore, the grid's resources must adhere to agreed-upon standards to remain open and scalable. A popular standard grid computing toolkit is called Globus 3.0 (15). The grid computing provides the user with the following services: computational, data, application, and knowledge. For these reasons, grid computing infrastructure is the ideal technology for large-scale enterprise PACS and teleradiology combined model implementation.

Using the concept of grid infrastructure along with the DICOM standards and IHE workflow profiles, the PACS and teleradiology combined model shown in Fig. 6 can be extended to an enterprise level model conceptually shown in Fig. 7. Grid computing is still in its infancy for medical imaging application. The concept shown in Fig. 7 would require several years before it can materialize.

SOME IMPORTANT ISSUES IN TELERADIOLOGY

Relationship between Teleradiology Technologies and Operation

There are two sets of trade-off parameters in teleradiology. The first set relates to the operation consisting of image quality, reading turn-around time, and cost; and the second set relates to technologies used in the operation including image capture, workstation, compression, communication, and data security requirements. Table 8 shows the relationship between these two sets of parameters.

Image Data Security

In image data security, the patient confidentiality as well as image integrity are important. Since teleradiology uses a public communication method to transmit images that have no security, the question arises as to what type of protection one should provide to assure the patient's confidentiality, and the authentication of the sender. The second issue is the image integrity. After the image is created in digital form, can we assure that the image created has not been altered either intentionally or unintentionally during the transmission? To guarantee patient

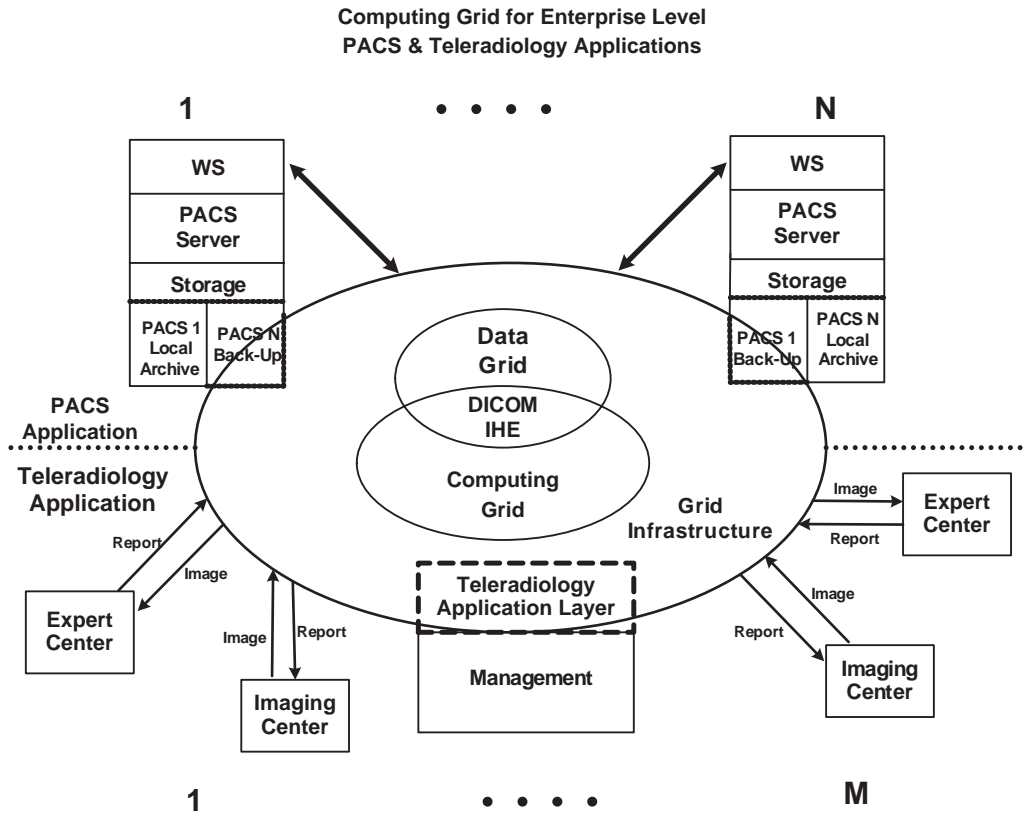


Figure 7. Enterprise level large-scale PACS and teleradiology combined model using the grid computing technology. The center ellipse of the figure is the grid computing infrastructure consisting of the data and computation grid with the DICOM standard and IHE workflow profiles. The data grid handles the image data, and the computational grid takes care of the workflow and management (see Chapter Section 19.2.2.4, Ref (4) for the concept of grid computing and the functions of the data grid.) The top row above the dotted line is the enterprise PACS, which consists of several PAC systems 1, . . . , N rectangles under numerals. The bottom row under the dotted line is the pure teleradiology model described in Fig. 5. The connection between PACS and teleradiology in the enterprise is not a straight line of data communication, instead it goes through the grid computing infrastructure for resource allocation and management, as well as image data acquisition and distribution.

confidentiality and image authenticity, methods such as hardware and software firewalls used routinely in information technology can be set up. To protect image integrity, data encryption and digital signatures can be used. These techniques have been in the domain of defense research for many years, which can be modified for teleradiology application. If high security is imposed on image data, it will increase the cost of decryption and decrease the

easy access due to many layers of passwords. The trade-off between cost and performance, confidentiality, and reliability has become a major socioeconomic issue in teleradiology. Since altering a digital image is fairly easy in today’s computer technology, developing methods to protect the integrity of image data is essential in teleradiology applications.

Medical–Legal Issues

There are four major medical–legal issues in teleradiology: privacy, licensure, credentialing, and malpractice liability. The ACR (American College of Radiology) Standard for Teleradiology adopted in 1994 defines guidelines for “qualifications of both physician and nonphysician personnel, equipment specifications, quality improvement, licensure, staff credentialing, and liability. Guidelines to these topics, although much is still uncertain, have been discussed extensively by others (16–19). It is important that these issues should be considered thoroughly before a teleradiology operation is set up.

Table 8. Relationship between Technologies and Teleradiology Operation

	Image Capture	WS	Compression	Communication	Image Integrity
Image quality	X	X	X		
Turn-around time	X	X	X	X	X
Cost	X	X	X	X	X

TRENDS IN TELEMEDICINE AND TELERADIOLOGY

The concept of telemedicine and teleradiology originated in the 1970s, however, technology was not ready for real clinical applications for teleradiology until several years ago. As teleradiology is being integrated into daily clinical service, the associated socioeconomic issues discussed also surface. The trends in teleradiology are to balance the cost with the requirements of image quality, and turn-around time for the service. Costs are affected by technology used in image capture, workstation, image compression, communication, and image security. We see that teleradiology will become a necessity in medical practices of the twenty-first century, and will be an integral component of telemedicine as the not so distant method for healthcare delivery.

Teleradiology uses the Web and Internet technologies. Issues that must be resolved immediately are how to lower the communication cost and to bundle textual with image information effectively and efficiently to assure efficient operation, and image security. For the former, Internet 2 appears to be an excellent candidate, and for the latter, ePR (electronic Patient Record) will evolve as a potential winner. [4, Ch 21].

BIBLIOGRAPHY

1. Stahl JN, Zhang J, Zeller C, Pomerantsev EV, Lou SL, Chou TM, Huang HK. Tele-conferencing with dynamic medical images. *IEEE Trans Inform Tech Biom* 2000;4(1): 88–96.
2. Zhang J, Stahl JN, Huang HK, Zhou X, Lou SL, Song KS. Real-time teleconsultation with high resolution and large volume medical images for collaborative health care. *IEEE Trans Inform Tech Biom* 2000;4(1):178–185.
3. Stahl JN, Zhang J, Chou TM, Zellner C, Pomerantsev EV, Huang HK. A new approach to tele-conferencing with intravascular ultrasound and cardiac angiography in a low-bandwidth environment. *RadioGraphics* 2000;20: 1495–1503.
4. Huang HK. *March, PACS and Imaging Informatics: Principles and Applications*. Hoboken, NJ: John Wiley & Sons; 2004.
5. Huang HK. Teleradiology technologies and some service models. Volume 20, *Computerized Medical Imaging and Graphics*. 1996. p 59–68.
6. DICOM Standard 2003, <http://www.dclunie.com/dicom-status/status.html#BaseStandard2001>
7. HL7 Version 3.0: Preview for CIOs, Managers and Programmers, http://www.neotool.com/company/press/199912_v3.htm#V3.0_preview
8. Stahl JN, Tellis W, Huang HK. Network latency and operator performance in teleradiology applications. *J Digital Imag* 2000;13(3):119–123.
9. Huang HK. 2003, Research trends in medical imaging informatics. In: Hwang NHC, Woo SLY, editors. *Frontiers in Biomedical Engineering based on the Proc WCCBME (World Congress for Chinese Biomedical Engineers Proceedings)*, Chapt. 17, New York: Kluwer Academic Publisher; 2002; p. 269–281.
10. Carr C, Moore SM. IHE: A model for driving adoption of standards. *Comp Med Imaging Graphics* 2003;Issues 2–3: 137–146.
11. Wang J, Huang HK., Three-dimensional image compression with wavelet transform. In: Bankman IN, editor-in-chief, Rangayyan RM, Woods RP, Robb RA, Huang HK, editors. *Hanbook of Medical Imaging*. Academic Press; 2000; Chapt. 52, p. 851–862.
12. Zhou X, Huang HK. Authenticity and integrity of digital mammography image. *IEEE Trans Medical Imaging* 2001; 20(8):784–791.
13. Huang HK. Enterprise PACS and image distribution. *Comp Med Imaging Graph* 2003;27(2–3):241–253.
14. Bernman F, Fox G, Hey T. *Grid Computing*. Hoboken, NJ: John Wiley & Sons; 2003.
15. GlobusToolkit 3 Core White Paper, <http://www-unix.globus.org/toolkit/documentation.html>
16. James AE, Jr, James E, III, Johnson B, James J. Legal considerations of medical of medical imaging. *Leg Med* 1993; 87–113.
17. Berger SB, Cepelewicz BB. Medical-legal issues in teleradiology. *Am J Roentgenol* 1996;166:505–510.
18. Berlin L. Malpractice issue in radiology-teleradiology. *Am J Roentgenol* 1998;170:1417–1422.
19. Kamp GH. Medical-legal issues in teleradiology: A commentary. *Am J Roentgenol* 1996;166:511–512.

See also COMPUTER-ASSISTED DETECTION AND DIAGNOSIS; RADIOLOGY INFORMATION SYSTEMS.

TEMPERATURE MONITORING

PETER FRASCO
Mayo Clinic Scottsdale
Scottsdale, Arizona

INTRODUCTION

The notion that illness and fever (elevation of body temperature above normal) are linked has been known since the time of Hippocrates and Galen. However, the concept of temperature as a quantifiable vital sign (like pulse rate and blood pressure) that could be measured and recorded is a relatively recent phenomena. Although the thermometer is as ancient as the microscope and older than the stethoscope, its use as an instrument for physical diagnosis in medicine is a relatively recent phenomenon (1).

The thermometer was invented in the fourteenth century, but its use in clinical medicine did not become commonplace until the early twentieth century (2). In 1625, Sanctorio Sanctorius described a device that was used to measure oral temperature. The practice of measuring body temperature with mercury in a glass thermometer began in the early eighteenth century with the work of the Hemann Boerhaave in the Hollan and his students in Vienna (1,2). One of his students, Anton DeHaen, noted changes in temperature with shivering or fever. He also described an increase in heart rate with increased body temperature (1). His contemporaries were unimpressed and the thermometer was largely neglected until the nineteenth century. In 1791, K.A. Wunderlich established the normal range of body temperature from 36.3 to 37.5 °C after recording nearly 1 million readings in 25,000 patients (1,2).

The most plausible explanations for the relative delay from the discovery that body temperature could be measured and monitored to the routine use to the technology relate to the complexity of the early instruments, which were nearly 12 in. (30 cm) long and required ~20 min to record a single measurement. In 1870, T.C. Allbut produced a thermometer that was portable (6 in. (15 cm) long) and reliable (2). It could record a temperature within 5 min. The Allbut thermometer was the forerunner for the mercury in glass thermometer in use presently.

In addition, the late nineteenth century was also an era of intense interest in the specific organ systems and the instruments with which to study them. During this time, the concept of disease shifted away from a holistic model toward a more organ-specific model. Illness was defined by the foci of alterations of function and structure. Overall body heat, as measured by the thermometer, which represented a general phenomena, did not fit easily into this local, organ specific disease concept.

This article will focus upon the technical aspects of temperature monitoring. This emphasis on basic science of the instrumentation of temperature monitoring will be balanced with a discussion of the clinically important topics of heat loss and heat conservation in the surgical patient.

DEFINITIONS

Heat

Heat is the form of energy that is transferred across a boundary of a system at a given temperature to another system at a lower temperature by virtue of the temperature difference between the two systems. This transfer of energy can occur through radiation, conduction, convection, and evaporation. The standard unit of heat (in the International System of Units, SI) is the calorie, which is the amount of energy needed to raise the temperature of 1 g of water by 1 °C. The British thermal unit (Btu), which is not frequently used in medicine, is defined as the amount of heat needed to raise the temperature of 1 lb of water by 1 °F (see discussion of temperature scales below). Heat is transferred from the substance at the higher temperature to the substance at the lower temperature.

Temperature

Temperature is defined as a measurement of the thermal state of a matter, which determines whether it will give heat to another substance, object or energy source, or receive heat from it (3,4). Temperature is a measure of the kinetic energy of the molecules or atoms of a substance. This energy is directly proportional to the velocity of the particles. Temperature will increase as heat energy is added and will decrease as heat energy is lost.

Body Temperature

Body temperature is best defined as the measure of the total kinetic energy within the body. This temperature represents the net thermal effect of total body heat produc-

tion and heat loss. Body temperature will vary in different parts of the body depending on perfusion, exposure, metabolic activity, local heat gain, or local heat loss. The physiologically important temperature is the “core temperature”, which represents the temperature of the body’s vital organs (brain, heart, lungs, gut). Any true change in body temperature (ΔT) represents an imbalance in the dynamic between production and loss and can be defined by the following formula (5):

$$\Delta T = \frac{\text{heat production} - \text{heat loss}}{\text{body mass} \times \text{specific heat}}$$

SCALES

Temperature scales are constructed by defining two points based upon a predictable and preferably linear change in the physical property of a given substance at a constant pressure, assigning temperature values to each and then defining the unit of increment between the fixed points. The relationship between temperature and the physical property can be defined as follows:

$$t(x) = ax + b$$

where t is the temperature of the substance. This temperature changes as the property x of the substance changes. The constants a and b depend on the substance used (e.g., mercury, ethanol, copper). The constants are also determined by specifying two points on the scale.

Kelvin

A reading of 1 K [named after William Thompson, a.k.a. Lord Kelvin (1824–1907)] is defined as $1/273.16(3.6609 \times 10^{-3})$ of the thermodynamic temperature of the triple point of pure water (3). The triple point of water is the point at which the solid, liquid, and gaseous phases of water are in equilibrium. Absolute zero, the absence of all heat, when the pressure of the ideal gas is zero, is a temperature of 0 K. The triple point of water in the Kelvin scale is 273 K. Kelvin is the SI unit of temperature.

Centigrade

The centigrade scale was first described by Carolus Linnaeus. The freezing point of water was set at 0 and the boiling point at 100.

Celsius

The celsius temperature scale was originally defined by Anders Celsius. He set the boiling point of water at 0 and the freezing point at 100. As such, the Celsius scale was the reverse of the Centigrade scale. In 1948, the centigrade scale was replaced with a newly defined Celsius scale that was based upon setting the triple point of water at 0.01 °C and the boiling point of water at 99.975 °C. The single degree increments in the Celsius scale are equal in magnitude to those in the Kelvin scale. To convert from celsius to kelvin the following formula can be used

$$K = ^\circ C + 273$$

Table 1. Characteristics of Common Thermometers

Type	Liquid Expansion	Resistance Coil	Thermistor	Thermocouple	Liquid Crystal	Infrared
Accuracy	±0.2 °C	±0.1 °C	±0.1 °C	±0.1 °C	±0.4 °C	±0.3 °C
Sites	Oral, Rectal	Skin	All sites	All sites	Skin	Oral Rectal Otic
Cost	Inexpensive	Expensive	Inexpensive	Moderate	Inexpensive	Very Expensive
Design	Expansion	Electrical	Electrical	Electrical	Chemical	Radiation

Fahrenheit

The fahrenheit [named after Gabriel Fahrenheit (1686–1736)] scale (°F), like the Kelvin scale, is also based upon the triple point and boiling point of water. The scale was originally calibrated with a mercury thermometer using a mixture of salt, ice and water as the zero point. The second point was obtained when the salt was eliminated from the ice and water. This was set at 30 °F. The boiling point of water is 212 °F on this scale. The freezing point of water was adjusted to 32 °F to allow the interval between the triple point and boiling point to be a more rational number (180).

To convert from Fahrenheit to Celsius the following formula can be used

$$^{\circ}\text{F} = (1.8 \times ^{\circ}\text{C}) + 32$$

TYPES OF THERMOMETERS

A thermometer measures temperature of a system in a quantitative way (Table 1).

NONELECTRICAL METHODS

Changes in Physical Dimensions (See Table 1)

Liquid Expansion Thermometers. The nonelectrical methods of thermometry are loosely based upon the second perfect gas law, Charles' or Guy Lussac's law, which states that at constant pressure the volume of a given mass of gas varies directly with the absolute temperature (4). As heat is added to a substance and temperature increases, the volume of liquids and gases increase. Mercury and ethanol are the most commonly used materials for the expansion based or liquid in glass thermometers.

Ethanol is an alternative to mercury in glass thermometers. Although cheaper than mercury, ethanol thermometers may be unsuitable for high temperatures since ethanol boils at 78.5 °C. Ethanol thermometers have a range from –75 to 120 °C. This is unimportant in the clinical setting of measuring and monitoring body temperature but may be important in other aspects of medicine. In addition, the scale of the ethanol-based thermometer may be less linear than that of the mercury thermometer (3,5).

The design of the liquid expansion thermometer has changed little in last century. The most sensitive location of the liquid in glass thermometer is the bulb, where the largest volume of liquid exists. However, the entire thermometer is temperature sensitive. These thermometers are manufactured with a constriction at the lower portion of the mercury column near the bulb. This prevents the

liquid from retracting into the bulb and allows the maximum temperature measured to be displayed until the thermometer is shaken.

Liquid expansion thermometers are impractical for continuous use due to the inherent rigidity of the device, the risk of breakage and the typically slow response (1–3 min) compared to the techniques listed below.

Bimetallic. The sensing element is comprised of two dissimilar metals bonded together in a coil, spiral, or disk. One end of the coil is attached to a lever, the other end is fixed to a point within the device. As temperature changes, the metals expand (with heat) or contract (with cold) by different amounts and the coil tightens or loosens, respectively, and moves a lever over a scale. These thermometers are not very accurate, but are relatively stable over time, require little maintenance, and are inexpensive. Bimetallic thermometer sensitivity to small changes in temperature can be increased by using a long strip wound into a tight coil. The most common clinical application for the bimetallic spring thermometer is use with thermostat devices, temperature alarm devices and operating room temperature monitoring.

The radius of curvature for a bimetallic thermometer is inversely proportional to the difference between the bonding temperature for the strip and the temperature being measured (3,5).

$$R = \frac{t\{3(1+m)^2 + (1+mn)[m^2 + (1/mn)]\}}{6(\alpha_2 - \alpha_1)(T - T_0)(1+m)^2}$$

where t = total thickness of the bimetallic strip; m = ratio of thicknesses (low/high expansion materials); n = ratio of Young moduli of elasticity (low/high expansion materials); α_1 = lower coefficient of thermal expansion (1/°C); α_2 = higher coefficient of thermal expansion (1/°C); T = temperature, °C; T_0 = initial bonding temperature, °C.

CHANGES IN ELECTRICAL PROPERTIES (SEE TABLE 1)

Resistance Thermometer

The conductivity of any metal depends on the movement of electrons through its crystal lattice. Resistance to this movement of electrons varies with temperature. Resistance temperature detectors (RTDs) utilize metallic conductors with positive coefficients of resistance or positive temperature coefficients (PTC). As temperature increases, resistance increases almost linearly. The metals that have nearly linear PTC include platinum, tungsten, nickel and nickel alloys. Each metal has a specific resistivity, ρ , which

varies directly with temperature (3,5–7).

$$\rho_T = \rho_0 [1 + \alpha(T - T_0)]$$

$$\text{Resistance}(\Omega) = \frac{(\rho)L}{A}$$

where L = metal wire length and A = cross-sectional area.

The RTD response relationship is nearly linear and is defined by the following equation (3–7):

$$R = R_0[1 - a(T - T_0) + b(T - T_0)^2]$$

Although the response is nearly linear throughout a wide range of temperature, it is linearity over the smaller temperature interval that is important in clinical medicine. In this scenario, resistance varies according to the following formula:

$$R_t = R_0 (1 + \alpha^*T)$$

Where a and b are calibration constants for resistor material and purity, T is measured temperature, and R_0 reference resistance measured at T_0 .

A simple RTD system is comprised of a metal resistor (e.g., platinum wire fashioned into a coil), a source of electrical potential and an ammeter calibrated to indicate temperature. Platinum is the preferred component due to its resistance to corrosion and large positive coefficient of resistance. Adding a Wheatstone bridge with an array of resistors increases the sensitivity of an RTD system.

The RTDs are the most accurate, most stable of the electrical methods for temperature measurement and are nearly linear over a relatively wide range of temperatures. They are, however, slow and expensive compared to thermocouples and thermistors (3,5). In addition, unlike thermocouples, an external current source is required to produce a voltage drop across the sensor. This source of current is a source of self-heating of the RTD if the current is not limited.

Thermistor

A thermistor (Fig. 1) is similar in principle to a resistance thermometer in that the ability to measure temperature depends on the changes in resistance of various metals in

response to temperature. There are some important differences. Unlike RTDs, which are made conductors with a positive coefficient to resistance, thermistors are made from semiconductor materials that have a large negative coefficient of resistance. Therefore, as temperature increases, the resistance within the thermistor decreases.

$$\text{Resistance}(\Omega) = A^{(B/\text{absolute temperature})}$$

The relationship between resistance and temperature is nonlinear. In addition, thermistors operate within a relatively high resistance range (1 kΩ–100 kΩ) compared to RTDs.

A digital readout thermistor-based thermometer is illustrated in Fig. 1.

Thermistors are made from heavy metal oxides (cobalt, iron, nickel, manganese, zinc). The metal is shaped into a bead that is sealed into a small measuring tip to which electrodes are attached. The tip can be sealed into a glass tube, cardiac catheter or stainless steel probe. Thermistors are accurate to $\pm 0.5^\circ\text{C}$ over a range of -80 to 150°C (3). Other advantages of thermistors include relatively small size, low production cost, reproducibility, very high sensitivity, and resolution and relative insensitivity to shock and vibration. Compared to RTDs, however, thermistors are less stable and are more susceptible to internal heating or self-heating. In addition, recalibration may be necessary because the resistance of the metal oxide increases over extended periods of time for unclear reasons. Calibration is defined by the following equation (3,8):

$$\text{Resistance} = R_0 e^{\beta(1/T - 1/T_0)}$$

where R_0 = reference resistance measured at T_0 ; T = measured temperature; β = material constant.

Thermocouple

When heat is applied to one end of a metallic conductor, electrons at this “hot” junction acquire increased thermal energy relative to those electrons at the unheated or “cold” junction. The electrons diffuse from the hot junction to the cold junction, and in doing so lose this thermal energy. Heat is conducted along the conductor and an

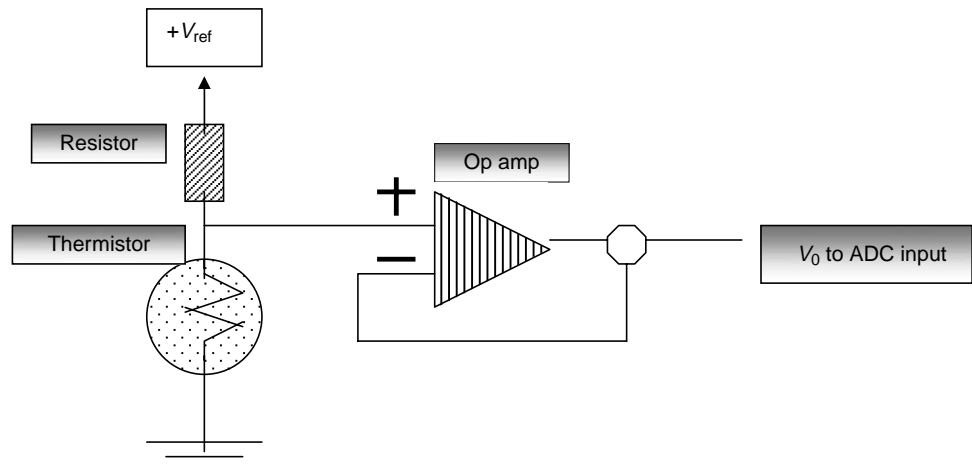


Figure 1. Digital thermistor thermometer ADC.

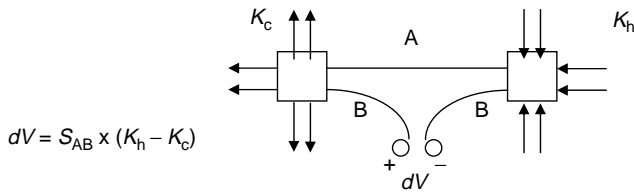


Figure 2. Seebeck principle.

accumulation of electrons at the cold end results in the production of an electric field between the ends of the metal (4). A small voltage is produced at the junction point (in an open circuit) of two dissimilar metals when heat is applied. The electromotive force or voltage across the terminals of an open circuit comprised of two dissimilar metals is proportional to the temperature difference of the two junction temperatures. Platinum, platinum–rhodium, nickel–chromium, and nickel–aluminum can be used as thermocouples. The thermocouple is enclosed in a metal or ceramic shield that protects the device from corrosion and other environmental factors. This voltage does not depend on the temperature along the metals between the junctions.

The thermocouple is based upon the Seebeck principle (Figs. 2 and 3). In 1821, T.J. Seebeck discovered that current is produced in a closed-circuit comprised of two dissimilar metallic conductors when heat is applied to one of the two junctions or if a temperature difference exists between the two junctions. Not every combination pair of metals is acceptable for thermocouple usage, since the direction and magnitude of the current produced is a function of the thermal properties of the metals that comprise the circuit and the temperature difference between the junctions.

In Fig. 2, dV is the voltage difference, K_c is the temperature of the cold junction, K_h is the temperature of the hot junction, S_{AB} is the factor of proportionality.

If the potential difference is to be measured, a second junction is required to produce a complete electrical circuit. This second junction will have its own thermoelectric electromotive force. The electromotive force generated is quantitatively and linearly related to the temperature difference between the two junctions. In order for the thermocouple to be used as a thermometer, this second

junction (reference junction, see Fig. 3) must be either maintained at a constant temperature or possess some form of mechanical compensation. The reference junction can be maintained at a constant temperature by immersion in an ice bath or kept at a precisely controlled, constant temperature. Mechanical compensation can be obtained by using a bimetallic spring attached to a coil suspension system. This suspension system functions as the galvanometer. The bimetallic spring is selected according to the range of the instrument. In solid-state instruments, an electrical zero can be used or a thermistor or RTD is used to monitor the reference or cold junction. The signal from the thermistor or the RTD is used to compensate for the cold junction temperature. Figure 3 illustrates a schematic of a thermistor and the relationship between temperature and voltage at the measuring junction.

Thermocouples are versatile, inexpensive, small, and durable. They are accurate within a wide temperature range and can be manufactured into a wide variety of physical forms. The low thermal capacity of thermocouples allows for a rapid response time for measurement. An added advantage over RTDs and thermistors is that thermocouples are self-powered. However, compared to the other contact sensors [thermistors, RTDs, and integrated circuit thermometers (ICs)], thermocouples are less sensitive and stable. Other disadvantages include necessity for compensation of the reference junction and, compared to RTDs, a relatively nonlinear voltage to temperature relationship.

Changes in Chemical Phase

Quartz Crystal Thermometry. Liquid crystals (LCs) have the optical properties of a single crystalline solid, but possess the mechanical properties of a liquid. Temperature changes can affect the color of certain liquid crystals. Crystals at specific temperatures, when exposed to light, will scatter some of the light that in turn produces iridescence. This property allows LCs to be used for temperature measurement. Liquid crystals are broadly categorized as either lyotropic or thermotropic. Lyotropic crystals are used in the production of soaps and detergents and react to the type of solvent with which they are mixed (9).

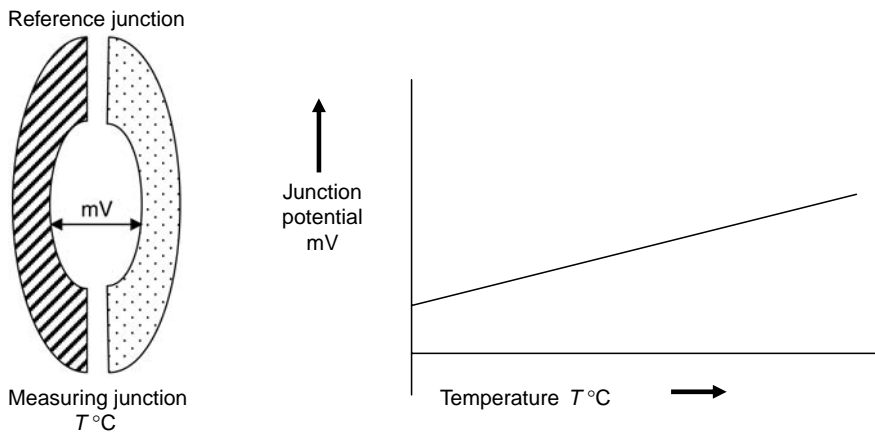


Figure 3. Thermocouple.

Thermotropic liquid crystals react to changes in temperature or pressure (9). Thermotropic crystals are further categorized as either isotropic or nematic (9). Isotropic crystals are random in their arrangement, while nematic crystals have a definite pattern or order. Nematic liquid crystals can be arranged in layers (smectic) or in spirals (cholesteric) (9). Temperature indicators can be constructed by embedding microencapsulated cholesteric liquid crystals into adhesive strips (3,8,9). These crystal devices can be applied to any accessible or visible location. These devices are more useful as trend devices during surgical procedures since they are only accurate to 0.5°C (3,8). Other disadvantages of LC thermometry include a lack of interfaces to other hemodynamic or respiratory monitors and inherent subjectivity, as colors may be interpreted differently by individual observers.

CHANGES IN EMITTED THERMAL RADIATION (SEE TABLE 1)

Infrared Thermometers

The electromagnetic spectrum is divided into a number of wavelength regions called bands. Light at frequencies less than red, which is at the low frequency edge of the visible portion of the spectrum, are called infrared (IR) and are not visible to the human eye. The IR band of the spectrum covers wavelengths between 0.7 and 1000 μm (4). All bodies with a temperature >0 K radiate energy in the IR band. This heat energy induces electron vibrations that cause electromagnetic emission. This energy travels like light (as an electromagnetic wave or photon) in all directions. The frequency of this emitted radiation is temperature dependent. The amplitude of the emission is dependent on the emissivity of the substance. Emissivity is the ratio of the energy radiated by an object at a given temperature to the energy radiated by a blackbody (perfect radiator) at the same temperature. Emissivity depends on the surface finish, color, aging, and oxidation state of the substance in question. For example, a highly polished metallic object would have a high reflectivity and a low emissivity.

Radiation striking a surface is reflected, absorbed and/or transmitted.

$$\rho(\text{reflectivity}) + \varepsilon(\text{emissivity}) + \tau(\text{transmissivity}) = 1$$

$$\varepsilon(\text{emissivity}) = \alpha(\text{absorbivity})$$

Infrared thermometers measure the amount of IR energy emitted from the object of interest. Contact with the surface or object of interest is not required. Infrared thermometers are composed of a lens (collection of energy emitted from an object), a sensor (thermal, photoelectric or photon detector), and a signal converter (converts thermal energy into an electrical signal). Infrared thermometers have very rapid response times (milliseconds), do not require contact with an object or surface (avoid contamination), are simple to use, and require little if any maintenance.

$$T_b = \frac{[\chi(N_T - N_{T0}) + T_0^4]}{4}$$

A basic IR thermometer is comprised of four parts: (1) a waveguide that collects and focuses the energy emitted by the target, (2) a pyroelectric sensor that converts the energy to an electrical signal, (3) a microprocessor that adjusts emissivity allowing thermometer calibration to match the emitting characteristics of the object being measured, and (4) a sensor temperature compensation circuit that ensuring that temperature variations within the thermometer are not transferred to the final output. which are fed through the amplifier, multiplexer (MUX), and analogue-to-digital converter (ADC) to the microprocessor for processing and display. The microprocessor that handles emissivity adjustment also performs temperature compensation and calculates the patient temperature.

CLINICAL APPLICATIONS

With the exception of monitoring temperature during anesthesia and surgery or during care in the intensive care unit, temperature is measured and recorded intermittently. During surgery, according to the guidelines established by the American Society of Anesthesiologists (www.asahq.org), the capability to monitor temperature must be readily available. In clinical situations where changes in body temperature are anticipated (long abdominal (10) or thoracic surgical procedures, pediatric) or intended (cardiopulmonary bypass (8,11), neurosurgical procedures), temperature must be monitored. In these situations, temperature is measured continuously with either internal (invasive) or external devices.

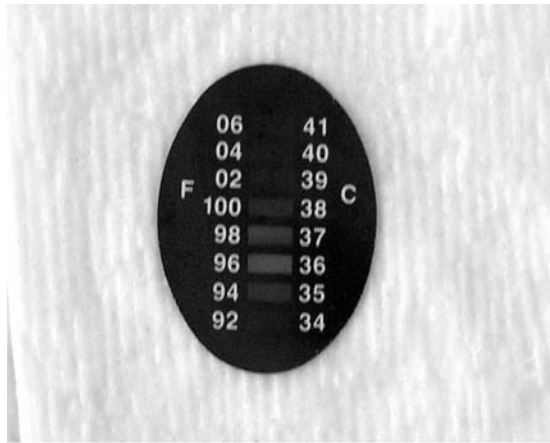
Sites

Accepted measurement sites of core temperature fall into four categories, with distinct advantages, disadvantages, and practical limitations: peripheral (skin); brain (tympanic, nasopharyngeal); visceral (bladder, esophageal, rectal); blood [Pulmonary artery catheter (PAC), esophageal].

There are distinct advantages and disadvantages for each site. These can be broadly categorized into ease of access, risk-benefit analysis, accuracy in measuring core temperature, and precision (5,8,12). For example, the pulmonary artery temperature clearly and accurately reflects core and cardiac temperature (except during the early transition periods during hypothermic cardiopulmonary bypass with extracorporeal circulation) (11), but requires access to the central circulation.

Skin

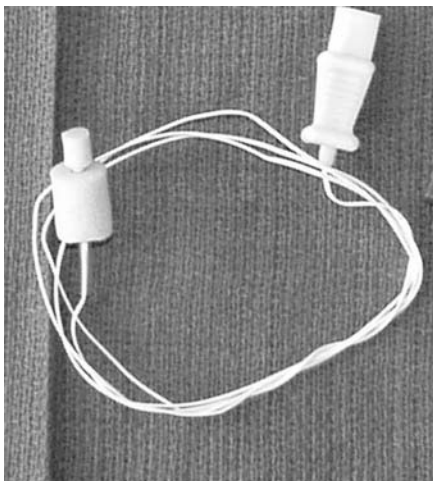
The skin is easily accessed for temperature measurement, but skin temperature can correlate poorly with core temperature. Disposable LCs or reusable metallic disks attached to thermistors or thermocouples can be used (see Fig. 4a). The LC devices are commonly applied to the forehead where fat distribution is minimal and where regional blood flow is adequate. The reusable probes can be placed in a number of locations including the back, chest or abdominal wall, axilla, or distal extremities. Measurement of skin temperature is also affected by moisture



(a)



(b)



(c)

Figure 4. (a) Liquid-crystal disk and contact-type (skin) thermistor. (b) Infrared otic (tympanic membrane) thermometer. (c) Contact-type thermistor for tympanic membrane temperature.

(sweat) and pressure applied to the probe. An increase in the amount of pressure on the probe will increase temperature measured. The unpredictable relationship between skin and core temperature during surgery and anesthesia makes the skin probe unreliable as an accurate marker for core temperature. Skin probes can be useful as a convenient and essentially risk-free trend monitor (as long as environmental temperature and skin perfusion are constant).

Oral Cavity

The mercury or ethanol in glass thermometers can be used in the oral cavity. As noted previously, the response time is slow relative to other techniques. Thermistors or thermocouples can be manufactured into probes that can be inserted into the oral cavity for rapid, intermittent temperature measurement. In addition, probes can be produced to allow continuous monitoring in the oral cavity. The oral cavity temperature can be up to 1.1 °C lower than simultaneous measurements from the pulmonary artery catheter (5,12).

Tympanic Membrane

The tympanic membrane is a useful and accessible site for IR thermometry. It is a relatively flat, uniformly textured structure that is (1) supplied by branches of the external carotid artery and (2) close to the internal carotid artery, which is the main supply of blood to the brain. As such, it is an acceptable location for core temperature measurement that accurately reflects hypothalamic temperature. The IR probe (see Fig. 4b) (with the aforementioned components) is configured into an otoscope-like device. This probe is covered with a disposable cone-shaped probe cover. The temperature is then displayed on a liquid crystal or light emitting diode display. They are, however, impractical for continuous measurements in medicine due to size, shape, and design of the current devices.

Contact type probes can be produced using thermocouples or thermistors. These probes can be used for continuous measurement of temperature during surgery and anesthesia. The probes are usually manufactured with a foam insulator so the tip of the probe is the only site of contact and measurement. Alternatively, a cotton or felt ball may be inserted to insulate and stabilize a wire probe (see Fig. 4c). There is little lag in the tympanic temperature during core cooling and warming studies, in contrast to most of the other available sites.

Nasopharynx

The nasopharyngeal probes are similar in construction (thermocouple or thermistor) to the tympanic membrane probes. The nasopharyngeal temperature is an adequate marker for brain temperature as long as the probe is placed posterior to the soft palate. The probe is quite uncomfortable in the awake patient. In addition, in the intubated patient, the probe is subject to airflow currents if leakage around the endotracheal tube occurs. These currents can adversely affect the accuracy of the probe in the nasopharynx.

Esophagus

Midesophageal temperature can be measured using a thermistor or thermocouple attached to a flexible probe. The probe can also be integrated with a stethoscope that can be used to monitor heart and breath sounds during anesthesia and surgery.

Position within the esophagus is important in predicting the accuracy of esophageal temperature as a marker of core temperature. Temperature can vary from 1 to 6 °C, depending on the position of the probe in the esophagus. For example, if the probe is inserted in a more distal location (eg, stomach), the temperature measured may be higher than the true core temperature due to warming caused by the normal metabolism in the liver, which is adjacent to the stomach. If the probe is positioned in more proximal position (adjacent to the trachea), accuracy relative to core temperature is affected by cooling secondary to ventilation with room temperature gases via the endotracheal tube. In addition, the probe may be cooled by infusion of cold fluids for irrigation into a chest incision for thoracic or cardiac procedures. The temperature will be affected by the cooling and warming phase of cardiopulmonary bypass with extracorporeal circulation. Optimal location of an esophageal temperature probe in an adult is between 35 and 45 cm from the nostrils.

Urinary Bladder

The bladder temperature correlates well with other measures of core temperature. Bladder temperature is recorded with specially modified urinary catheters. These catheters are commonly used to monitor urine output during surgical procedures or during treatment in the intensive care unit. A thermistor or thermocouple is attached to the patient end of the catheter.

Changes in bladder temperature during hypothermic cardiopulmonary bypass with extracorporeal circulation will occur later than changes in nasopharyngeal, tympanic membrane, and pulmonary artery temperature.

Rectum

The rectal mucosa was previously used as a site for core temperature measurement. It is now clear that the rectum is a relatively peripheral site. As such, it does not accurately reflect core temperature. In fact, at times, rectal temperature may exceed core temperature due to the heat produced by metabolism of the fecal bacterial flora. At other times, the presence of feces in the rectum may insulate the probe from contact with the rectal mucosa. In addition, the rectal veins receive blood from the lower extremities. This blood can have a cooling effect on the rectal temperature.

Rectal probes are seldom used for continuous temperature monitoring due the inaccuracies listed above. In addition, awake patients may find them uncomfortable.

Intermittent rectal temperature measurements are often made in the pediatric population using mercury or ethanol in glass thermometers. In addition, a digital thermistor or thermocouple probe may be used in this location.

Intravascular

The most accurate location for measuring core temperature is the central circulation. The inferior vena cava and the superior vena cava drain into the right atrium, and this mixed-venous blood enters the right ventricle and exits via the pulmonary artery to the pulmonary circulation. The pulmonary artery temperature can be measured with a thermistor that is positioned at the tip of a pulmonary artery (or Swan-Ganz) catheter. This thermistor is actually intended for use in calculation of the cardiac output using the thermodilution technique, display of the temperature is a byproduct of this design.

The pulmonary artery catheter is inserted via an introducer or sheath that is placed into one of the large veins of the body (subclavian, internal jugular, or femoral) typically by percutaneous cannulation. While the temperature of mixed-venous blood accurately reflects true core temperature, there are variables that contribute to the accuracy of any given measure. The pulmonary artery temperature is affected by local heating or cooling during the rewarming phase or the hypothermic phase of cardiopulmonary bypass, the instillation of warm or cold irrigation fluids into the chest, and application of ice or slush to the pericardium. In the absence of these factors, the pulmonary artery temperature is an accurate monitor for core temperature.

Insertion of the pulmonary artery catheter involves significant risks including infection, hematoma, pneumothorax, and arrhythmia and is not appropriate for use solely as a temperature monitor. However, the cases in which pulmonary artery catheters tend to be used, including cardiac surgery, major vascular, and major abdominal surgery, tend also to be associated with the potential for profound changes in core temperature.

Temperature Regulation

The body's core temperature may change without any true change in the total heat content due to redistribution of heat to or from the periphery (7,13–15). This occurs most commonly during surgical procedures due to peripheral dilatation secondary to anesthesia.

The temperature control center is located in the anterior hypothalamus. The hypothalamic regulatory center receives afferent input from peripheral skin receptors and initiates appropriate responses, such as reduced heat loss through skin vasoconstriction, increased heat production through raised muscle tone, or shivering (7,13–15). Impairment of thermoregulatory control can occur due to a number of factors including (7,13–15): anesthetic drugs; hypoxemia; extremes of age; shock (Blood loss or hypovolemia); hypothyroidism; hypoglycemia; malnutrition; extreme exertion; central nervous system (CNS) dysfunction.

Heat production varies by age and gender. At rest, the average male produces 1 kcal·kg⁻¹ of heat per hour and the average female 0.93 kcal·kg⁻¹·h⁻¹ (16,17). Heat production occurs primarily by metabolic activity in the liver and skeletal muscles. Heat production in skeletal muscle increases with voluntary activity (movement and exercise) and shivering (see below). With exercise, heat production can increase to 3–9 kcal·kg⁻¹·h⁻¹ (16,17). This heat is transferred from the muscles to the blood supply of the muscular bed. This blood then

enters the liver raising core temperature. The specific heat of a human is $0.83 \text{ kcal}\cdot\text{kg}^{-1}\cdot\text{C}^{-1}$, with a 70 kg male having to gain 58.1 kcal to raise core temperature by 1°C (16,17). With basal heat production of only $1 \text{ kcal}\cdot\text{kg}^{-1}\cdot\text{h}^{-1}$, endogenous heat production, if normal, would still raise body temperature $1.2^\circ\text{C}\cdot\text{h}^{-1}$; assuming sufficient insulation to prevent any other heat loss from any other mechanisms.

The loss of body heat during anesthesia and surgery is the most significant contributing factor to the development of hypothermia. Heat loss occurs as a consequence of loss through multiple mechanisms, including (10,13,14): evaporative (15–30%); conductive (20–30%); convective (15–30%); radiant (30–50%).

Evaporation is the conversion of a liquid into a vapor at a temperature below its boiling point. Evaporation is always accompanied by a reduction in temperature. Molecules of water at the skin surface and mucosal surfaces in the airways and viscera with enough heat energy to overcome the cohesion of neighboring molecules will vaporize. The average energy of the remaining water molecules will decrease reducing the surface temperature. The intraoperative evaporative loss occurs via the skin, the surgical incision, and, most importantly, the lungs, secondary to controlled ventilation of cold, dry gases. Evaporative losses are dependent on the surface area of the exposed surface, minute ventilation, relative humidity, and airflow velocity. These losses can be reduced by increasing the humidity of respiratory gases. The anesthetic drugs, which produce peripheral and cerebral vasodilatation, certainly exaggerate the normal evaporative skin loss of heat and prevent peripheral vasoconstriction to conserve heat. However, sweating, which increases the amount of moisture available, increases evaporative losses. It is estimated that a general anesthetic reduces the shivering threshold to $<34.5^\circ\text{C}$, and muscle relaxants will completely abolish this protective response (13,14).

Conduction is the transfer of heat energy by random atomic or molecular motion between two objects. The transfer always occurs down the temperature gradient from the warmer surface to the colder substance. Conductive loss occurs from (1) placing the patient on the room temperature operating room table and (2) patient contact with the room temperature surgical instruments (18). Conduction is less important than radiation, evaporation, and convection clinically since the layer of air that surrounds a body insulates against heat loss unless air movement is present (see Convection below).

The administration of intravenous fluids, cold blood bank products, as well as surgical wound irrigation provides additional significant conductive heat losses (19). The thermal stress of infusing 1 L of unwarmed crystalloid is $\sim 17 \text{ kcal}$ and the stress from a liter of unwarmed bank blood is $\sim 30 \text{ kcal}$.

Calculation of the thermal stress from intravenous fluids or transfusions can be made with the formula:

$$\text{IV "Lost" kcal} = (T_c - T_f) \times V$$

Where T_c = body temperature, T_f = fluid temperature, and V = volume of fluid infused in liters. Hence, the administration of 3 L of room temperature (20°C) crystalloid to a patient represents a 61 kcal challenge whose core temperature is 37°C ($[37-20^\circ\text{C}] \times 3 \text{ L}$).

Convection is the transfer of heat secondary to air currents. Air adjacent to the skin is warmed by conduction. This warmed is less dense and rises. As air flows over the body, the current carries heat away from the body. Convective heat flux is defined by the following equation (4,5,20):

$$Q_{cv} = \gamma(T_s - T_g)$$

where γ is the proportionality constant ($\text{J}\cdot\text{g}^{-2}\cdot\text{K}^{-1}$), T_s is the surface temperature, and T_g is the air temperature.

Forced convective heat loss is caused by gas flow caused by external means (fan or pump). Free convective loss occurs due to the gas flow that occurs secondary to temperature differences (4,5,18,20).

Clinical convective losses occur secondary to the requirement of maintaining the air conditioner and/or the air handling system at the colder settings due to multiple layers of sterile clothing worn by the surgical staff. In addition, the rates of operating room air turnover may be significantly higher than in other ambient settings. This increase in air turnover serves to reduce the likelihood of infection, but increases convective losses. Convective losses can be reduced by trapping the layer of air between the skin and external environment with a barrier, such as a thermal blanket or forced air device (21).

Radiation heat loss occurs due to the transfer of heat in the form of electromagnetic energy. The magnitude of heat transfer is dependent on the surface area of the emitting object (4,13,18). It is not dependent on the presence of a material medium therefore no direct contact is required.

SUMMARY

Iatrogenic hypothermia predisposes the patient to profound physiological consequences, including delayed recovery from anesthesia, increased oxygen consumption, increased vascular resistance, cardiac instability and potential ischemia or infarction, coagulopathy, diminished patient satisfaction, and increased recovery room costs. In considering the need to provide the best patient outcomes, hypothermia induced complications can and should be prevented. Prevention of hypothermia requires monitoring of body temperature. The devices that can be used vary in accuracy, convenience and degree of invasiveness. Appropriate choices will influence patient outcome.

BIBLIOGRAPHY

1. Lyons A, Keiner M. The seventeenth century: anatomical and physiological advances: The thermometer. In: Lyons A, Petrucelli R, editors. *Medicine: An Illustrated History*. New York: Abradale Press; 1987. p 437–439.
2. Pearce J. A brief history of the clinical thermometer. *QJM* 2002;95(4):251–252.
3. Liptak B. *Temperature Measurement*. 3rd ed. Radnor (PA): Chilton Book Company; 1993. p 134.
4. Halliday D, Resnick R, Walker J. *Fundamentals of Physics*. 7th ed. Hoboken (NJ): Wiley; 2005.
5. Parbrook G, Davis P, Parbrook E, editors. *Basic Physics and Measurement in Anaesthesia*. 3rd ed. Oxford: Butterworth-Heinemann Ltd.; 1990. p 344.

6. Chang H. *Inventing temperature: measurement and scientific progress*. Oxford: Oxford University Press; 2004. p 286.
7. Szocik J, Barker S, Tremper K. *Fundamental principles of monitoring and instrumentation*. In: Miller R, et al., editors. *Anesthesia*. Philadelphia: Churchill Livingstone; 2000. p 1053–1077.
8. Cork R. *Temperature monitoring*. In: Blitt C, Hines R, editors. *Monitoring in Anesthesia and Critical Care Medicine*. New York: Churchill Livingstone; 1995. p 543–556.
9. Collins P. *Liquid Crystals: Nature's Delicate Phase of Matter*. 2nd ed. Princeton: Princeton University Press; 2002. p 204.
10. Joachimsson P, Hedstrand U, Tabow F, Hansson B. Prevention of intraoperative hypothermia during abdominal surgery. *Acta Anaesthesiol Scand* 1987;31(1):330–337.
11. Phillips P, Skov P. Rewarming and cardiac surgery: a review. *Heart Lung* 1988;17(5):511–520.
12. Cork R, Vaughn R, Humphery L. Precision and accuracy of intraoperative temperature monitoring. *Anesthesia Analg* 1983;62:211–217.
13. Sessler D. Mild perioperative hypothermia. *N Engl J Med* 1997;336:1630–1637.
14. Sessler D, Rubinstein E, Moayeri A. Physiologic responses to mild perianesthetic hypothermia in humans. *Anesthesiology* 1991;75(4):594–610.
15. Sessler D, Schroeder M. Heat loss in humans covered with cotton hospital blankets. *Anesth Analg* 1993;77:73–77.
16. Guyton A, Hall J. *Textbook of Medical Physiology*. 10th ed. Philadelphia: Saunders; 2000. p 1064.
17. Schafer J. *Body temperature regulation*. In: Johnson L, editor. *Essential Medical Physiology*. San Diego: Elsevier; 2003. 921–932.
18. Fallacaro M, Fallacaro N, Rachel T. Inadvertent hypothermia. Etiology, effects and prevention. *AORN Journal* 1986;44(1): 54-7–60-1.
19. Evans J, et al. Cardiovascular performance and core temperature during transurethral prostatectomy. *J Urol* 1994;152: 2025–2029.
20. Bejan A. *Convection Heat Transfer*. 3rd ed. New York: Wiley; 2004. p 728.
21. Augustine S. Hypothermia in the post anesthesia care unit. *J Post Anesthesia Nursing* 1990;5:254–263.

See also BIOHEAT TRANSFER; INTEGRATED CIRCUIT TEMPERATURE SENSOR; MONITORING IN ANESTHESIA.

TEMPOROMANDIBULAR JOINT. See TOOTH AND JAW, BIOMECHANICS OF.

TENDON, PROPERTIES OF. See LIGAMENT AND TENDON, PROPERTIES OF.

TENS DEVICES. See TRANSCUTANEOUS ELECTRICAL NERVE STIMULATION (TENS).

THERMISTORS

PEDRO LOPES DE MELO
State University of Rio de
Janeiro
Brazil

INTRODUCTION

The normal body temperature of human beings is considered to be constant $\sim 37^\circ\text{C}$. Heat energy is stored in the body

and it is essentially constant, as long as we are alive. Most of the heat produced in the organism comes from deep organs, mainly the liver, brain, and heart, as well as the skeletal muscles during exercise (1). A major part of this heat is dissipated at the skin surface by radiation, convection, evaporation, and respiration. However, as the environment temperature varies frequently, the body presents mechanisms to generate and loose heat, controlling and regulating the body temperature. It is done almost exclusively using feedback mechanisms that operate by means of regulatory centers located at the hypothalamus. Thus, if the environment temperature decreases, the body generates more heat and tries to keep it, while the heat generation is decreased and part of the body heat is dissipated when the environment temperature increases. Disease states introduces modifications in this equilibrium that can be indirectly evaluated by body temperature. That is the reason why it is one of the more important physiological parameter, measured virtually in all patients in every hospital bed, being used as a clinical index of disease, as well as an aid for diagnostic and prognostic purposes. Temperature can be measured at the human body in two basic ways: at skin surface and by systemic measurements. Although, in practice, systemic measurements are much more commonly used, both kinds provide valuable diagnostic information. Skin temperature is the result of the relationship between the heat supplied by blood surface circulation, the environmental temperature, and the air circulation around the area at which the measurement is being taken. Skin temperature data are usually obtained from sensors in contact with the skin's surface. However, the use of self-adhering foam patches to affix temperature sensors to the skin appears to procedure artificially higher mean temperatures (2). Systemic temperature is the temperature of the internal regions of the body. This parameter can be measured by temperature sensing devices placed at the mouth, in the rectum, or under the armpits. The oral temperature of a health person is $\sim 37^\circ\text{C}$. The rectal temperature is typically 0.5°C higher than the oral temperature, while the under-arm temperature is about 1°C lower than the oral one. Some clinical applications may need more specific measurements, as skull and core temperatures, for example. Skull temperature may be obtained by placing the thermal probe into the nasopharynx near the base of the skull, while core temperature may be obtained from a probe inserted into the esophagus. The basic characteristics of devices used in temperature measurements are ease of operation, cleaning and sterilization, and guarantee of patient safety. It is also important to mention the size and thermal mass of the probe, since these determine the disturbance imposed by the measurement and the speed of response.

The simplest phenomenon for temperature sensing is, perhaps, thermal expansion, which is the basis of the mercury-in-glass thermometer. In clinical routine measurements, however, electronic temperature recording and display are advantageous, since they permit the use of automatic continuous measurement of temperature and remote applications by connecting to computerized systems. Another convenient characteristic of these devices is their small size. There are numerous ways of measuring temperature electronically, most commonly by transducers

based on temperature dependence of electric resistance (RTDs and thermistors) and thermoelectric effects (thermocouples). Thermistors offer a high sensitivity and degree of interchangeability at lower cost than either RTDs or thermocouples. It makes them ideal for healthcare products that incorporate sensor probes that can be discarded after using and replaced with new probes of the same specification, without recalibration. In this section, the attention will be focused on thermal-sensitive resistors, their properties, and the basic instrumentation used in temperature measurements. Examples of commercial devices dedicated to temperature measurements are also discussed, as well as systems for clinical and research applications in several branches of medicine to measure flow, detect the presence of fluids, and evaluate the properties of tissue based on temperature measurements using thermistors.

THEORY

Thermistor (the contracted name of thermally sensitive resistor) is a general term used for both positive and negative temperature coefficient types of semiconducting thermal transducers. These devices are constructed of ceramic materials whose electrical conduction properties are temperature sensitive. At a fixed environmental temperature, a thermistor exhibits a specific ohmic value. However, if the environment temperature varies, this resistance changes. Thermistors are able to sense temperatures from -50 to 300 °C, which is a small range compared with metal wire sensors. A platinum wire, for example, can be used to measure temperature from -160 to 1000 °C. This range, however, is much greater than most of the medical temperature measurement tasks would require.

Thermistors usually have high negative temperature coefficients (NTC), resulting in a decrease of the thermistor resistance with increasing temperature. The negative temperature coefficient in nonmetallic materials (silver sulfide), was first observed by Michael Faraday in 1833, but it was not until 1940 that thermistors were developed that were able to produce reproducible results. The NTC thermistors are typically made of transition-metal oxides. The most usual oxides are those of manganese, nickel, cobalt, and iron (3,4). In the basic fabrication process, a mixture of two or more metal oxide powders are first combined with suitable binders, and are molded to a desired geometry. Then the products are dried and sintered at an elevated temperature. The units are finally coated with an epoxy layer for final protection and stabilization. Varying the types of oxides used, their relative proportions, the sintering atmosphere, and the sintering temperature, a wide range of resistivities and temperature coefficient characteristics can be obtained. More detailed descriptions of NTC manufacturing techniques can be obtained in Refs. 3–5. Since their first use in practical electronic thermometer in the early 1950s, thermistor technology has been enhanced continuously, resulting in improvements that probably situate this device as the most widely used temperature transducer for medical applications nowadays. These devices have been developed to be very sensitive to variations in temperature (~ -3 to -5 °C), present excellent

long-term stability ($\pm 0.2\%$ of nominal resistance value per year), and be small in size. Because of this small size, these sensors present a fast response to variations in fluid temperature. Thermistors are also relatively inexpensive.

Beyond the advantages of high sensitivity, interchangeability, and low cost cited previously, another major advantage offered by thermistors is that, unlike RTDs and thermocouples, thermistors are virtually unaffected by lead resistance, since they are high resistance devices. Specifically comparing them with thermocouples, we observe that thermistors can be used with less complex and expensive instruments. It happens because the thermal electromotive force (EMF) values produced by thermocouples are around a few microvolts per degree, requiring high gain and low noise amplification of the signal. Moreover, thermocouple demands additional circuits for compensation of cold junction temperature (4). There are many physical configurations in which thermistors are found, varying from very small bead thermistors, that are spherical and have diameters as small as 0.1 mm, to large flat disks having diameters of several centimeters. Some of these configurations are described in Fig. 1.

Thermistors with positive temperature coefficient (PTC) may also be constructed, by sintering barium and strontium titanate mixtures. These thermistors are often called switching thermistors because of their resistance–temperature characteristics. As temperature increases, the zero-power resistance of this device remains essentially constant until reaching the switching temperature or Currie point, where there is a sharp upward increase in the resistance. The switching temperature can be from -20 to $+125$ °C. The PTC thermistors are frequently used as thermostats to sense and regulate oven temperature (6) and for circuit protection.

THERMISTOR TERMINOLOGY

Dissipation constant (DC or δ) is the amount of power required to raise the temperature of the thermistor 1 °C above the surrounding temperature in steady-state conditions (6). It means that the resistance changes by an equivalent of 1 °C for each dissipation constant rating ($\text{mW} \cdot \text{°C}^{-1}$) for the selected device. It depends on the heat transfer from the thermistor to its surroundings (by conduction through the leads, free convection in the medium and radiation), the relative motion of the medium in which the thermistor is located, as well as the thermal conductivity. A typical value of the dissipation constant of a thermistor with a 0.24 cm outer diameter, coated with epoxy or phenolic layers, is $2 \text{ mW} \cdot \text{°C}^{-1}$ in still air (7). Its parameter increases with thermistor mass and in water is ~ 5 – 10 times the value measured in air.

Maximum operating temperature is the maximum body temperature at which the thermistor will operate for an extended period of time with acceptable stability of its characteristics. This temperature can be the result of internal or external heating, or both, and should not exceed the maximum value specified (6).

Self-heating is a process observed when a current flowing through a thermistor causes sufficient heating

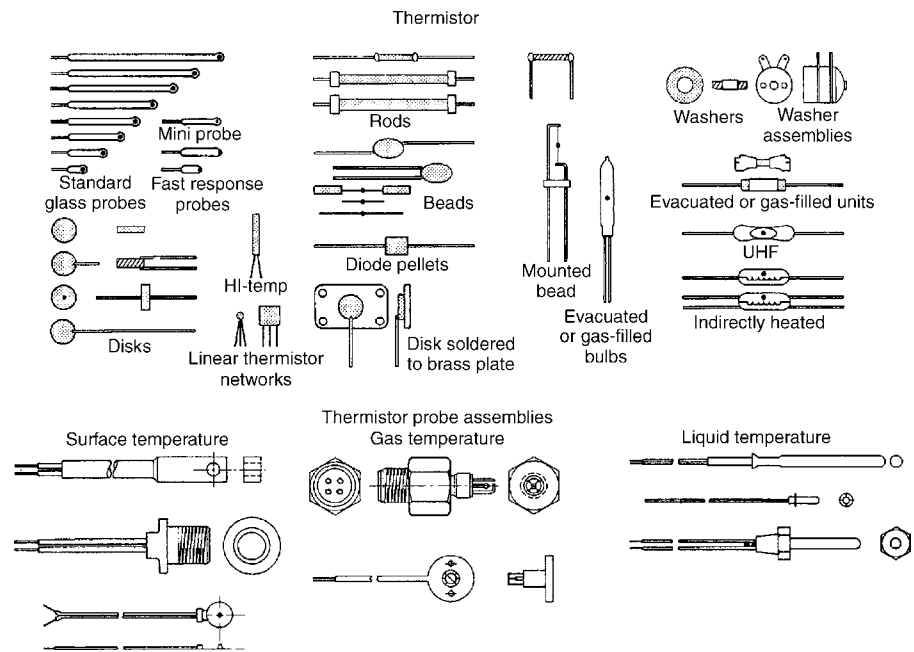


Figure 1. Some thermistor configurations (from *Interfacing Sensors to the IBM PC*, by W. J. Tompkins and J. G. Webster, 1988 Reprinted by permission of Prentice-Hall).

by power dissipation to raise the thermistor's temperature above the ambient (6). As the effects of self-heating are not always negligible (or may even be intended), a distinction has to be made between the characteristics of an electrically loaded thermistor and those of an unloaded thermistor. The properties of an unloaded thermistor are also termed as zero-power characteristics. This effect depends on the thermal dissipation factor (δ) and the geometry of the thermistor itself. In general, the smaller the device, the smaller is the permissible current before self-heat.

Stability is the ability of a thermistor to retain specified characteristics after being subjected to designated environmental or electrical test conditions ($^{\circ}\text{C}\cdot\text{year}^{-1}$). This parameter is dependent on environmental conditions (e.g., humidity, excessive temperature, and thermal shock), which should be minimized to guarantee stability (7). Physical reasons for this may be thermal stress causing a change in concentration of lattice imperfections, oxygen exchange with the environment (with unprotected, nonglass-encapsulated thermistors), or diffusion in the contact areas of metalized surface contacts. To enhance long-term thermistors stability, they are usually subjected to an ageing process directly after manufacture (8). This results in chemically stable devices that are not significantly affected by aging, exhibiting typically $<0.02^{\circ}\text{C}\cdot\text{year}^{-1}$ of thermometric drift.

Zero-power resistance (R_0) is the direct current (dc) resistance value of a thermistor at a specified temperature, with negligible electrical power to avoid self-heating (6).

Temperature coefficient of resistance (α) is a useful measurement of the thermistor's sensitivity and is defined as the relative change in resistance referred to change in temperature at a specified temperature ($\% \cdot ^{\circ}\text{C}^{-1}$) under zero-power conditions (6,8).

$$\alpha = \frac{1}{R_T} \frac{dR}{dT} \quad (1)$$

Because the relationship between resistance and temperature is not linear, α is a function of temperature and it is usually specified over the temperature range where the temperature variation is expected. One of the most important characteristics of a thermistor is, without question, its extremely high temperature coefficient of resistance. A typical value is $-4.3\% \cdot ^{\circ}\text{C}^{-1}$ at 37°C .

Thermal time constant (TC or τ) describes the heat inertness of thermistors (5) and it is the time (s) required for a thermistor to change 63.2% of the total difference between its initial and final body temperature, when subjected to a step function change in temperature under zero-power conditions. This parameter is directly affected by the mass of the thermistor, the thermal properties of the medium surrounding the device, the thermal coupling to the environment, the motion of the medium, the conduction through the leads, and the radiation losses. An epoxy coated thermistor with a 0.24 cm outer diameter will typically have a time constant of 0.75 s in stirred oil and 10 s in still air (7).

Interchangeability tolerance: The rated thermistor parameters values are subject to manufacturing tolerances. They are determined by the composition and structure of the various metal oxides being used in the device production. The result will be a variation from unit to unit within a production lot, as well as from lot to lot. Interchangeability tolerance is the value of how far a specific family of devices may be from the nominal resistance versus temperature curve. For example, if a family of devices has an interchangeability tolerance of $\pm 1.0^{\circ}\text{C}$ over the range from 0 to 70°C , it means that, for this range, all devices of this family are within $\pm 1.0^{\circ}\text{C}$ of the resistance versus temperature curve. This feature results in accurate temperature measurements to $\pm 1.0^{\circ}\text{C}$, no matter the substitution of thermistors. Modern thermistor technology results in the production of devices with tight zero-power resistance tolerances. Over the medical temperature range, interchangeable tolerances to $\pm 0.1^{\circ}\text{C}$ are typical (9,10).

Maximum power rating is the maximum power (mW) that a thermistor will dissipate for an extended period of time with acceptable stability of its characteristics (11).

Standard reference temperature is the thermistor body temperature at which nominal zero-power resistance is specified and is usually 25 °C (11).

RESISTANCE-TEMPERATURE CHARACTERISTICS

Thermistors are one member of the family of resistive temperature sensors. Unlike the other members of this family, which presents linear relationship between resistance and temperature (platinum, nickel and cooper, e.g.), thermistors are very nonlinear. This can be seen in Fig. 2, where the characteristics of some typical NTC thermistors and a platinum RTD are compared, showing the higher sensitivity to temperature changes of the first ones (3). In thermistor literature, the most frequently used characteristic relationship between the thermistor resistance, and the ambient temperature is

$$R_T = R_0 \exp \left[\beta \left(\frac{1}{T} - \frac{1}{T_0} \right) \right] \quad (2)$$

where R_T is the thermistor resistance (Ω) at temperature T , R_0 is the thermistor resistance (Ω) at temperature T_0 , which is the standard reference temperature (K, $K = ^\circ\text{C} + 273.15$), and β is the material constant for thermistor (K). A typical thermistor resistance may vary from 5000 Ω at 0 °C, until 100 Ω at 150 °C, while the reference temperature T_0 is usually taken as 298 K (25 °C). Equation 2 can be rearranged to solve for β :

$$\beta = \frac{T T_0}{T - T_0} \ln \frac{R_0}{R_T} \quad (3)$$

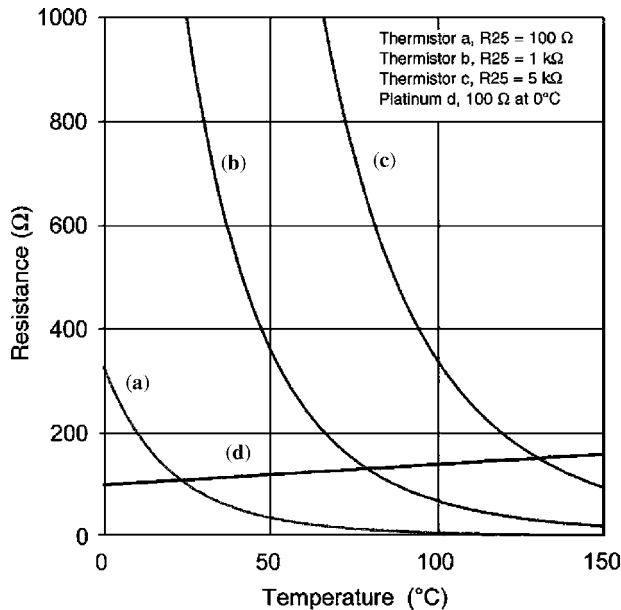


Figure 2. Resistance-temperature relationship of thermistors (a,b,c) and RTD (d). Over the presented range, NTC thermistors offer higher sensitivity to temperature changes compared to RTD. (Reproduced from (3) with permission).

The constant β is usually between 2000 and 5000 K and increases slightly with temperature. Its value is determined by the material properties and it is usually calculated and specified by thermistor manufacturers using two temperatures over a given range. Applying the definition of the temperature coefficient of resistance (eq. 1) in equation 2 results in

$$\alpha = -\frac{\beta}{T^2} \quad (4)$$

It is important to point out that equations 2–4 are valid only over small temperature spans. The Steinhart–Hart equation is, however, more accurate over wider temperature ranges.

$$\frac{1}{T} = a + b(\ln R_T) + c(\ln R_T)^3 \quad (5)$$

where T is the temperature in K, a , b , and c are the coefficients derived from measurements. The Steinhart–Hart equation is an empirically developed polynomial that has been determined to be the best mathematical expression for resistance-temperature relationships of NTC thermistors and probes (12). Solving for resistance, when temperature is known, the form of the equation changes to

$$R_T = e \left[\left(-\frac{\chi}{2} + \left\{ \frac{\chi^2}{4} + \frac{\psi^3}{27} \right\}^{1/2} \right)^{1/3} + \left(-\frac{\chi}{2} - \left\{ \frac{\chi^2}{4} + \frac{\psi^3}{27} \right\}^{1/2} \right)^{1/3} \right] \quad (6)$$

where, $\chi = (a - 1/T)/c$ and $\psi = b/c$. The a , b , and c coefficients, can be solved measuring the thermistor resistance (R_1, R_2, R_3) at three different temperatures (T_1, T_2, T_3) and using simultaneously equation 5. The data calculated by equations 5 and 6 will be accurate to better than ± 0.01 °C, when $T_1 \leq -40$ °C, $T_3 \leq 150$ °C, $|T_1 - T_2| \leq 50$ °C and $|T_2 - T_3| \leq 50$ °C. Parameters T_1 , T_2 , and T_3 are evenly spaced and at least 10 °C apart.

VOLTAGE-CURRENT CHARACTERISTICS

If a constant electrical power is applied to the thermistor its temperature will first increase considerably, but this change will decline with time. After some time a steady state will be reached, where the power is dissipated by thermal conduction or convection. The voltage drop on the thermistor as a function of the flowing current under conditions of thermal equilibrium is

$$VI = \delta(T_T - T_A) \quad (7)$$

If the dissipation constant (δ) variations are negligible for a determined medium and established conditions, and the resistance-temperature characteristic is known, the static current-voltage characteristics can be obtained. Since $V = R_T I$, where R_T is the temperature dependent NTC resistance,

$$I = \sqrt{\frac{\delta(T_T - T_A)}{R_T}} \quad (8)$$

and

$$V = \sqrt{\delta(T_T - T_A) R_T} \quad (9)$$

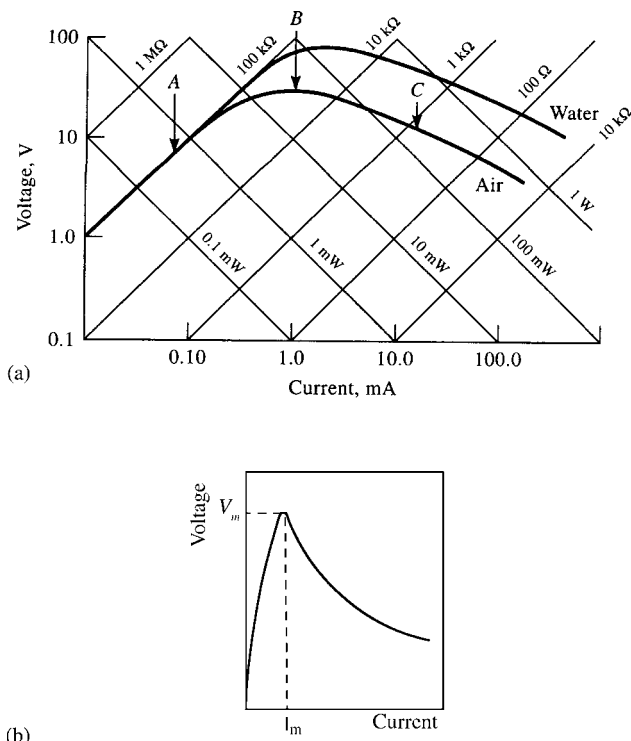


Figure 3. Thermistor voltage versus current characteristics with the device in air and water on log-log coordinates (A), and characteristics on linear coordinates (B). Figure (A) is from *Interfacing Sensors to the IBM PC*, by W. J. Tompkins and J. G. Webster, 1988 (Reprinted by permission of Prentice-Hall) and Figure (B) is from *Ceramic Sensors: technology and applications*, T. G. Nenov and S. P. Yordanov, 1996, (courtesy of Technomic Inc, Pennsylvania, USA).

using the above equations the voltage–current curves can be plotted on log–log coordinates, where lines of constant resistance have a slope of +1 and lines of constant power have a slope of –1 (Fig. 3a) (6,13). For some applications, it is more convenient to plot the voltage–current characteristic on linear coordinates (Fig. 3b) (5,13).

The voltage–current characteristics of a NTC thermistor has three different sections: first, for small currents, the amount of power dissipated in the thermistor is negligible, and the voltage–current characteristic will describe a constant resistance that is equal to the zero-power resistance of the device at the specified ambient temperature (section A in Fig. 3a). In this condition, the temperature of the thermistor is that of the surroundings. This curve section is usually used when thermistors are employed as temperature sensors. As the current continues to be increased, in the second curve section, there is a rise in the thermistor temperature above that of the surrounding medium since this power cannot be completely removed from the thermistor and the effects of self-heating become evident, resulting in a decrease in its resistance (section B in Fig. 3a). In this condition, more current flows due to decreasing resistance (6). Subsequent incremental increases in current introduce a corresponding decrease in resistance. Hence, the slope of the voltage–current characteristic ($\Delta E/\Delta I$) decreases

with increasing current. This continues until a current value is reached (I_m), for which the slope becomes zero and the voltage reaches a maximum value (V_m) (Fig. 3b). As the current is increased above the value of I_m , the third curve section (section C in Fig. 3a) is entered. The slope of the characteristic curve continues to decrease and the thermistor exhibits a negative resistance characteristic (5). This last section of the operating area of NTC thermistors, when they are self-heated, is used in applications such as liquid level detection, air flow detection, and thermal conductivity measurement. However, care should be taken when designing a circuit for self-heated application, because the thermistor is vulnerable to destruction in this region by thermal runaway. This can be avoided by passing constant currents coherent with the thermistor dissipation characteristics. As the power dissipated is proportional to I^2R , when the current is constant, further increases in temperature causes decreased resistances and decreased power, protecting the device.

Anytime a thermistor is applied for temperature measurement, it is very important to avoid its self-heat. Since a thermistor resistance changes with the temperature, this self-generated heat will change this resistance value, producing an erroneous reading. For example, if the dissipation constant of the thermistor selected is $5 \text{ mW}\cdot\text{C}^{-1}$ and the power dissipated by the device is $20 \text{ mW}\cdot\text{C}^{-1}$, then a 4°C error is induced due to the effect of self-heating. This effect is more pronounced when dealing with a still fluid (i.e., neither flowing nor agitated), because there is less carry-off of the heat generated. This kind of problem does not arise with thermocouples, that are essentially zero-current devices. To maintain a higher degree of accuracy, the temperature error caused by self-heating should be an order of magnitude less than the required sensor accuracy. As an example, if the power dissipation constant of a thermistor is $\sim 2 \text{ mW}\cdot\text{C}^{-1}$ in still air, in order to keep the self-heat error $< 0.1^\circ\text{C}$ the power dissipation must be $< 0.2 \text{ mW}$. Very low current levels are required to obtain such a low power dissipation factor. This mode of operation is usually called zero-power sensing.

THERMAL CHARACTERISTICS

The power dissipated by a thermistor (P) is equal to the rate at which thermal energy (H) is supplied to the thermistor. This is the same as the rate at which energy is lost from the thermistor to its surroundings (H_L), plus the rate at which energy is absorbed (H_A) (13,14).

$$P = \frac{dH}{dt} = \frac{dH_L}{dt} + \frac{dH_A}{dt} \tag{10}$$

The rate at which thermal energy is lost from the thermistor is proportional to the temperature raise of the thermistor,

$$\frac{dH_L}{dt} = \delta(T_T - T_A) \tag{11}$$

The dissipation constant (δ) is typically measured under equilibrium conditions. It is not a true constant, since it varies slightly with temperature and with an increase in

temperature. The following relation can express the rate at which thermal energy is absorbed by the thermistor to produce a specific amount of rise in temperature:

$$\frac{dH_A}{dt} = sm \frac{dT_T}{dt} \quad (12)$$

where s is the specific heat and m is the mass of the thermistor. Thus, the heat-transfer equation for an NTC thermistor, at any instant after the application of power to the circuit, can be expressed as

$$P = RI^2 = \frac{dH}{dt} = \delta(T_A - T_T) + sm \frac{dT_T}{dt} \quad (13)$$

The thermal transient conditions at turn-on is given by the solution of equation 13, which is obtained considering P constant:

$$T_T - T_A = \frac{P}{\delta} \left[1 - \exp\left(-\frac{\delta}{sm}t\right) \right] \quad (14)$$

It means that when a significant amount of power is dissipated in a thermistor, its body temperature will rise above the ambient temperature as a function of time. In steady-state condition, when thermal equilibrium is achieved ($dT_T/dt = 0$ in equation 13, or when $t \gg ms/\delta$ in equation 14), the rate of heat loss is equal to the power supplied to the thermistor (remind equation 7, where $VI = \delta(T_T - T_A)$). When self-heating is negligible ($P \cong 0$), the heat-transfer equation 13 can be rewritten,

$$\frac{dT_T}{dt} = \frac{-\delta}{sm}(T_T - T_A) \quad (15)$$

which can be solved to

$$T_T = T_A + (T_I - T_A) \exp\left(\frac{-t}{\tau}\right) \quad (16)$$

where T_I is the initial body temperature and τ is the thermal time constant ($\tau = ms/\delta$), which depends on the same environmental factors as δ .

All of the preceded discussions of thermal properties of NTC thermistors have been based upon a single device structure with a single time constant. When these devices are encapsulated into sensor housings, the simple exponential response no longer describes adequately the system response. The mass of the housing and the thermal conductivity of the materials used in the sensor housing will normally increase the system dissipation constant, increasing the thermal response time. In this case, the thermal properties are somewhat difficult to predict by mathematical modeling, usually requiring experimental tests of the system to obtain data on the resulting response time and dissipation constant (13). In general, the thermal response time may be reduced keeping the thermal resistance between the actual temperature sensor and the tissue being measured as low as possible.

THERMISTOR LINEARIZATION

The inherent nonlinearity of the resistance versus temperature characteristics of thermistors is rather troublesome in many applications. Even if we are interested in

temperature measurements close to the body temperature, with variations of only a few degrees, the thermistor must be linearized. This can be accomplished basically in three ways: modifying the transducer circuitry, using analog circuits, or using digital techniques (15).

In the first option, the output of the transducer can be linearized over a limited temperature range with the addition of series or parallel resistors (6,16,17). It causes the voltage or the resistance of a simple fixed resistor thermistor to have zero error along a linear temperature scale at three equidistant points (Fig 4). The series resistance (R_s) required to make the conductance-temperature characteristic of a thermistor approximately linear may be calculated as follows (16,18):

$$R_s = R_{T_m} \left(\frac{\beta - 2T_M}{\beta + 2T_M} \right) \quad (17)$$

Where R_{T_m} is the resistance of the thermistor at the mid-scale temperature T_M . As can be observed in Fig. 4, this procedure reduces the sensitivity of the transducer. However, as the sensitivity of a thermistor is relatively high, the reduction is often a satisfactory tradeoff. The temperature coefficient of the series combination (α_s) is given by (6)

$$\alpha_s = \frac{-(\beta/T_M)^2}{(G_{T_M}/G_s) + 1} \quad (18)$$

Where G_{T_M} is the conductance of the thermistor at the mid-scale temperature and G_s is the conductance of the series resistance. An alternative method is to define the low, mid, and high end of the desired temperature range at three

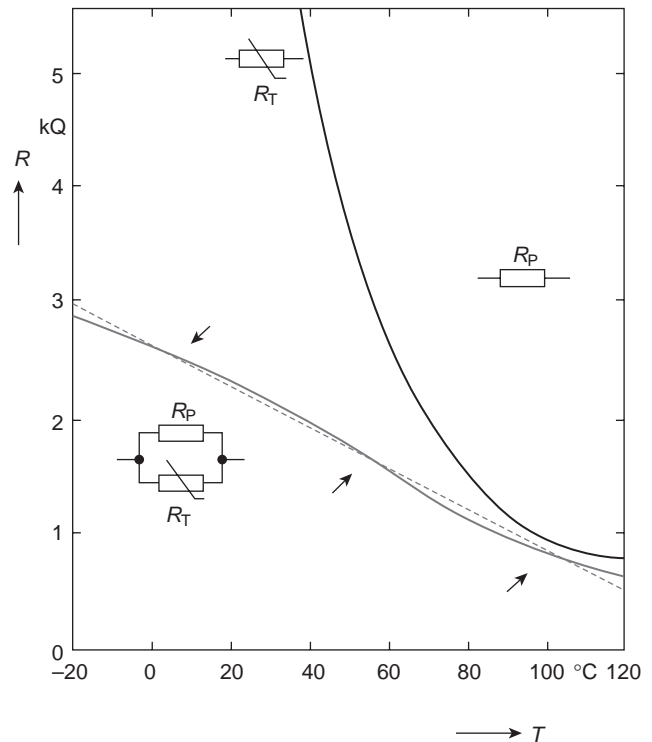


Figure 4. Linearization of a thermistor using a parallel resistor. (Adapted from (17) with permission).

equidistant points, and calculate R_s by (19,20):

$$R_s = \frac{[R_{T_M}(R_{T_{LO}} + R_{T_{HI}}) - (2R_{T_{LO}}R_{T_{HI}})]}{[(R_{T_{LO}} + R_{T_{HI}}) - (2R_{T_M})]} \quad (19)$$

where R_{LO} , R_{T_M} , and R_{HI} are the thermistor resistance at the low, mid, and high end of the range, respectively. In this procedure, the addition of series resistor R_s forces the equivalent resistance of the fixed resistor thermistor to have zero error along a linear temperature scale in the three points chosen. The temperature range of the application determines the maximum error. For example, if the range is taken to be -50°C to 100°C , the errors are 0 at -50 , 25, and 100°C , and the errors elsewhere are distributed in an S-shaped. If the temperature range is reduced, the errors become smaller; being 2.0°C over a 60°C range, 0.05°C over a 30°C range, and 0.01°C over a 10°C range. Another way to linearize thermistors includes the use of a Wheatstone bridge. In most applications, the bridge consists of a linear thermistor voltage divider and a fixed resistor voltage divider, as described in Fig. 5a (21). This circuit is designed to produce 1 V at 25°C and 200 mV at 45°C , with an output voltage that is linear within $\pm 0.06^\circ\text{C}$ in this temperature range (Fig. 5b). Hoge (22) showed that circuits based on resistors (serial, parallel, or bridge) are equal in their ability to linearize thermistor

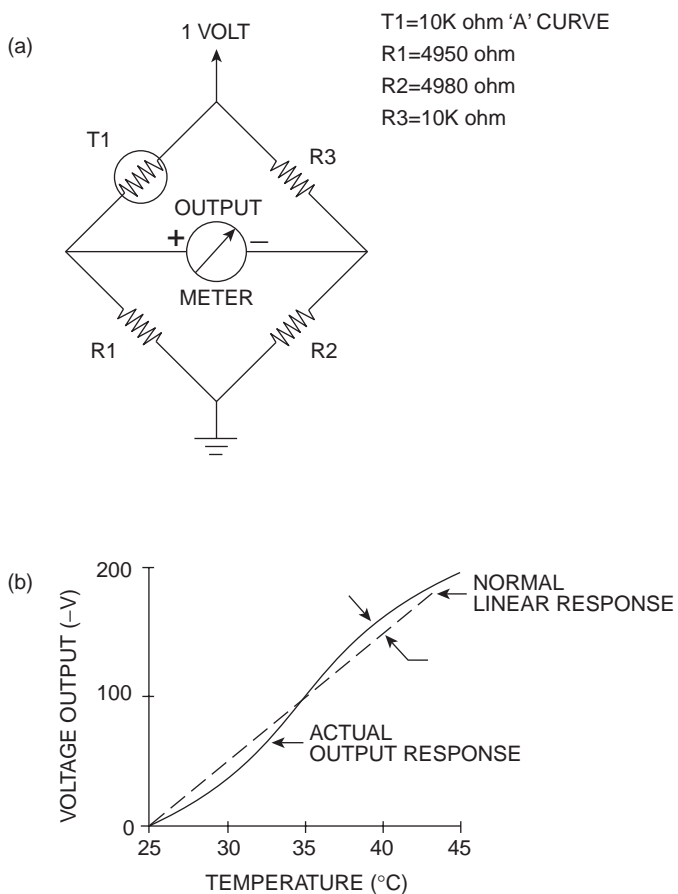
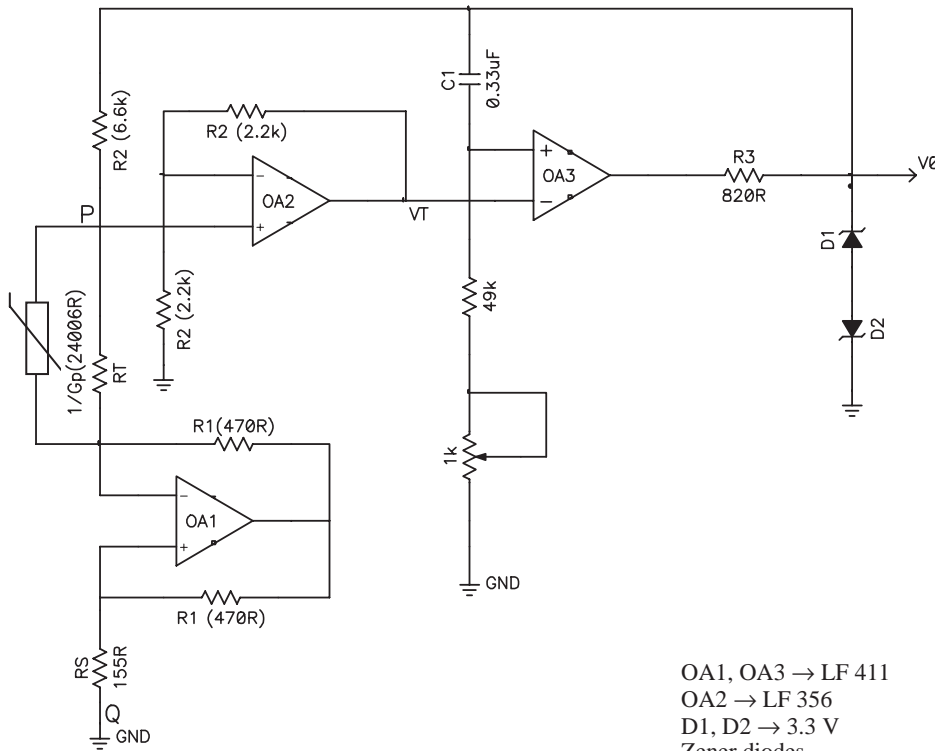


Figure 5. Use of a thermistor in a Wheatstone bridge for temperature measurements (A) and respective output voltage as a function of temperature (B) (courtesy of Alpha Sensors, 2121 Palomar Airport Rd., Carlsbad, CA 92009, USA).

characteristics. Thus, the choice of a linearizing circuit should be made on other grounds, such as simplicity or convenience.

Theoretically, an improved linearization over wider temperatures can be achieved using more complex circuits. For many applications, it can be a better option to linearize the transducer at some point of the analog process, as in cases where no digital processing is used. It is also true when limited processing capability and/or memory are available, and the analog processing can be done simply and at low cost. Analog circuitry using piecewise-linear approximations has been developed to be used with temperature transducers (23). However, these circuits are complex and costly, and are not usually used in practice. Voltage-to-frequency converters (24), logarithmic (25), and temperature to frequency (26) circuits, were proposed for this objective. In 1990 Slomovitz and Joskowicz (27) compared the quality of these active linearizing circuits. They concluded that the errors (2, 1, and 0.7K , respectively) were in the same order as those obtained in single resistor circuits. The authors attributed the origin of the errors to the fact that all of them were based on the simple exponential approximation. More recently, Kaliyugavardan et al. (28) proposed a method for linearization of thermistor response using series-parallel resistors based on a new four-constant curve fit method, which resulted in a temperature-to-frequency converter that provides accuracy better than 0.2K . The same researchers obtained even more accurate results by using the circuit described in Fig. 6, which works essentially as an astable multivibrator. Experimental results obtained using a standard thermistor in the range of $35\text{--}95^\circ\text{C}$ revealed a peak error less than $\pm 0.1\text{K}$ (29). Recently, a signal-conditioning circuit by generating a compensating, pseudologarithmic response function was proposed (30).

If the data are to be digitized and processed digitally, as soon as possible, it probably makes sense to perform any needed linearization in the digital domain. The techniques used in this case allow linearization to be done much more efficiently and accurately in software, and eliminate the need of tedious manual calibration using multiple and sometimes interacting trimpots. The principal techniques involve look-up tables and computational algorithms (20,31,32). A look-up table may be constructed, for example, by using a read only memory (ROM) hardwired to the analog-to-digital converter output. Each level from the converter corresponds to an address in ROM, and the word stored in that address is the linearized value. It can be constructed by using the device characteristics, available from the manufacturer. The Steinhart-Hart equation can be used to create the table, which implements a third-order linearization formula that provides high accuracy. If memory is limited, and the measurement is made infrequently, but rapid mathematics is available (as, e.g., in digital signal processing systems), a mathematical function that approximates the inverse of the nonlinear relationship, or the difference between the ideal signal and the actual signal, can be derived and stored in program memory. Then, whenever the measurement is made, the processor computes the correct value, based on its mathematical relationship to the measured input variable (23). In system



OA1, OA3 → LF 411
 OA2 → LF 356
 D1, D2 → 3.3 V
 Zener diodes

Figure 6. Linear temperature to frequency converter (Reproduced from (29) with permission).

including microcontrollers, in many cases the mathematical functions may contain complex polynomial and exponential functions, placing a great burden on the program memory, RAM, and execution speed of most low cost devices. Digital piecewise linear interpolation may be a good choice for sensor linearization in such systems due to its fast execution speed, reduced program memory requirements, and easy of implementation (33).

ELECTRONICS FOR TEMPERATURE MEASUREMENTS

In the circuit project, simulation and analysis environment, SPICE subcircuits for thermistor modeling are useful, allowing for a realistic simulation of thermistor parameters for all standard analysis [transient, alternating current(ac), and dc]. Useful subcircuits have been described by Keskin (14), Hagerman (34) and Wangenheim (35), and are also commercially available (<http://www.catenauk.com>; <http://www.intusoft.com/products>). As discussed earlier, there are a variety of circuits in which thermistors may be used for temperature measurements. Caution must always be taken, however, to insure that the power dissipated in the thermistor is held at a minimum and that the current flow is insufficient to cause self-heating, since temperature measurements require that the thermistor be operated in a zero-power condition. The most common technique implies the use of a constant current source, and the measurement of the voltage developed across the thermistor.

Current Sources

In the section dedicated to the thermal characteristics of thermistors, it was shown that the use of current source

would prevent thermal runaway in thermistors. The simplest circuit that can approximate a current source is a high voltage source with a high resistance connected in series (voltage divider), as can be seen in Fig. 7. In this circuit, the output voltage is taken across the fixed resistor, and, the higher this resistor in relation to the thermistor resistance, and the higher the voltage, the closer this circuit will conform with an ideal current source. From the plot of the output voltage (Fig. 7b) it can be observed that there is a range of temperatures where the circuit is reasonably linear with good sensitivity. However, this simple circuit does not act like an ideal current source, since modifications in the thermistor resistance introduces alterations in current. An even better approximation to the desired current source would be obtained using standard circuits based on operational amplifiers (16,36). Integrated current sources are also useful, allowing for the configuration of regulated current sources of varying magnitudes. Fig. 8 shows a typical example, which is based on a device containing two low current regulators (37). One of current regulators supplies $100\ \mu\text{A}$ to the thermistor. The temperature of the thermistor is converted into a voltage that is increased by R_3 , filtered by R_2 - C_1 and amplified by U_{1B} . The second current source is used to provide the reference voltage in combination with R_1 and U_{1a} . This circuit is a useful framework for thermistor temperature measurements using analog-to-digital converters (37).

Wheatstone Bridges

The wheatstone bridge is a widely used means of measuring temperature using thermistors, since the bridge aids the linearization of the NTC (Fig. 5). The condition for the

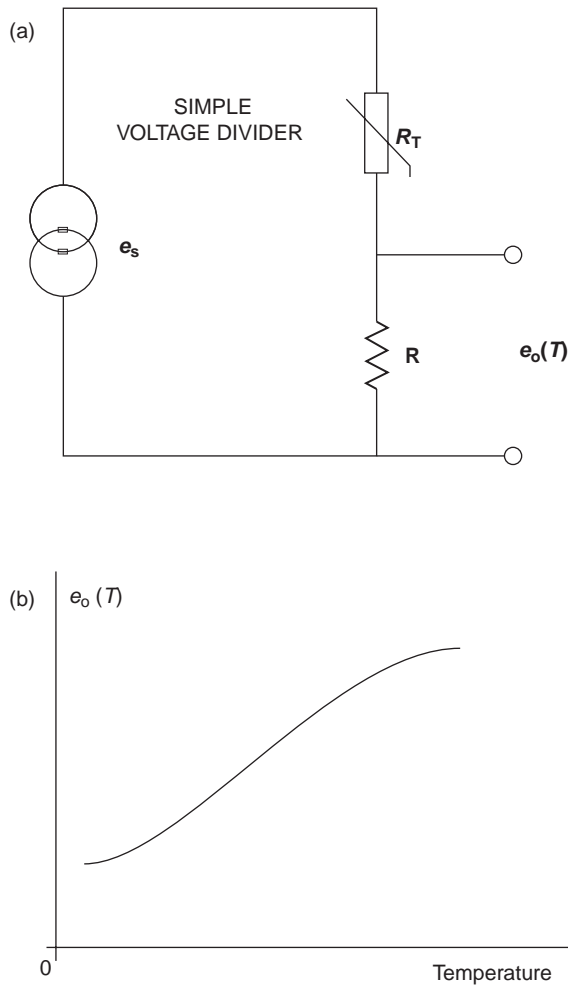


Figure 7. Simple series circuit to approximate a constant current source (A) and corresponding output voltage (B).

balance of the bridge is $T_1R_2 = R_1R_3$. In a normal operation, R_3 is usually a variable resistor that is used to zero the bridge. If the bridge is excited with a constant dc voltage, the bridge output may be displayed in a dc meter, calibrated directly in temperature units. In medical care, it is often necessary to accurately determine very small temperature differences. For example, it may be necessary to measure the difference in temperature between two regions of an organ. In cases involving differential temperature measurements, matched thermistors are placed in the two voltage dividers. In Fig. 5, the second thermistor replaces R_3 . With this configuration, temperature differentials as close as 0.001°C can be readily detected.

Temperature-to-Frequency or Temperature-to-Time Interval Converters

For applications in which digital processors or microcontrollers are used for data acquisition and signal processing, the transducer response must assume a form suitable for conversion to digital format. Temperature-to-frequency or temperature-to-time interval conversions are convenient methods to measure temperature in this case, since they permit an easy and low cost interface, no ADC is needed and only one bit of input is necessary (6). Another advantage is that the optical isolation circuits used in this kind of application presents lower cost than that used in linear systems. Fig. 9 shows two simple circuits in which the frequency of oscillation varies with the temperature. Because a thermistor resistance varies with temperature, the RC time constant will change accordingly. And since the 555 timer determines the frequency corresponding to the RC time constant, the actual resistance is proportional to the number of counts recorded by the 555 timer (Fig. 9a). The frequency, or the number of counts in a given time window, can be easily converted to a temperature value (38,39). In the circuit described in Fig. 9(b), the thermistor is placed in the feedback loop of a hysteresis-based oscillator, and the output frequency is related to temperature,

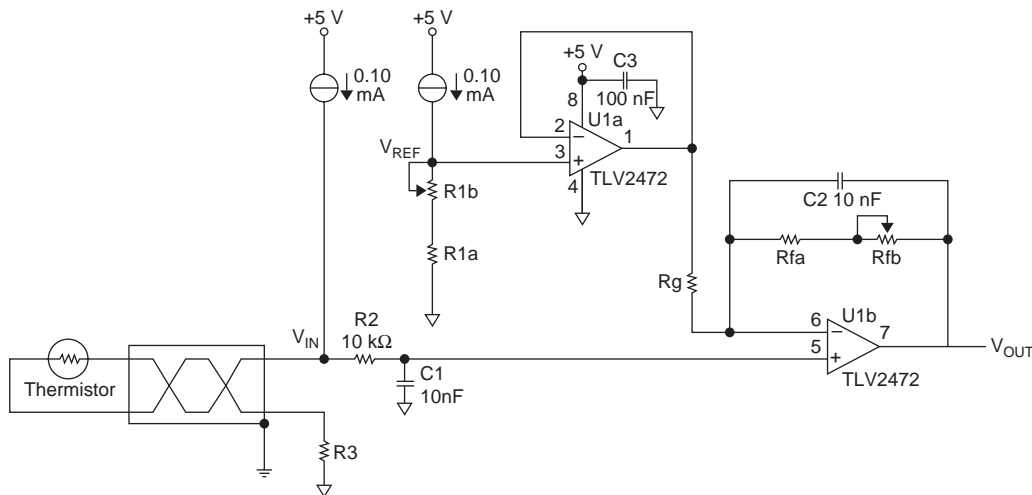


Figure 8. Circuit containing two integrated low current regulators. Courtesy of Texas Instruments Inc. 13532 N. Central Expressway M/S 3807 Dallas, Texas, USA (Adapted from (37) with permission).

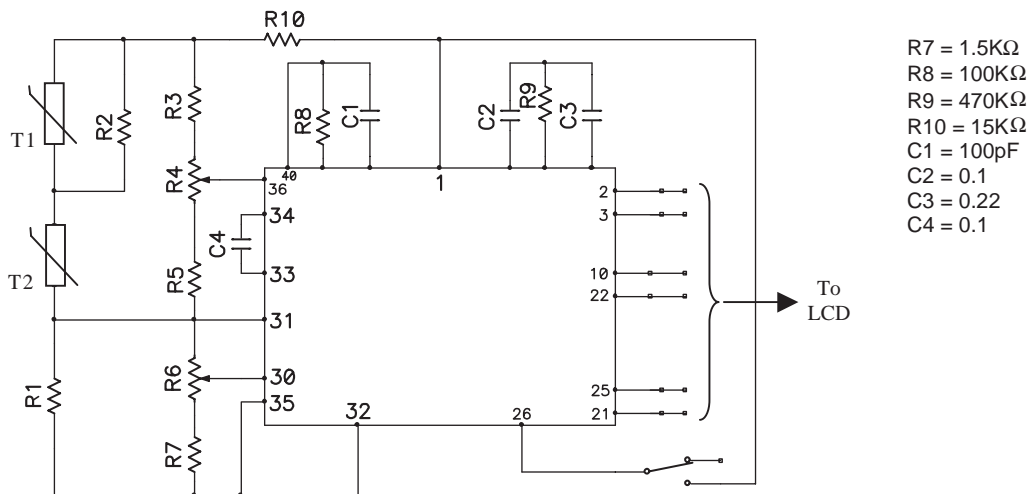


Figure 11. Digital thermometer based on 3-1/2 digits voltmeter (courtesy of Alpha Sensors, 2121 Palomar Airport Rd., Carlsbad, CA 92009, USA).

High Resolution Measurements

Very low modifications in temperature can be sensed by using lock-in amplifiers (40). These amplifiers are used to measure the amplitude and phase of repetitive ac signals buried in noise. It is achieved by their ability of acting as a narrow bandpass filter, which removes unwanted noise while allowing through the signal that is to be measured. The ac frequency of the signal to be measured is used as a reference signal to set the passband region of the filter, and must be supplied to the lock-in amplifier along with the unknown signal. Thus, ac must excite the thermistor. Studying perfusion changes associated with cerebral blood flow, Wei et al. (40) were able to detect temperature changes of 0.001 °C using lock-in amplifiers, Wheatstone bridges and matched thermistors.

Measurements without Physical Contact

Evans and Hajnayebi (41) developed a temperature acquisition system using proximity telemetry. The system consists of two units, a module that can be worn by the patient containing a temperature-to-pulse width converter (astable oscillator), and a handheld interrogator, which incorporates a radio frequency generator and an automatic-gain controlled (AGC) temperature readout (Fig. 13). Usually in

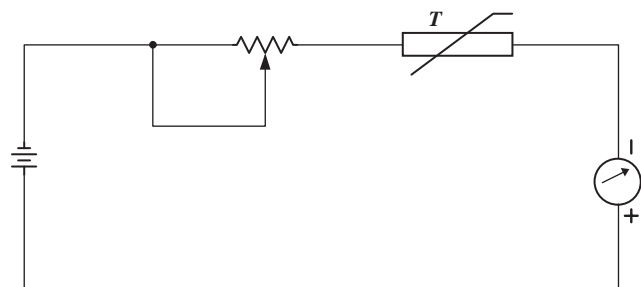


Figure 12. A simple low-cost circuit for temperature measurements based on a microammeter (from *Interfacing Sensors to the IBM PC*, by W. J. Tompkins and J. G. Webster, 1988, Reprinted by permission of Prentice-Hall).

measuring a patient’s temperature, the clinician must wait until the thermometer probe reaches body temperature and then log the readout. In some patients, however, frequent temperature measurement is essential, as, for example, in severe anemic patients undergoing slow blood flow transfusion where it may be necessary to monitor the body temperature each 15 min, for periods of up to 48 h. In such circumstances, the physical disturbance to the patient becomes extremely unpleasant. The system described in Fig. 13 does not require physical contact, allowing for thermistor to be interrogated by a handheld unit held in close proximity to the patient. Compared with conventional electronic clinical thermometers, such design has an additional advantage, since it does not require a separate display module for each patient.

Systems Based on Microcontrollers

The advent of low cost microcontrollers provides the design engineer new design possibilities for medical temperature measurement. Microcontrollers are comprised of a built-in microprocessor, analog-to-digital converter, RAM, and several digital inputs–outputs. The complete system utilizes the microcontroller, multiplexer, EPROM, digital display,

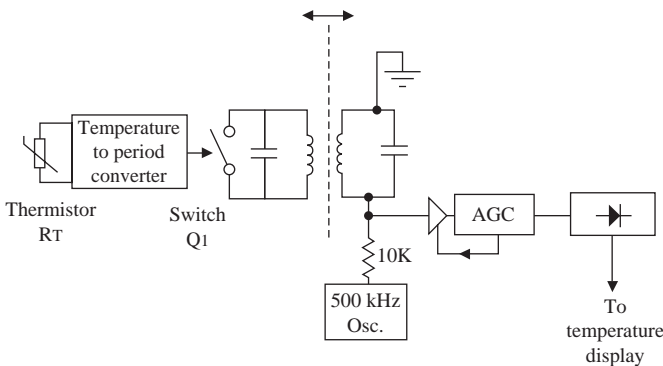
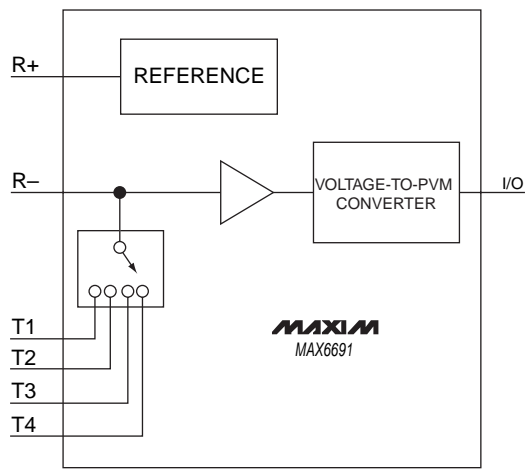
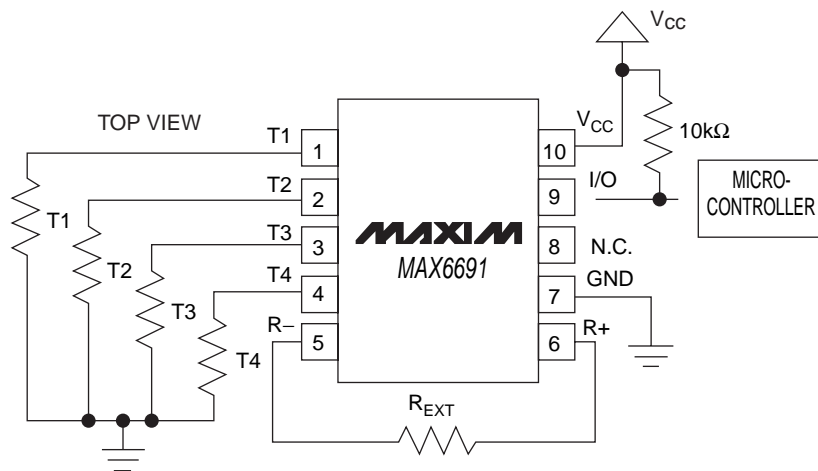


Figure 13. Proximity clinical thermometer hardware. Reprinted from (41), with permission from Elsevier.



(a)



(b)

Figure 14. Functional diagram (A) and typical application circuit (B) of the MAX6691, four-channel temperature-to-pulse-width converter. Copyright Maxim Integrated Products (www.maxim-ic.com). Used by permission.

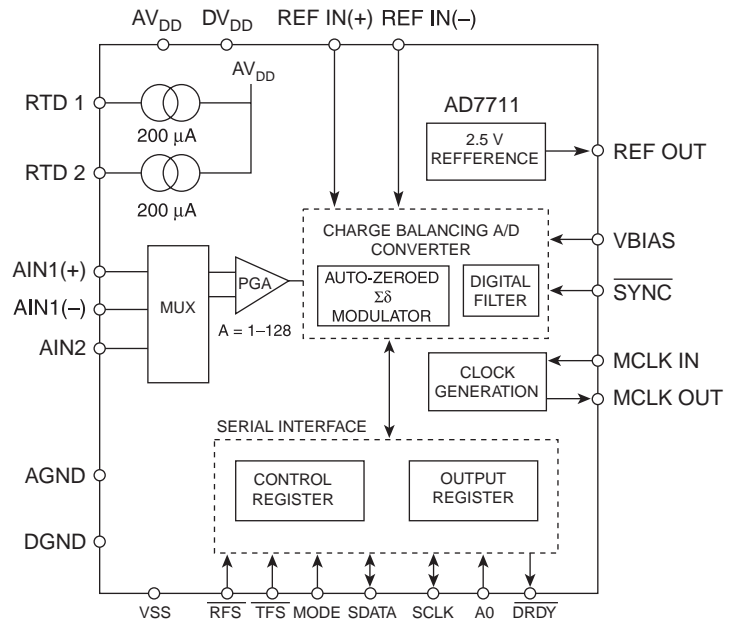
keypad, and display driver, being programmed in assembler language (www.microchip.com; www.motorola.com). These systems are relatively inexpensive to produce, yet offer high temperature accuracy and various software-controlled outputs (39,42). For example, a microcontroller system utilizing thermistor sensors can monitor the temperature in several locations in a patient. After programming, the microcontroller converts the resistance of the thermistor into a temperature reading by using the resistance versus temperature algorithm based on the Steinhart–Hart equation (33) or a look-up table, as discussed before.

Integrated Circuits

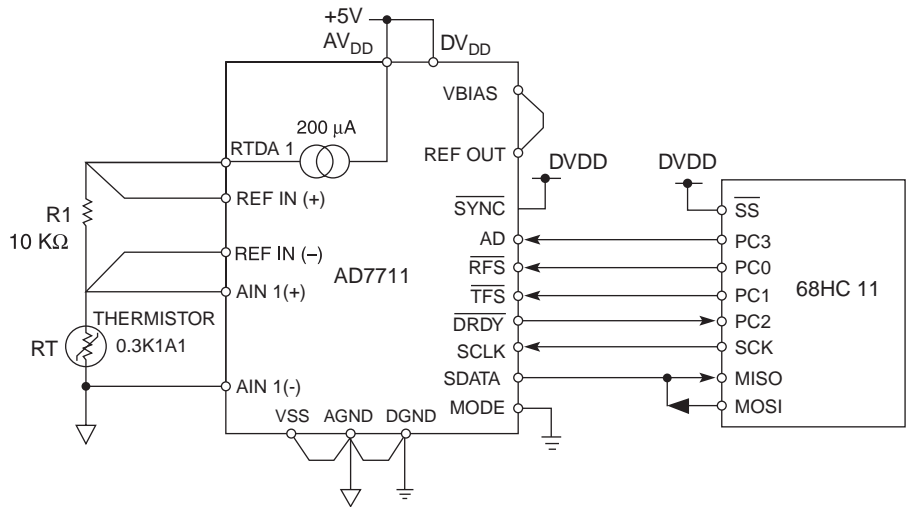
Recently, the electronic industry made available integrated circuits specifically dedicated to use with thermistors (43) and well suited for this task (19). The MAX6691 is a four-channel thermistor temperature-to-pulse-width converter that measures the temperatures of up to four thermistors and converts them to a series of pulses whose widths are related to the thermistors temperatures (43). This device can be readily connected to a variety of microcontrollers with a simple single-wire interface. Operating under the supervision of a microcontroller, the MAX6691 powers the thermistors only

when a measurement is under course. This minimizes the power dissipation in the thermistors, virtually eliminating self-heating. In the intervals between conversions, the MAX6691 falls into a $10\ \mu\text{A}$ (max) sleep mode, where the voltage reference is disabled and the supply current is at its minimum. In handheld healthcare systems, where power is at a premium, it may be an interesting characteristic. These integrated circuits also have internal voltage reference that isolates thermistor from power-supply noise, as described in the functional diagram presented in Fig. 14a.

The AD7711 is an integrated circuit from Analog Devices with signal conditioning and A/D conversion functions that is well suited for temperature measurement applications using thermistors (19). This integrated circuit includes a programmable gain amplifier, current sources and a voltage reference on the chip, allowing thermistors to be excited in either a constant current or voltage mode (Fig. 15a). The reference input is also differential, allowing ratiometric operation on the front end. The component can achieve >16 bits of peak-to-peak resolution and update rates of 100 Hz. Filtering is also provided as part of the $\Sigma\Delta$ process. This on-chip filtering may be useful when transducer excitation frequencies must be removed from the input signal. In addition, the filter profile also provides



(a)



(b)

Figure 15. Functional diagram (A) and typical application circuit (B) of the AD7711, 24-Bit Sigma-Delta, Signal Conditioning ADC. Reproduced from Ref. (19) with permission.

notches with 120 dB attenuation that can be placed at 50 or 60 Hz, minimizing line frequency components from the system. An example of the AD7711 use is presented in Fig. 15b. The on-chip 200 μ A current source acts as the excitation for the thermistor and generates the reference for the converter. In this case, the circuit is fully ratiometric, ensuring that changes in the excitation current will not affect the performance of the circuit. The diagram also shows a serial interface to a 68HC11 microcontroller, which may control the A/D converter, take data readings, and run a linearization algorithm.

Personal Computer Interfacing

Thermistors can be integrated to data acquisition systems and an IBM PC or compatible computer, in order to imple-

ment a flexible system for application is medical thermometry (6,44). These PC-based DAQ systems requires some signal conditioning hardware to interface the thermistor to the measurement device, such as a plug-in DAQ board (Fig. 16). The DAQ board or module performs the analog-to-digital conversion. A system to interface the thermistor and its output signal to this stage should include an excitation current or voltage source, signal amplification, low pass filtering and isolation signal conditioning hardware, which serves to electrically isolate the thermistor sensors from the measurement system, protecting the patient. Concerning the last topic, it is important to point out that electrical-safety codes and standards in health-care facilities must always be strictly followed (45). Several of useful circuits to perform these tasks were discussed earlier. A typical plug-in DAQ board has eight to 16 analog

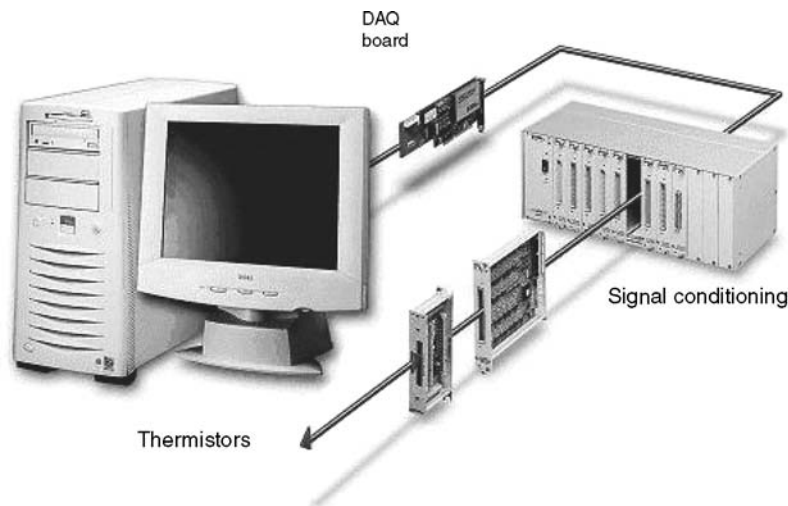


Figure 16. Block diagram of a PC-based DAQ system for temperature measurements with thermistors (courtesy of National Instruments Corporation, Austin Texas, USA).

input channels. External multiplexing systems can be included if it is necessary to expand the number of input channels that can be connected to a DAQ board. The multiplexing system sequentially switches multiple channels to a single-input channel of the DAQ board. Software choices for controlling the data acquisition system include a general purpose programming language (e.g., C, Delphi, or Pascal, under DOS or Windows). Alternatively, the LabVIEW environment can be used to control the system. In this case, special routines for implementing the Steinhart–Hart equation to convert measured voltages from thermistors into temperatures are available.

EXAMPLES OF COMMERCIALY AVAILABLE THERMISTORS AND THERMISTOR PROBES

In the earliest days, thermistors acquired a reputation for being unpredictable and unstable devices, due to problems in manufacture and in application. Nowadays, however, with increased user familiarity with these devices and modern manufacturing technologies, they can be used with a great deal of confidence (15). A large number of companies produce NTC thermistors. Moreover, there is a great variety in the constructional characteristics and parameters of these components, depending on their use. In this section, a survey will be made concerning NTC thermistors for medical use manufactured by some companies. Detailed information can be found in available catalogs of the manufacturers or at the WWW pages listed at the end of this section.

Matched interchangeable thermistors eliminates the need for individual resistance temperature calibration, as well as permits the standardization of circuit components and simplification of design and replacement problems. Honeywell/Fenwall makes devices with interchangeability to within 0.2°C , from 0 to 70°C (Uni-Curve series, Fig. 17a) (10). Highly interchangeable thermistors are also manufactured by Alpha Sensors. These devices present tolerances to $\pm 0.1^{\circ}\text{C}$ over the medical temperature range (21).

The Honeywell/Fenwall LTN (Linear Thermistor Networks, Fig. 17b) series are designed to produce a resistance change or voltage output that varies linearly with tem-

perature (typical maximum linearity deviation of 0.256°C from -30 to 50°C). They consist of one twin thermistor and two precision resistors, and are also available in probe assemblies, providing a sensitivity that is hundred folds greater than that of thermocouples.

The standalone thermistor element is relatively fragile and cannot be placed in a rugged environment. In order to overcome this problem in biomedical applications, thermistor probes, which are thermistor elements embedded in protective tubes, are widely used. For example, Fig. 17c shows a pediatric probe for oral or rectal use (Ysi 423) with the sensor located at the tip of a semiflexible nylon tube. These probes can be disposable (Ysi 4400) or reusable (Ysi 400 ou 700). When disinfections are critical, autoclavable probes are also available for skin, esophageal, or rectal use in adult or pediatric measurements (Ysi 400AC probes, Fig. 17d).

For research or medical applications requiring small size and rapid response, as, for example, long-term subcutaneous measurement, thermal dilution, and flow measurement studies, Ysi manufactures catheter probes with time constant of 0.2 s, electrically isolated and operating in the temperature range of 0 – 70°C . Honeywell/Fenwall manufactures Small Bead Thermistors, that offer ultrafast time response (T.C. 1 s maximum) and are highly sensitive to electric power ($\delta = 0.1$ $\text{mW}\cdot^{\circ}\text{C}^{-1}$ minimum). They are suited for usage in low heat capacity applications and their micro size (0.36 mm outer diameter) makes them adequate for using in extremely small assemblies, (e.g., catheter and hypodermic needles). In addition, they are also adequate for used in self-heat applications. If a standard probe does not fulfill the needs, custom design is available by several manufacturers (please, see WWW pages at the end of this article).

SOME FIELDS OF CLINICAL APPLICATIONS OF NTC THERMISTORS

As pointed out before, thermistors are probably the most widely used transducer for medical temperature measurements. Their characteristics have been contributing to

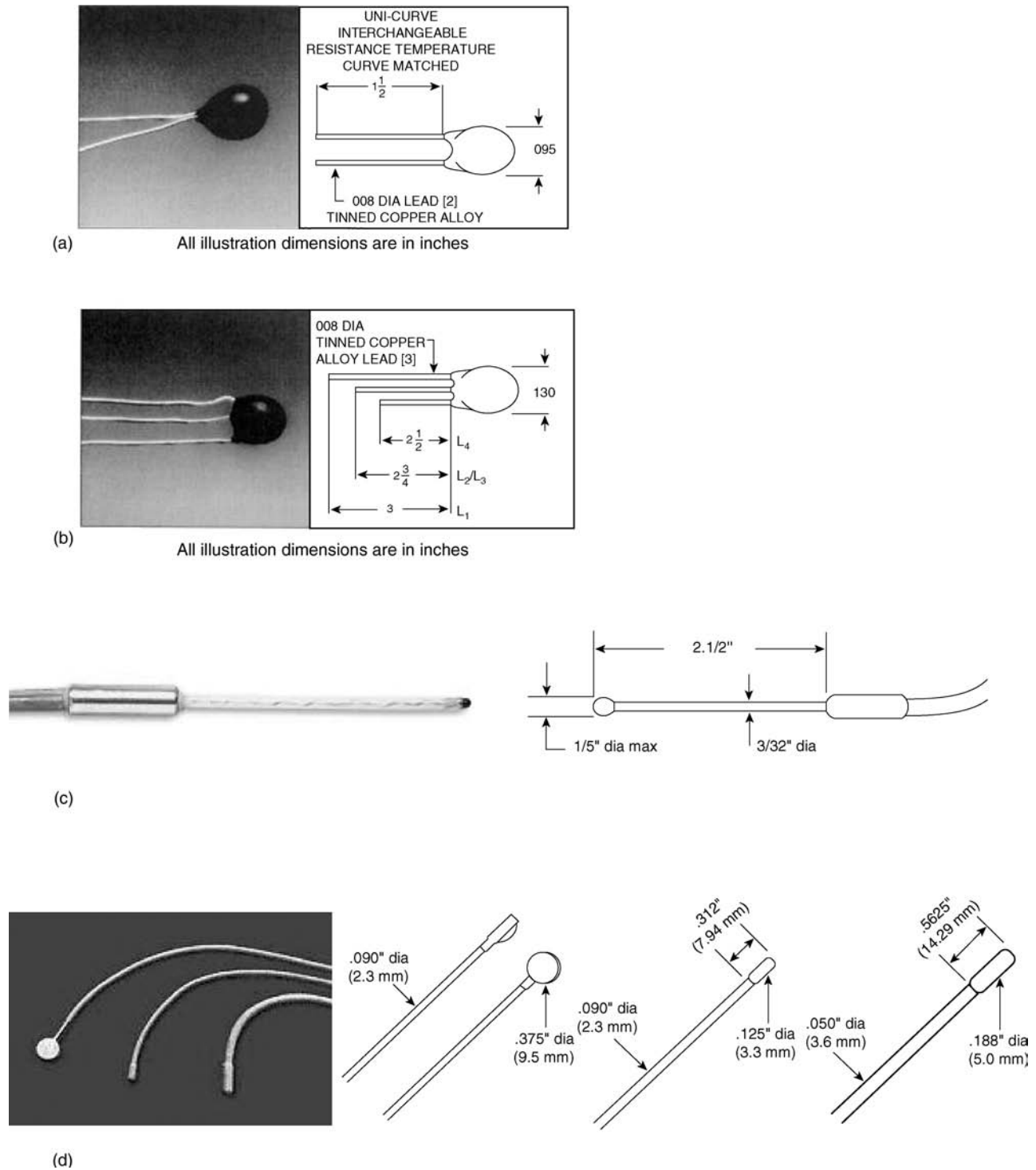


Figure 17. Examples of thermistors commercially available. Highly interchangeability devices (Uni-Curve® series) (A), linear thermistor networks (LTN® series) (B) (courtesy of Honeywell International Inc. 101, Columbia Road Morristown, NJ 07962, USA). Pediatric probe for oral or rectal use (C), and autoclavable probes (D) for skin, esophageal or rectal use in adult or pediatric measurements (courtesy of Ysi, 1700/1725 Brannum Lane, Yellow Springs, Ohio, 45387, USA).

facilitate clinically difficult measurements, that includes testicular temperature measurements in reproductive medicine (46), hypothermia (47), transcutaneous measurements during cardiopulmonary bypass (48), conti-

nuous monitoring of preterm infants (49), and personal heat strain monitoring in occupational medicine (50). However, beyond their utility in temperature measurements, these devices have found widespread use in a variety of

systems for clinical and research applications to measure flow, thermal conductivity and diffusivity of biomaterials, and to detect the presence of liquids. Therefore, this section intends to have a brief discussion about some of these applications, showing the branches of medicine where technology based on thermistor is used, and how this technology can assist each of them.

Cardiovascular Monitoring

Cardiac output, the volume of blood ejected by the heart each minute, is a key parameter in cardiovascular medicine, which is used to obtain diagnostic information about the heart and for continuous monitoring of heart function in critically ill patients. The thermodilution method uses thermistor type catheters to estimate this parameter. The tip of a Swan–Ganz catheter is inserted into a large vein, typically one in the right side of the neck, and advances through the right heart into the pulmonary artery (Fig. 18). Other sites can be used for catheter insertion (e.g., the groin or the arm). A cold saline or a dextrose solution, whose volume and temperature are known, is injected into the blood stream through one of the catheter lumens. The solution mixes with the blood in the right atrium and is diluted as it is carried downstream to a thermistor located at the surface of another catheter lumen. At the thermistor location, the temperature of the blood-injected mixture is measured over a period of time, and then the cardiac output (efficiency) is computed from this temperature–time response data.

At the level of the capillary network, the tissue blood flow (perfusion) is a primary factor in the local transport of heat, drugs, oxygen, nutrients, and waste products. This fundamental parameter holds the key to the diagnosis and subsequent treatment of numerous medical problems. The high sensitivity to small changes in temperature shown by

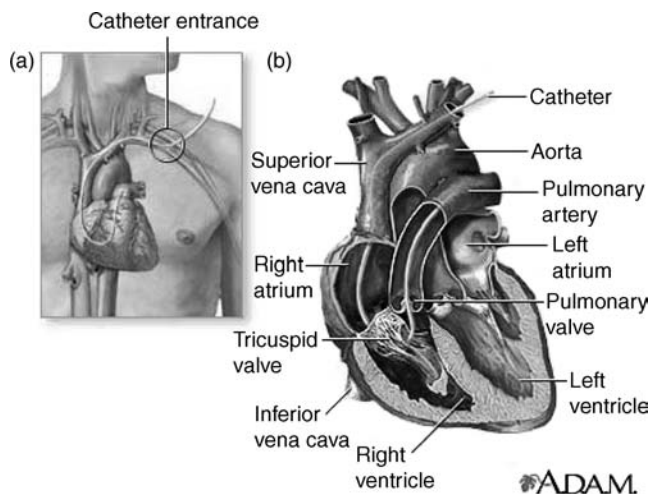


Figure 18. Placement of the thermistor-based catheter in the thermodilution method of estimating cardiac output. Overview (A), and a detail of positioning in the heart (B). This method involves the passage of a catheter into the right side of the heart to obtain diagnostic information about the heart and for continuous monitoring of heart function in critically ill patients. From Medical Encyclopedia (Medline Plus), reprinted by permission.

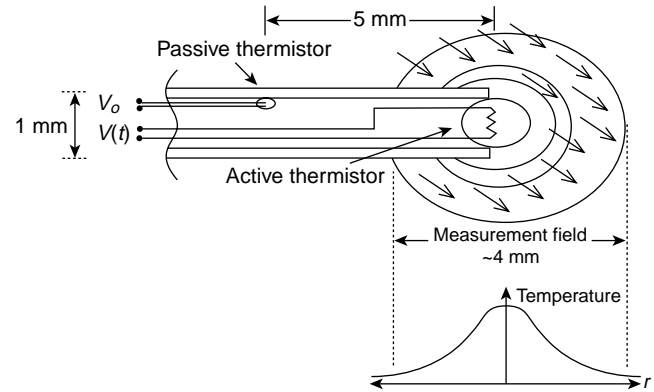


Figure 19. Example of thermal diffusion probe. It can be seen an active, heated thermistor, at the probe tip which produces a thermal measurement field that is dependent of the tissue thermal properties and the perfusion. The passive thermistor, mounted proximal to the probe tip, monitors the baseline temperature variations. Reproduced from (51) with permission of the IFMBE.

a thermistor can be employed to sense the small amounts of heat involved in the thermal techniques developed for the measurement of tissue blood flow (51). Minimally invasive probes constructed around two thermistors are usually applied in these techniques. An example of such probes is presented in Fig. 19. In these probes, one thermistor is used in the self-heat mode, operating as both a heat source and a temperature sensor. Changes in perfusion cause changes in its temperature, which is used as an indirect index of blood flow. As the baseline body temperature fluctuates, a second thermistor with the same size and electrical characteristics is used to measure and compensate for the changes in reference temperature. Systems like these may help clinicians in several application areas, providing early warning for ischemic events, targeting therapy rapidly and accurately, monitoring of patients in organ transplantations, and evaluating tumor and cerebral blood flow (51).

Respiratory Measurements

Obstructive and restrictive respiratory diseases present a huge public health problem. Prevalence rates of asthma and chronic obstructive pulmonary disease are $\sim 4\%$ each. These diseases are characterized by airflow limitation that results from modifications in lung parenchyma and airways. Thermal convection flowmeters measures the local speed of a fluid by measuring the heat loss from a heated element in the flow path (52,53). These instruments are used in respiratory medicine for gas flow analysis, and commonly thermistors are used as the sensing elements. They are also known as thermistor pneumotachometers (16). In these instruments, a small thermistor is placed in the flowing fluid. It operates in the self-heat mode in order to maintain an average temperature above that of the surrounding fluid. This is accomplished using a feedback circuit (16,52,53), as described in Fig. 20. A change in temperature tends to cause a variation in the thermistor resistance, which in turn affects the amplifier output voltage and the current through the sensor. Considering the

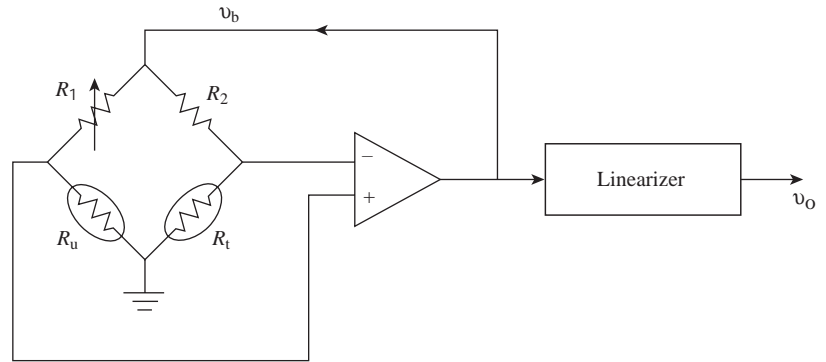


Figure 20. Feedback circuit used in thermal convection flowmeters. From *Medical Instrumentation: Application and Design*, by J. G. Webster. Reprinted by permission of John Wiley and Sons, Inc.

amplifier as presenting infinite constant gain and bandwidth, the bridge is always statically maintained balanced, and hence the thermistor resistance and temperature are constant. The sensor resistance is maintained constant equal to R_1 under any operating condition. This circuit operates satisfactory if ambient temperature is constant. If changes in ambient temperature are expected, a second unheated thermistor can be included in the circuit to compensate it (R_t in Fig. 20). One of the main advantages of this configuration is that the high negative gain feedback divides the sensor time constant by a factor equal to the loop gain, improving the frequency response (52). Thermistors lose heat at a rate dependent on the local mass flow, temperature, specific heat, kinematic viscosity, and thermal conductivity of the fluid. Thus, when a patient exhales through a breathing device in which the thermistor is mounted, it is cooled and the circuit needs to provide more electrical power to keep the thermistor temperature constant. This power is proportional to the airflow. When gas flow properties are sufficiently constant, the output voltage of these circuits is a nonlinear function of mass-flow rate only. Linear mass-flow relationships may be obtained using a linearization stage (Fig. 20), based on analog or digital implementations of piecewise linear or polynomial approximations (53).

Since the thermistor is cooled equally for both directions of velocity, the system with a single sensor provides an output of the same polarity, independent of the flow direction. This feature limits the use of these sensors to unidirectional flow, which can be satisfactory in some applications, as for example, in forced expiratory testes. Directional sensitivity can be provided by putting multiple sensors at separate points along the flow.

Speech Therapy

Spoken language is a highly developed skill. It involves the use of parts of the brain that deal with hearing, understanding speech, sound production, and conversion of the thoughts into speech. Patients suffering from cleft palate or similar defects exhibit a type of speech that is characterized by excessive nasal escape of air and by an abnormal resonance compared with normal speakers. This may be treated by plastic surgery, palatal prostheses, speech therapy, or even a combination of these treatments (54). Trained listeners and also experienced therapists have been already used to detect the presence of nasal air escape. These evaluations are, however, subjective and

hinder objective comparisons with previously made assessments, especially if these were made by another therapist (55). In order to allow a qualitative and quantitative analysis of nasal air escape, anemometers based on thermistors were proposed (54–56). Fig. 21 is a cross-section showing details of the device. The thermistor is positioned in the longitudinal path of the nasal airflow. The system uses a Wheatstone bridge; in one arm of the bridge a thermistor is connected, while in the opposite arm a second thermistor is used, as temperature-compensation element. In the practical use of this kind of system, clear differences were obtained when nonnasal and nasal subjects were asked to talk a nonnasal word (cheese) and a nasal word (missing) (54). The system can provide a numerical figure of merit to indicate the extent of the defect and the effectiveness of a treatment, as, for example, palatal training.

Sleep Medicine

One of the most used methods to sense breathing patterns is to detect airflow using a nasal thermistor sensor. The principle here is that of exhaled air ($\sim 37^\circ\text{C}$) being slightly warmer than inhaled air (room temperature), resulting in modifications in the thermistor resistance correlated with the respiration rate. The small size of the thermistor contributes to prolonged, minimally intrusive measurements of breathing pattern, which are particularly important for respiratory surveillance in newborn intensive care, biofeedback studies, and circadian rhythm analysis (57).

These characteristics also made thermistors a traditional device in the diagnostic of sleep-disordered breathing (SDB) (58), which is a widespread disease estimated to be present in $\sim 2\text{--}4\%$ of middle-aged adults. In these patients, the pharyngeal airway narrows with sleep onset, initially producing harsh respiratory breathing with some evidence of inspiratory flow limitation. Then, with further narrowing snoring is generated, and finally there is a complete collapse producing a full obstructive sleep apnea (OSA). Even if upper-airway obstruction is incomplete (hypopnea), increased upper airway resistance will still cause clinical symptoms similar to OSA because of respiratory effort-related arousals. Although rare compared to obstructive events, sleep apnea episodes can also be present in the absence of upper airway obstruction. In this case, known as central sleep apnea (CSA), the apnea events result from a decreasing in central controller output to the inspiratory pump muscles. The classic daytime manifestation of SDB is the excessive sleepiness, although other

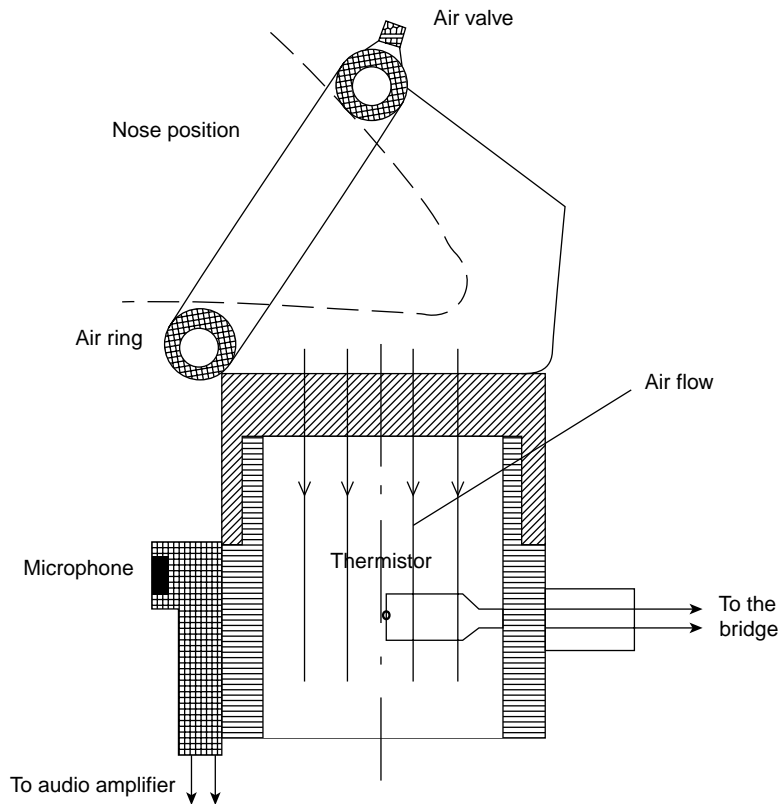


Figure 21. Cross-section showing a thermistor in a mask used by speech therapists to evaluate nasal air escape (Reproduced from (54) with permission of the IFMBE).

symptoms, such as unrefreshing sleep, poor concentration and fatigue are commonly reported. Automobile and industrial accidents have been associated with a poor quality of sleep as well as hypertension and intellectual deficits. The severity of the disease is defined in terms of the apnea/hypopnea index (AHI). This measurement reflects the average number of apneas plus hypopneas observed per hour of sleep, and it is usually derived by identifying and counting each respiratory disturbance, with subsequent division of the sum by the number of hours slept. This way, scientific and diagnostic studies of sleep-disordered breathing are critically dependent of the performance of the measuring device used to detect abnormal respiratory events. Thermistors usually used in SDB diagnosis presents a long time constant (58). This way, they are not able to provide a good characterization of fast events like hypopneas. The effect of the thermistor high measurement time constant can be seen in Fig. 22, in which a simultaneous acquisition of the thermistor and a fast-responding system were done in a awake subject during spontaneous breathing. Fig. 22 shows that ventilatory details were lost by the thermistor system. This loss can be explained by the lowpass filter action of the thermistor, in which the higher order harmonics of the respiration process are discarded. This characteristic limits the accuracy of the clinical diagnosis based in thermistor in sleep studies, because it may introduce underestimation of hypopneas events. It happens because, during hypopneas, the respiratory flow tends to change from a quasisinusoidal (normal breathing) to an almost square-wave pattern. Similar limitation to describe dynamic respiratory events was also observed for thermistors used to monitor air stream temperatures in

exercising asthmatic patients (59). Recently, Togawa et al. (60) presented a technique that improved the accuracy of fluctuating temperature measurement by thermistors. This technique seems to present a great potential to improve the performance of these devices in the monitoring of fast events, a field still open to research.

Evaluation of the Thermal Properties of Tissue

The evaluation of thermophysical properties of tissue assume importance with the increasing use of hyperther-

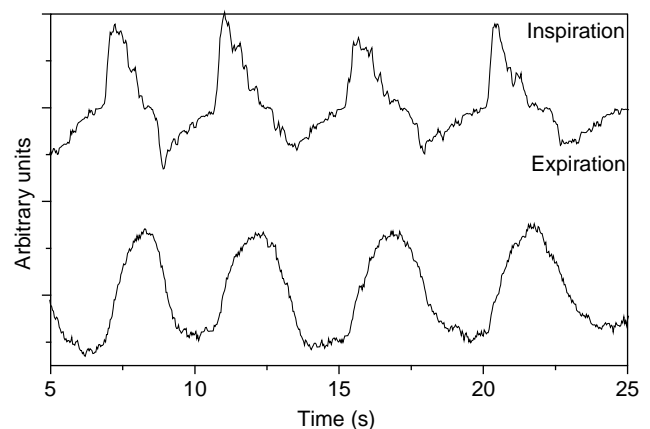


Figure 22. Typical signal morphology obtained in studies with an awake subject during spontaneous breathing by the thermistor (bottom trace) and NPR systems (top trace). Note that respiratory details were lost by the thermistor signal (Reproduced from (58) with permission—© 2004, American Institute of Physics).

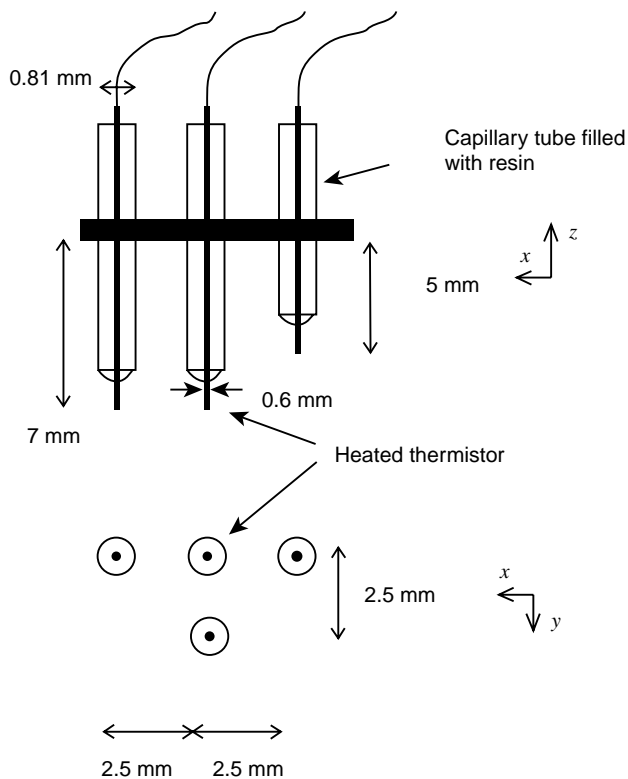


Figure 23. Thermistor thermal probe for the evaluation of thermal conductivity and the thermal diffusivity of biomaterials (Reproduced from (61) with permission—© 2001 IEEE).

mia and therapeutic procedures based on heat delivery, such as radiofrequency, microwave, laser, and ultrasound (61). The thermal conductivity and the thermal diffusivity of biomaterials are measured using thermal probe techniques. Various thermal diffusion probe techniques have been developed from Chato's first practical use of the thermal probe (62). Physically, for all of these techniques, heat is introduced to the tissue at a specific location and is dissipated by conduction through the tissue and by convection with the blood perfusion. The similarity of these techniques is the use of a thermistor bead, either as a heat source or a temperature sensor. The probes are usually constructed by placing a miniature thermistor at the tip of capillary tubes. Fig. 23 shows one example of probe assembly (61). Four probes were used with one heated thermistor and three sensing thermistors in three different locations. The distribution of three sensing thermistors orthogonal to the central heating thermistor permits the determination of the directional nature of thermal properties. The probe is positioned invasively within the tissue of interest. Electrical power is delivered to the heating thermistor and the resulting temperature rise is measured by the sensing thermistors. An empirical relationship between the power delivered by the first thermistor and the temperature raise recorded by the sensing thermistors is used to evaluate the thermal conductivity along the line joining each of the thermistors. The delay between the application of a power pulse in the heated thermistor and the temperature pulse in the sensing thermistors is calculated by cross-correla-

tion and used to determine the diffusivity of the material in each direction. This system allowed the authors to measure the thermal properties of a swine left ventricle *in vivo*, as well as to investigate the effect of ablation. The thermal conductivity and the diffusivity of the tissue dropped after ablation. These data will help to generate three-dimensional (3D) thermal models of the ablation process.

Geriatrics

In the geriatric setting, the urinary incontinence is present in 7–15% of the elderly. The prevalence in elderly living in chronic care hospitals and nursing homes is considerably higher, up to 62% (63). The high sensibility and small size of the thermistor contribute to its application in the management of urinary incontinence, allowing for the development of monitors that can help to reduce elderly reliance on incontinence pads, both improving their quality of life, and reducing the cost of the treatment (64). These monitors are based on the changes in temperature produced by the urinary incontinence event, being developed for long-term ambulatory monitoring (63,64) and for home healthcare (65).

BIBLIOGRAPHY

1. Guyton AC, Hall JE. Textbook of Medical Physiology. 10th ed., Philadelphia: Saunders; 2002.
2. Buono MJ, Ulrich RL. Comparison of mean skin temperature using "covered" and "uncovered" contact thermistors. *Phys Meas* 1998;19:297–300.
3. Lavenuta G. Negative temperature coefficient thermistors. *SENSORS*, May 1997, p. 46.
4. Wang CC, Akbar SA, Madou MJ. Ceramic based resistive sensors. *J Electroceramics* 1998;2(4):273–282.
5. Nenov TG, Yordanov SP. *Ceramic Sensors: Technology and Applications*. Pennsylvania: Technomic; 1996.
6. Lin CH, Jadvar H. Interfacing temperature sensors. In: Tompkins WJ, Webster JG, editors. Chap. 7. *Interfacing Sensors to the IBM PC*. Englewood Cliffs, (NJ): Prentice-Hall; 1988. pp. 183–224.
7. Anonymous (1998). Technical Notes, [Online]. Alpha Sensors, Inc. Available at <http://www.alphasensors.com/technical.html>. Accessed 2005, March 11.
8. Anonymous (2002). General technical information [Online]. Epcos-Electronic Parts and Components. Available at <http://www.physics.leidenuniv.nl/edu/courses/Experimentele%20Natuurkunde/NTC.pdf>. Accessed 2005, March 11.
9. Anonymous (2001). YSI Precision Medical Probes and Accessories Catalog, [Online]. Ysi Corporation. Available at <http://http://www.advindsys.com/ysi.htm>. Accessed [2005, March 11].
10. Anonymous (No date). Thermistors catalog, [Online]. Honeywell Corporation (Fenwall). Available at <http://content.honeywell.com/sensing/hss/thermal/product/thermistors.asp>. Accessed 2005, March 17.
11. Anonymous (No date). Thermistor glossary, [Online]. RTI Electronics Inc. Available at <http://http://www.rtie.com/ntc/glossary.htm>. Accessed 2005, March 11.
12. Anonymous (2004, December, 14). Application Notes, [Online]. Ametherm Inc. Available at <http://www.ametherm.com/Applications/>. Accessed 2005, March 11.

13. Anonymous (No date). NTC Thermistors, [Online]. Thermometrics Inc. Available at <http://www.thermometrics.com/assets/images/ntcnotes.pdf>. Accessed 2005, March 11.
14. Keskin AU, Yanar TM. Steady-state solution of loaded thermistor problems using an electrical equivalent circuit model. *Meas Sci Technol* 2004;15:2163–2169.
15. Sheingold DH. *Transducer Interfacing Handbook*. Norwood (MA): Analog Devices; 1981.
16. Normann RA, *Principles of Biomedical Instrumentation*. New York: John Wiley & Sons; 1988.
17. Anonymous (2002). Application Notes [Online]. Epcos—Electronic Parts and Components. Available at <http://www.epcos.com/inf/50/db/ntc02/00290045.pdf>. Accessed 2005, March 11.
18. Beakly WR. The design of thermistor thermometers with linear calibration. *J Sci Instrum*. 1951;28:176.
19. O'Grady A. Building a more perfect union: combining thermistors and high-resolution $\Sigma\Delta$ A/D converters, *Sensors Online* Available at www.sensorsmag.com/articles/0100/42/main.shtml, Accessed 2000.
20. Lyons P, Waterworth P. The use of NTC thermistors as sensing devices for TEC controllers and temperature control integrated circuits, technical report, Betatherm Ireland Ltd.
21. Anonymous (1998). Application Notes, [Online]. Alpha Sensors, Inc. Available at <http://www.alphasensors.com/tappnotes.html>. Accessed 2005, March 11.
22. Hoge HJ. Comparison of circuits for linearizing the temperature indications of thermistors. *Rev Sci Instrum* 1979;50(3).
23. Sheingold DH. *Nonlinear Circuits Handbook*. Norwood, (MA): Analog Devices; 1976.
24. Sundqvist B. Simple, wide-range linear temperature-to-frequency converters using standard thermistors. *J Phys E Sci Instrum* 16:261–264.
25. Patranabis D, Ghosh S, Bakshi C. Linearizing transducer characteristics. *IEEE Trans Instrum Meas*. 1988;37:66–69.
26. Sengupta RN. A widely linear temperature to frequency converter using a thermistor in a pulse generator. *IEEE Trans Instrum Meas* 1988;37:62–65.
27. Slomovitz D, Joskowicz J. Error evaluation of thermistor linearizing circuits. *Meas Sci Technol* 1990;1:1280–1284.
28. Kaliyugavaradan S, Sankaran P, Murti VGK. A new compensation scheme for thermistor and its implementation for response linearization over a wide temperature range. *IEEE Trans Instrum Meas* 1993;42(5):952–956.
29. Kaliyugavaradan S, Sankaran P, Murti VGK. Hardware linearization of thermistor response using series-parallel resistors for temperature-to-time conversion. *Meas Sci Technol* 1994;5:786–788.
30. Woodwards S. Pseudologarithmic thermistor signal conditioning spans wide temperature range, *EDN* 2005;50(1):66–67.
31. Ghosh D, Patranabis D. Software linearization of thermistor type nonlinearity. *IEE Proc Circuits Devices systems* 1992;139(3):339–342.
32. Finnie B. Software linearization techniques for thermistors. *Electronic Design* Available at www.elecdesign.com/Articles/Index.cfm?AD=1&ArticleID=6338. Accessed 1998.
33. Day J, Bible S. (2004). Piecewise linear interpolation on PIC 12/14/16 series microcontrollers, application note no. 942. [Online]. Microchip Technology Inc. Available at <http://ww1.microchip.com/downloads/en/AppNotes/00942A.pdf>. Accessed 2005, March 11.
34. Hagerman J. Model thermistors with spice. *Elect Design* 1991;39(5):85.
35. Wangenheim L. SPICE subcircuit models thermistors. *EDN* July 1997.
36. Boylestad RL, Nashelsky L. *Electronic Devices and Circuit Theory*. Englewood Cliffs (NJ): Prentice-Hall; 2001.
37. Bishop J. Thermistor temperature transducer-to-ADC application. *Analog Appl J* (Texas Instr Inc). 2000; 44–47.
38. Joyce D. (No date). Practical aspects of thermistors, Technical note no. 124. [Online]. Zworld Inc. Available at <http://www.zworld.com/support/techNoteswhitePapers.shtml>. Accessed 2005, March 11.
39. Anonymous (March, 2004). A high resolution/precision thermometer using ST7 and NE555, [Online]. ST Microelectronics, Application note no. AN1755. Available at <http://mcu.st.com/familiesdocs-15.html>. Accessed 2005, March 11.
40. Wei D, Saidel GM, Jones SC. Thermal method for continuous measurement of cerebral perfusion. *Med Biol Eng Comput* 1994;32(5):481–488.
41. Evans NE, Hajnayebi HR. Clinical temperature acquisition using proximity telemetry. *J Biom Eng* 1991;13:83–86.
42. Baker B. Thermistors in single supply temperature sensing circuits, AN685, Microchip Technology Inc., 1999.
43. Anonymous (2002). Four-channel termistor temperature-to-pulse-width converter, datasheet, [Online]. Maxim Integrated Products. Available at <http://www.maxim-ic.com>. Accessed 2005, March 11.
44. Potter D. (June, 1995). Measuring temperature with thermistors – a tutorial, [Online]. National Instruments Inc., Application note no. 65. Available at <http://www.seas.upenn.edu/courses/belab/ReferenceFiles/Thermistors/an065.pdf>. Accessed 2005, March 11.
45. Olson WH. Electrical safety. Webster JG, editor. *Medical Instrumentation: Application and Design*. 3rd edition. New York: John Wiley & Sons; 1998.
46. Harrison RM, Smith SD, Roberts JA. Testicular temperatures measured by thermistor probe and contact thermography. *Fertility Sterility* 1990;54(1):173–174.
47. Thoresen M, Whitelaw A. Cardiovascular changes during mild therapeutic hypothermia and rewarming in infants with hypoxic-ischemic encephalopathy. *Pediatrics* 2000;106:92–99.
48. Sakuragi T, Mukai M, Dan K. Deep body temperature during cardiopulmonary bypass. *Br J Anaesth* 1993;71(4):583–585.
49. Dollberg S, Rimon A, Atherton HD, Hoat SB. Continuous measurement of core body temperature in preterm infants. *Am J Perinatol* 2000;17(5):257–264.
50. Muir IH, Bishop PA, Lomax RG, Green JM. Prediction of rectal temperature from ear channel temperature. *Ergonomics* 2001;44(11):962–972.
51. Martin GT, Bowman HF. Validation of real-time continuous perfusion measurements. *Med Biol Eng Comput* 2000;38:319–325.
52. Webster JG. Measurement of flow and volume of blood. Webster JG, editor. *Medical Instrumentation: Application and Design*. 3rd edition. New York: John Wiley & Sons; 1998.
53. Primiano Jr. FP. Measurement of the respiratory system. Webster JG, editor. *Medical Instrumentation: Application and Design*. 3rd edition. New York: John Wiley & Sons; 1998.
54. Besar SS, Kelly SW, Greenhalgh PA. Nasal airflow measurement using a compensated thermistor anemometer, Part 1 system description and quantitative analysis. *Med Biol Eng Comput* 1989;27:628–631.
55. Besar SS, Kelly SW, Greenhalgh PA. Nasal airflow measurement using a compensated thermistor anemometer, Part 2 Computer signal processing and quantitative analysis. *Med Biol Eng Comput* 1990;28:127–132.
56. Mirlohi HR, Kelly SW, Manley MCG. New technique for assessment of velopharyngeal function. *Med Biol Eng Comput* 1994;32:562–566.
57. Jovanov E, Raskovic D, Hormigo R. Thermistor-based Sensor for circadian rhythm evaluation. *Biomed Sci Instr* 2001;37:493–497.

58. Mesquita Jr JA, Melo PL. A respiratory monitoring system based on pressure measurement for the analysis of sleep breathing disorders: Static and dynamic errors reduction and comparisons with thermistors and pneumotachographs. *Rev Sci Instrum* 2004;75(3):760–767.
59. Clary AL, Fouke JM. Fast-responding automated airway temperature probes. *Med Biol Eng Comput* 1991;29:501–504.
60. Tagawa M, Kato K, Ohta Y. Response compensation of thermistors: Frequency response and identification of thermal time constant. *Rev Sci Instrum* 2003;74(3):1350–1358.
61. Naresh CB, et al. Measurement of directional thermal properties of biomaterials. *IEEE Trans Biomed Eng* 2001; 48(2):261–267.
62. Chato JC. A method for the measurement of thermal properties of biological materials. Proceedings of the Symposium on Thermal Problems in Biotechnology New York, paper LCN068-58741, 16–25, 1968.
63. Hurk PRB, et al. Long-term ambulatory monitoring of urine leakage in the elderly: An evolution of the validity and clinical application of the thermistor signaling. *J Med Eng Technol* 1998;22(2):91–93.
64. Cusick G, et al. A system for logging incontinence events using a simple sensor. Proceedings of The Institution of Mechanical Engineers part H. *J Eng Med* 2003;217(H4): 30–310.
65. Tamura T, et al. A warning detector for urinary incontinence for home health care. *Biom Instrum Technol* 1995;29:343–349.

Further Reading

- Books: The following books contain several useful information concerning thermistors, its use and instrumentation. Tompkins WJ, Webster JG. Eds. *Design of Microcomputer-Based Medical Instrumentation*. New Jersey: Prentice Hall; 1981.
- Enderle J, Blanchard S, Bronzino J. *Introduction to Biomedical Engineering*. London: Academic Press; 2000.
- Doebelin EO. *Measurement Systems: Application and Design*. New York: McGraw-Hill; 1990.
- Cobbold RSC. *Transducers for Biomedical Measurements: Principles and Applications*. New York: John Wiley & Sons; 1974.
- Papers: Thermistor general characteristics and technology: Keskin AU. A simple analog behavioral model for NTC thermistors including self-heating effect. *Sensors and Actuators-A* 2005; 118-2:244–247.
- Veijola T. Electrothermal simulation models for NTC and PTC thermistors. *Proceedings of CSC 98 1998*;2, Piraeus: 950–955.
- Broughton MB. Analysis and design of almost-linear one thermistor temperature transducers. *IEEE Trans Instrum Meas* 1974. 23(1):1–5.
- Jung JS, et al. Reliability evaluation and failure analysis for NTC thermistor. *Int J Mod Phys B* 2003;17(8–9):1254–1260.
- Pathan ZB, Shaligram AD. A PC-based characterization set-up for reliable parameter testing of thermistor. *Indian J Pure Appl Phys* 1995;33(4):215–219.

Clinical Applications of NTC Thermistors

- Farré R, et al. Accuracy of thermistors and thermocouples as flow-measuring devices for detecting hypopneas. *Eur Resp J* 1998;11:179–182.
- Bhavaraju NC, et al. Measurement of directional thermal properties of biomaterials. *IEEE Trans Biom Eng* 2001;48(2):261–267.

- dos Santos I, et al. Measurement of ejection fraction with standard thermodilution catheters. *Med Eng Phys* 2002;24(5):325–335.
- Cui R, et al. A needle temperature microsensor for in vivo and real-time measurement of the temperature in acupoints. *Sensors Actuators A: Phys* 2005;119(1):128–132.

Technical Reports

- Doug C. Implementing Ohmmeter/Temperature Sensor. AN512, Microchip Technology Inc.
- Rodger R. Resistance and Capacitance Meter Using a PIC16C622. AN611, Microchip Technology Inc.
- Joyce D. (No date). Practical aspects of thermistors, Technical note no. 124. [Online]. Zworld Inc. Available at <http://www.zworld.com/support/techNoteswhitePapers.shtml>. Accessed 2005, March 11.

WWW Pages : Manufactures of Thermistors and Associated Products

- Honeywell/Fenwall: <http://content.honeywell.com/sensing/hss/thermal/>.
- Advanced Thermal Products: <http://www.atpsensor.com/>.
- Alpha Sensors: <http://www.alphatechnicsonline.com/>.
- Betatherm: <http://www.betatherm.com/products/index.php>.
- Precision Engineering: <http://www.pel-ltd.co.uk/english/spec.htm>.
- Thermometrics: <http://www.thermometrics.com/>.
- Yellow Springs Instruments Inc.: <http://www.ysi.com/temperature.htm>.
- Omega Engineering: <http://www.omega.com>.

See also HYPERTHERMIA, SYSTEMIC; TEMPERATURE MONITORING.

THERMOCOUPLES

JOAKIM WREN
DAN LOYD
Linköping University
Linköping, Sweden

INTRODUCTION

Accurate temperature measurement is of utmost importance in many medical, biological, and biomedical applications. Technological development has increased both the performance and the range of equipment for temperature measurement, but at the same time the appetite for better measurements has increased as well. Opposite to what might be expected, this in fact increases the demands on the choice and especially the use of equipment for temperature measurement, as aspects associated with the particular application becomes relatively more important.

Despite the large variety of sensors and associated instrumentation, there is a very limited amount of principles or “physics” behind the different sensor types, at least when the most widely used sensors are considered (1); thermocouples, thermistors, and resistance based sensors (e.g., Pt-100). A thermocouple is an active sensor meaning that by itself it can give rise to a detectable signal. Thermistors and resistance-based sensors are passive as they demand an outer-power source. What all these sensors

have in common is that they only measure their own temperature. Thus, they interact thermally with the object to be measured, which can give significant measurement errors.

In order to carry out measurements with sufficient accuracy at a reasonable cost, several demands are put on the equipment and its use. All sensors and measurement systems have their benefits and limitations, and this should be taken into account together with the aspects of the present measurement situation when the measurement equipment is chosen. This process often involves analysis of accuracy, stability, time dependence, and environmental factors, for example, exposure to electromagnetic (EM) and thermal disturbances.

THEORY

Temperature Measurement and the Laws of Thermodynamics

Temperature measurement is based on the zeroth and first laws of thermodynamics. The zeroth law, as the name implies, was stated after the first and second laws, but it was considered to be of such importance that it was denoted the zeroth law. It states that if two bodies are in thermal equilibrium with a third body, they are also in thermal equilibrium with each other. For a further thermodynamic discussion see Ref. 2. From a temperature measurement perspective, it is interesting to replace the third body by a thermometer, for which the zeroth law can be restated as two bodies are in thermal equilibrium if they have the same temperature even if they are not in contact. The law thus tells us what we know from intuition, namely, that systems of different temperatures and in thermal contact with each others strive to equalize their temperatures. This is a very fundamental statement, although in many situations we must consider other aspects as, for example, heat transfer inside the considered systems and between the system(s) and the(ir) surroundings.

The above discussion emphasizes the importance of the first law of thermodynamics, which states that energy is conserved. In reality, all bodies are exposed to a thermal environment at another and often varying temperature that gives a time-dependent heat flux between the bodies and their environment. The difference and/or the variation in temperature and corresponding heat flux may sometimes be small; often, however, they are significant and can be considered as one of the main sources of measurement errors, which will be discussed in the next section. The principle of conservation of energy does not only cause trouble in the form of measurement errors, it can also be beneficial to use in the frequently used “lumped heat-capacity method” (see, e.g., Ref. 3).

Thermocouples

Thermocouple temperature measurement is based on the electrothermal phenomena known as the Seebeck effect after its discoverer T.J. Seebeck in 1821. If two dissimilar metals are connected in a circuit, as in Fig. 1, a temperature dependent electromotive force (EMF) arises in the circuit. The magnitude of the EMF is dependent on the materials and the temperature at the junctions.

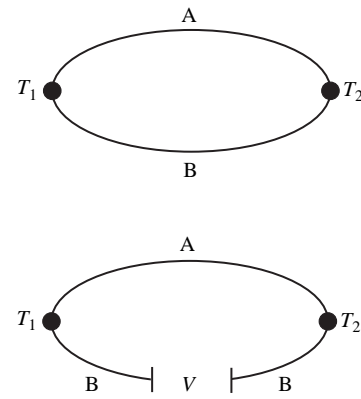


Figure 1. (a) Closed and (b) open thermocouple circuits with conductors A and B and junction temperatures T_1 and T_2 .

Since the junctions are electrically connected and subject to different electrical potentials, a current will flow in the circuit; thus, thermal energy is converted into electric energy (see, e.g., Ref. 4). The current flow makes the Seebeck effect related to the Peltier and Thomson effects (see, e.g., Ref. 5). Peltier discovered in 1834 that if a current flows in a circuit made of two dissimilar conductors as in Fig. 1, one of the junctions becomes cold and the other junction hot. Thomson later found that for a circuit subject to a temperature gradient (heat flux), heat is rejected in any point where current and heat flows in the same direction, and is absorbed where the flows are countercurrent.

If a voltmeter is inserted in the circuit, as in Fig. 2b, it is possible to measure the EMF and relate it to the junction temperatures, as the voltage is dependent on the difference between temperatures T_1 and T_2 . Temperature measurement using a thermocouple is in fact measurement of a temperature difference between the two junctions.

It is important that the voltmeter has a sufficiently high resistance to keep the current small in order to eliminate disturbances related to the Peltier and Thomson effects (see, e.g., Ref. 5). This is normally not a problem, as modern voltmeters have sufficiently high impedance for these effects to be completely negligible.

Before modern measurement systems became available, the thermocouple was normally used by keeping one of the junctions (the reference junction) at a known temperature; typically an insulated ice bath that is very close to 0°C , whereas the other junction (the sensor) was used for temperature measurement. The measured voltage is proportional to the temperature difference between the junctions,

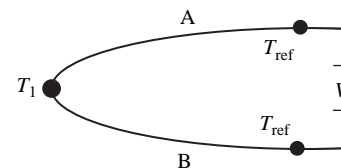


Figure 2. The simplest form of a thermocouple measurement circuit. The sensor measures temperature T_1 , whereas the reference junction temperature T_{ref} is integrated in the measurement system.

and since the ice bath temperature is known, the temperature of the other junction can be determined if the thermoelectric properties of the wires are known.

Today, the reference junction is more or less always integrated in the measurement system. The determination of the reference temperature is normally optimized for the instrument at room temperature (20–25 °C), although temperatures close to this interval can often be used accurately for most instruments. The circuit corresponding to a standard connection is given by Fig. 2, where the reference temperature is measured at the transition between the thermocouple leads (or extension leads) and the instrument connection. For two conductors A and B, the relation between the sensor temperature T_1 , the reference temperature T_{ref} and the measured EMF (voltage, V) follows the equation

$$V = \int_{T_{ref}}^{T_1} S_A(T)dT - \int_{T_{ref}}^{T_1} S_B(T)dT = \int_{T_{ref}}^{T_1} S_{AB}(T)dT \quad (1)$$

where $S_A(T)$, $S_B(T)$ and $S_{AB}(T)$ are the Seebeck coefficients (also called the thermoelectric power–sensitivity) for the respective conductors A and B, and the thermocouple AB. The EMF corresponding to more complex thermoelectric circuits is discussed in depth in Ref. 6.

If a commercial measurement system is not available or a designed circuit for inclusion in a larger system or stand alone is preferred, today there are special-purpose integrated circuits with built-in cold junction compensation (see, e.g., the example in Fig. 3).

Thermocouple Types

All metallic conductor pairs give rise to an electric potential, and thus have the potential to be used as a thermocouple. It is, however, convenient to use one of the standardized thermocouple types that has emerged, which offers obvious advantages, such as compatibility with commercial instruments, predetermined Seebeck coefficients, and a potential–temperature relationship and International Electrotechnical Commission (IEC)-specified tolerance classes.

There are about a dozen types of thermocouples that complement each other in terms of the measuring signal,

temperature range, and tolerance to different environments. Type K is the most commonly used and it is a good compromise of price and performance for many situations, although other types might be the choice for very high/low temperatures, hostile environments, and so on. Perhaps the most important difference between the various types is the output signal determined by the relative Seebeck coefficient of each material. Types S, B, and R contain platinum and are therefore more expensive than the others, but they are also more stable in, for example, e.g., oxidative environments.

There are organizations/national organizations that provide standards for thermocouples, but the standards of the IEC are internationally recognized and should be followed whenever possible. The IEC 584-1, last revised in 1995, contains reference tables and calculation polynomials for the output signals of the standardized thermocouple types as a function of temperature.

THERMOCOUPLE DESIGN

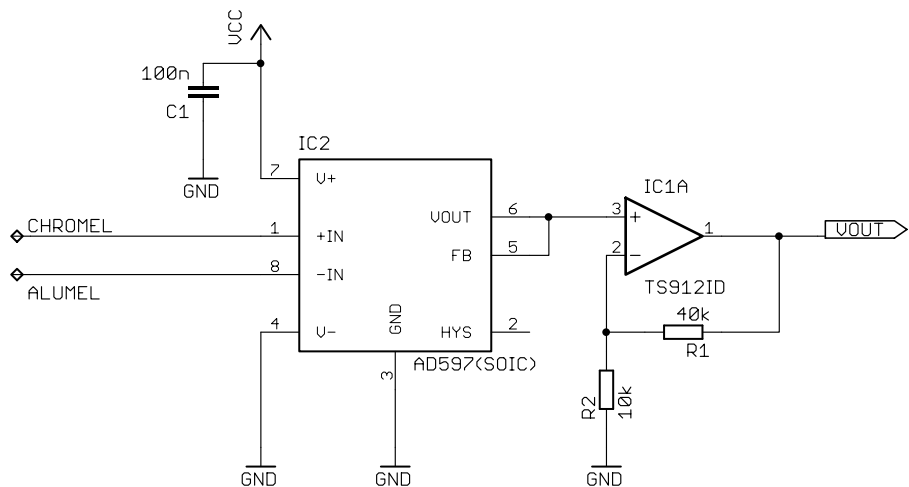
Wire Sensors

In its simplest form, a thermocouple element consists of a pair of wires that are connected together at a measuring junction. The junction must be electrically conducting and can be formed by soldering, crimping, or twisting the wires together, depending on what the situation requires. Wire sensors are mainly used to measure low temperatures in favorable environments. One limitation is imposed by the insulation material [e.g., poly(vinyl chloride) (PVC), nylon, Kapton or poly(tetra fluoroethylene) (PTFE)]. Another limitation is that the measuring junction is exposed to the atmosphere. The PVC insulation can withstand temperatures up to ~ 100 °C, whereas certain ceramic materials are tolerant of temperatures up to 1000 °C or more.

Sheathed Thermocouples

In sheathed thermocouples, the wires are normally insulated by densely packed magnesium oxide enclosed in a metal alloy suitable for the given thermocouple element.

Figure 3. An example of a circuit for temperature measurement using a thermocouple type K. The thermocouple controller AD597 (Analog Devices, Norwood, MA) contains a cold junction compensation circuit, and the OP-amplifier TS912ID (ST Microelectronics, Geneva, Switzerland) amplifies the output “V OUT” so that 0 V equals 0 °C and 5 V equals 100 °C with the present impedances. The thermocouple is connected to the “+IN” and “–IN” connections, and the low pass filtered voltage supply is connected at “V+”. There are several manufacturers of similar circuits; for details see the companies product-specific information.



Disadvantages associated with sheathed thermocouples are their relatively high prices and the absence of very thin sensor. Sheathed thermocouples have several advantages, such as (1) The metal sheathing is hermetically sealed, which makes the sensor useful in hostile environments. (2) It is tolerant to higher temperatures than wire sensor of corresponding size. (3) Sheathing can be easily bent to, and withhold, various shapes. (4) Vibration resistant.

Sheathing is available in different diameters and with exposed, grounded, and insulated measurement junctions (7,8). The exposed junction protrudes from the sheath and the tube is sealed, for example, by glass. The advantage is the shortest possible response time for sheathed thermocouples. However, several advantages of the fully enclosed sheathed thermocouple are lost; tolerance to high temperatures is one example. For the grounded junction, the wires are welded to the bottom of the sheathing tube. This gives a fairly fast response and good environmental protection, but sensitivity to ground currents. The insulated junction is not electrically connected with the sheath, which gives good protection against EM interference, but a slow temperature response.

MEASUREMENT ERRORS

Reliable temperature measurement is dependent on many factors that, more or less, inevitably leads to measurement errors. Consider, for example, a thermal treatment, such as radio frequency (RF) current ablation. During such a treatment, the temperature should ultimately be known in the entire heated area at every instant. However, this is obviously not possible. In practice, the temperature measurement is limited to a few single locations. In some situations, it might not even be possible to measure the temperature in any location inside the treatment area. In such cases, temperature information must be extracted from measurement outside of the treatment area, which gives space, time, and temperature uncertainties.

Insufficient Contact between Sensor and Measurement Object

Surface-mounted sensors are not only particularly sensitive to thermal shunting, but also to errors associated with an insufficient contact between the sensor and the measurement object. Suppose, for example, that a spherical sensor is placed in contact with a flat surface surrounded by air. Only a fraction of the circumference of the sensor will be in contact with the object to be measured, while the rest of the sensor is exposed to the temperature of the air. Thus, it is essential that the maximum possible contact surface be achieved between the sensor and the object to be measured, and/or that the sensor is thermally insulated from the surrounding.

Effects from adjacent hot or cold surfaces are especially treacherous as thermal radiation can effect the sensor temperature over relatively vast distances. An example is measurement of (hot) gas temperature in a chamber with cold walls; the walls will exchange energy with the sensor resulting in sensor temperature that is too low.

Thermal Shunting

A temperature sensor conducts heat, which gives what is called thermal shunting. The problem is particularly marked when the temperature is measured on pipes and other surfaces. The greater the temperature difference, the greater quantity of heat is “drawn out” of the body by the sensor. There is a sharp drop in temperature at the point where the sensor protrudes from the object to be measured. This is caused by a load being imposed on the object to be measured. The temperature drop is greatest when the object to be measured is a poor conductor of heat (low thermal conductivity). To overcome this measurement problem, the point at which the sensor leaves the body to be measured must be at a suitable distance from the measuring junction.

Temperature Gradients and Thermal Conduction

Heat losses through the probe occur when a sensor connects a warm zone with a colder one, especially if the sensor is the easiest path for the transfer of heat. Heat is conducted away from the object being measured to the probe. The heat flux gives a temperature drop and the sensor measures a lower temperature than the true value. Such losses can be effectively countered if we ensure that the thermal conductivity in the object and across the probe tip is much higher than it is along the sensor. In other words, more heat can be transferred to the sensor than conducted away by the sensor sheath.

Response Time

Since the sensor only measures its own temperature, there will almost always be a temperature deviation between the temperature of the sensor and its surrounding. In theory, the sensor temperature asymptotically approaches, but never reaches, the surrounding temperature. For practical purposes, the response time can be regarded as finite, but it can be both negligibly small and cause severe measurement errors.

The response time is defined as the time needed for the sensor to reach some fraction of a stepwise temperature change of the surrounding, often 90% or $1 - (1/e) = 67\%$. However, the response time of a sensor must be accompanied by information about the conditions under which the response time was obtained. The reason is that heat must be transferred from the surroundings to the sensor. This transfer takes some time, and the time needed differs greatly between different conditions, such as still air and stirred water. Too often, information of the conditions associated with a particular response time are left out, making the information useless. The response time is connected to the sensor and its surrounding and not the sensor itself.

Some critical factors affecting the response time of a sensor in contact with a body or medium include the following:

The heat capacity of the sensor. The greater heat capacity and mass, the longer response time.

The heat transfer in the materials. Air gaps and insulation, for example, reduce the heat transfer.

The contact surface between the sensor and the body or medium to be measured.

The response time can change over time; it can become shorter if for example, the junction of a sheathed probe is moved closer the sheath, or longer if, for example, a glue become brittle and porous and thereby gets different heat-transfer properties.

Self-Heating

Errors due to self-heating can be quite important in medical and biomedical applications depending on the small sensors often required (9). Resistance-based sensors and especially thermistors may be substantially affected by this type of error, but thermocouples are not affected as they are active sensors. All these types can, however, be affected by EM interference (see the sections Electromagnetic Interference and Electromagnetic Aspects).

Material Defects and Ageing

All types of thermocouple are subjected to varying degrees of wear and ageing depending on the environments in which they are used. It is therefore essential that all types of sensors are regularly inspected and calibrated. The type K thermocouple is the most widely used, and therefore the best documented (see, e.g., Ref. 7).

Use of Alternative Sensor Materials. Some thermocouples are expensive and it can therefore be tempting to use other materials as extension leads. As explained earlier, however, the Seebeck effect is created throughout the measuring circuit, so the use of other materials having a different Seebeck coefficient can give rise to a faulty output signal (7). There are materials, known as compensating leads, that have the same electrical properties as the thermocouple within limited temperature ranges, but it is always better to use thermoelectric material throughout the circuit. If this is not possible, compensating leads for the type of thermocouple used must be employed.

Alternative Probe/Junction Materials. Other materials are sometimes used for the measuring junction. A typical case is when the thermocouple wires are soldered to a tab, which is secured by screws at the required measuring point. This works if the tab has good electrical conductance and the temperature is the same at both ends of the wire. The wires must be as close together as possible so that they get the same temperature that constitutes the measured value.

Ageing. The ravages of time make their mark on all types of thermocouples. In general the process is accelerated by high temperatures ($\geq 200^\circ\text{C}$), vacuum, and various atmospheres. It is therefore of interest for only a limited amount of medical applications.

At high temperatures, short-range order (SRO) is a hysteretic phenomenon that is based on atomic migration inside the thermocouple, especially type K thermocouples.

Also type S thermocouples, which has wires of platinum–rhodium and pure platinum, are vulnerable to atomic migration. At high temperatures, the rhodium vaporizes and drifts across to the pure platinum wires, which results in a gradual fall in the output signal (see, e.g., Ref. 7.)

Connection Errors

Circuit Break (Open Circuit). A sensor wire has fractured, come adrift, or is making poor contact with the instrument. Modern instruments often trigger an alarm, for example, by Open appearing on the display.

Short Circuit. If the insulation has chafed and a short circuit occurs, a new measuring junction is created. The instrument will then display the temperature at the short circuit point, instead of at the tip of the probe. Sometimes, it can be very difficult to detect this type of error.

Reversed Polarity of Entire Measuring Circuit. If the polarity has been reversed, the instrument will also operate in reverse, that is, a temperature increase will be recorded as a temperature decrease.

Reversed Polarity within the Measuring Circuit. The extension lead must have the same polarity as the thermocouple wires. If the polarity of the thermocouple element is reversed, opposing voltages occur. The reading obtained will then be twice the temperature in the terminal head minus the temperature at the measuring junction.

Double Reversed Polarity. If the polarity of the extension lead has been reversed at both ends, the temperature at the ends will affect the output signal. The reading will be the temperature at the measuring junction less twice the temperature difference between the terminal head and the reference junction.

Electromagnetic Interference

Electromagnetic fields interact with materials inside the field. Thermocouples and other measurement sensors–probes are no exception, and the frequent occurrence of EM fields, especially in the medical–biomedical engineering sector, has contributed to many measurement errors.

The properties of an EM field change greatly with frequency. Interaction with the field is highly dependent on the geometry and the material of objects inside the field, which together with the great frequency spectrum of interest, makes evaluation of measurement interference multifaceted and intricate. Thus it is very difficult to predict and evaluate.

Measurement errors can arise due to direct effects, for example, induced current in the thermocouple wires (10). Remember that 1°C corresponds to $\sim 50\ \mu\text{V}$ for some typical thermocouples. Also, small currents can give significant measurement errors. Induced currents can also heat the wires and the measurement junction, leading to a too high temperature reading. Indirect disturbances can arise, for example, if the inserted probe disturbs the EM field. Such indirect effects are not only associated with metallic probes–material, but plastics as well (11).

THERMOCOUPLES IN MEDICINE AND BIOMEDICAL ENGINEERING

Thermal aspects, and thereby temperature measurement, are important in numerous applications in medicine and biomedical engineering. Thermocouple thermometry is still the dominant technique. Recent publications cover such varying fields as heating of pacemaker leads during magnetic resonance imaging (MRI) (12), estimation of brain protection during cardiopulmonary surgery (13), cryogen spray cooling (14), and thermal treatment using cryotherapy of breast fibroadenomas (15). Although the applications vary over a wide range, many of the fundamental questions are the same.

The following review discusses some applications in medicine and medical devices with a focus on temperature measurement during thermal treatment (treatment of disease and/or symptoms using heat or cold as one of the therapeutic agents). This is one of the major areas of interest for temperature measurement in medicine, and the application involves most of the interesting aspects, problems, and sources of errors associated with thermocouple thermometry. A more thorough review of the use and sources of errors of thermocouples in medicine can be found in Ref. 16, and some practical limitations are discussed in depth in Ref. 17.

Thermal Aspects

Methods, techniques, and equipments of thermal treatment cover a broad spectrum of applications, and the underlying medical as well as physical aspects are often substantially different. Many thermal and temperature measurement considerations are, however, equivalent, which is an important conclusion, as thermal aspects are often the dominant source of measurement errors (see, e.g., Ref. 18).

During thermal treatment, as in many other applications in medicine and elsewhere, it is often difficult to measure the temperature at the most important–interesting location(s). Furthermore, it is rarely a single or even a few temperature values that are solely of interest. More likely it is an entire temperature distribution: the temperature field. The temperature distribution is affected by many factors, and therefore is difficult to predict. During thermal treatment, time, temperature, and individually dependent blood perfusion are the most important and difficult aspects to consider. This is also the situation for other applications associated with human–animal temperature measurement at sites with a heterogeneous temperature distribution. Regardless of the application used, the relation between the measured temperature and the corresponding temperature distribution must always be taken under consideration.

An obvious example is functional neurosurgery using RF thermoablation. During such a treatment, the temperature is increased to 70–90 °C during 60 s in a small volume ($\sim 100 \text{ mm}^3$) surrounding the treatment electrode in order to destroy a malfunctioning tissue area. It is important to monitor the temperature in order to control the therapy, but practical difficulties have so far made intratissue

temperature measurement impossible. The thermocouple temperature reading is instead carried out inside the electrode that inevitably leads to both spatial and temporal temperature errors as both the maximum temperature and the interesting temperature distribution is located outside the electrode (19). The only possibility to carry out the treatment safely and efficiently is to map the temperature reading with the temperature distribution surrounding the electrode; experiments *in vivo*, *in vitro*, as well as modeling and computer simulations, can be useful tools (see, e.g., the review in Ref. 18).

Another major problem from a temperature measurement perspective is large temperature gradients that can give rise to measurement errors associated with uncertainty in probe positioning (20). Temperature gradients of $10^4 \text{ }^\circ\text{C}\cdot\text{m}^{-1}$ and even more are common during thermal treatment. The obvious effect is that an erroneous positioning of only a few tenths of a millimeter in such a steep gradient gives a measurement uncertainty of several degrees. This effect is particularly problematic since it by definition is an uncertainty and thus difficult to predict and compensate for.

Large temperature gradients also give multifaceted thermal conduction effects (21,22) as the probe in general has a different thermal conductivity compared with the measurement object (e.g., tissue). If the thermal conduction is larger in the probe compared with the tissue (as is the normal case), the probe will locally drain heat from the (heated) measurement location affecting the temperature distribution, while the probe itself experiences a temperature different from the adjacent tissue. Both these effects are highly time dependent; rapid temperature disturbances can initially be obtained during, for example, probe insertion and heating onset, although the probe-induced change of the temperature distribution can take much more time. Fortunately, it is possible to reduce the errors associated with these effects. The thermal conduction in the probe relative to the measurement object (e.g., tissue) can be estimated and compensated for, for example, by setting up a computer model that simulates the situation with and without the conduction effects (see the discussion in Ref. 18). Some theory of thermal conduction effects together with analytical, numerical, and experimental results, can be found in Ref. 23.

Electromagnetic Aspects

The use of EM energy for tissue heating purposes makes verification of the EM power deposition important. Since it is difficult to measure and model the EM distribution in general heterogeneous tissue and to obtain from this information an estimate of the corresponding heating effect, a more straightforward approach is by direct estimation of the heating deposition outgoing from the temperature measurement. This has been carried out using thermocouples in order to obtain information of the specific absorption rate (SAR), for example, in deep body regional hyperthermia (24).

When a probe is inserted in an EM field, a field perturbation followed by a changed power absorption in the vicinity of the probe might occur. This yields for all EM

fields, but the significance increases with the field strength that can cause troubles in, for example, thermal treatment and MRI. The volume occupied by a metallic material (probes) has substantially different electromagnetic properties compared with tissue. This can substantially change the field locally, and thereby also change the induced heating followed by an increased temperature.

Although thermocouples are not immune to either direct or indirect EM interference, the possibility to use very small sensors (conductor diameter ~ 0.1 mm) may for some applications keep this effect within reasonable or even negligible limits. In addition to the size, the shielding (11), the direction of the wires relative to the EM field (25), and the strength and frequency of the EM field, are all of importance to reduce EM interference. Perhaps the most serious source of direct EM interference is joule heating of the sensor-wires due to induced currents, which can cause heating of the sensor itself, as well as the surrounding tissue (21). Interference can also be reduced by an appropriate design of the measurement electronics and/or suppressed by filters (see, e.g., Refs. 10,16).

BIBLIOGRAPHY

- Doebelin EO. *Measurement Systems—Application and design*. 4th ed. New York: McGrawHill; 1990.
- Cengel Y, Turner R. *Fundamentals of Thermal-Fluid Sciences*. 2nd ed. New York: McGraw-Hill; 2005.
- Holman JP. *Heat Transfer*. 9th ed. New York: McGraw-Hill; 2002.
- Reed R. *Manual on the Use of Thermocouples in Temperature Measurement*. American Society for Testing and Materials; 1993.
- Dike PH. *Thermoelectric Thermometry*. Philadelphia: Leeds and Northrup; 1954.
- Moffat RJ. The gradient approach to thermocouple circuitry. *Temp Meas Control Sci* 1962;3(2).
- Pentronic AB (No date). Thermocouples. [Online]. Available at <http://www.pentronic.se/eng>. Accessed 2005.
- Omega Engineering Inc. (No date). Introduction to thermocouples. [Online]. Available at <http://www.omega.com/tech-ref/themointro.html>. Accessed 2005.
- Valvano JW, Nho S, Anderson GT. Analysis of the Weinbaum-Jiji model of blood flow in the canine kidney cortex for self-heated thermistors. *ASME—J Biomech Eng* 1994;116:201–207.
- Chakraborty DP, Brezovich IA. Error sources affecting thermocouple thermometry in RF electromagnetic fields. *J Microwave Power* 1982;17(1):17–28.
- Chan KW, Chou CK, McDougall JA, Luk KH. Changes in heating pattern due to perturbations by thermometer probes at 915 and 434 mhz. *Inter J Hyperther* 1988;4(4):447–456.
- Luechinger R, et al. *In vivo* heating of pacemaker leads during magnetic resonance imaging. *Eur Heart J* 2005;26(4):376–383.
- Kaukuntla H, et al. Temperature monitoring during cardiopulmonary bypass—do we undercool or overheat the brain? *Eur J Cardiothor Surg* 2004;26(3):580–585.
- Wangcun J, Aguilar G, Wang G, Nelson S. Heat-transfer dynamics during cryogen spray cooling of substrate at different initial temperatures. *Phys Med Biol* 2004;7(49).

- Littrup P, et al. Cryotherapy for breast fibroadenomas. *Radiology* 2005;234(1):63–72.
- Carnochan P, Dickinson RJ, Joiner MC. The practical use of thermocouples for temperature measurement in clinical hyperthermia. *Inter J Hyperthermia* 1986;1:1–19.
- van der Zee J, et al. Practical limitations of interstitial thermometry during deep hyperthermia. *Inter J Rad Oncol Biol Phys* 1998;40(5):1205–1212.
- Wren J. *On Medical Thermal Treatment—Modelling, Simulation and Experiments*. Ph.D. thesis dissertations. Linköpings universitet No. 763, 2002.
- Wren J, Eriksson O, Wårdell K, Loyd D. Analysis of temperature measurement for monitoring radio-frequency brain lesioning. *Med Biol Eng Comput* 2001;39:255–262.
- Wren J. Evaluation of three temperature measurement methods used during microwave thermotherapy of prostatic enlargement. *Inter J Hyperther* 2004;20(3).
- Constable RT, Dunscombe P, Tsoukatos A. Perturbation of the temperature distribution in microwave irradiated tissue due to the presence of metallic thermometers. *Med Phys* 1987; 14(3):385–388.
- Ryan TP, et al. Thermal conduction effects associated with temperature measurements in proximity to radio frequency electrodes and microwave antennas. *Inter J Rad Oncol Biol Phys* 1989;16:1557–1564.
- Samulski TV, Lyons BE, Britt RH. Temperature measurements in high thermal gradients: Analyses of conduction effects. *Inter J Rad Oncol Biol Phys* 1985;11:963–971.
- de Leeuw AAC, Crezee J, Lagendijk JJW. Temperature and sar measurements in deep-body hyperthermia with thermocouple thermometry. *Inter J Hyperther* 1993;9(5):685–697.
- Dunscombe PB, McLellan J. Heat production in microwave-irradiated thermocouples. *Med Phys* 1986;13(4):457–461.

See also TEMPERATURE MONITORING; THERMOMETRY.

THERMODILUTION. See CARDIAC OUTPUT, THERMODILUTION MEASUREMENT OF.

THERMOGRAPHY

HAIRONG QI
University of Tennessee
Knoxville, Tennessee
NICHOLAS A. DIAKIDES
Advanced Concepts Analysis, Inc.
Falls Church, Virginia

INTRODUCTION

The first documented application of infrared (IR) imaging in medicine was in 1956 (1), when breast cancer patients were examined for asymmetric hot spots and vascularity in IR images of the breasts. Since then, numerous research findings have been published (2–4) and the 1960s witnessed the first surge of medical application of the IR technology (5,6), with breast cancer detection as the primary practice. However, IR imaging has not been widely recognized in medicine nowadays, largely due to the premature use of the technology, the superficial understand-

ing of IR images, and its poorly controlled introduction into breast cancer detection in the 1970s (7).

Recently, advances in a couple of related areas have pushed forward series of activities to reappraise the role of IR imaging in medicine (7–12). These advances, including the development of the new-generation IR technology, smart image processing algorithms, and the pathophysiological-based understanding of IR images, will provide a cost-effective, noninvasive, nondestructive, and patient-friendly approach to health monitoring and examination, as well as to assisting diagnosis. These new developments are discussed in detail in this article.

Temperature is a long established indicator of health. The Greek physician, Hippocrates, wrote in 400 bc. “In whatever part of the body excess of heat or cold is felt, the disease is there to be discovered (13).” The ancient Greeks immersed the body in wet mud and the area that dried more quickly, indicating a warmer region, was considered the diseased tissue. The use of hands and thermometers to measure heat emanating from the body remained well into the sixteenth through the eighteenth centuries. Nowadays, we still rely on thermometers a lot when performing health examination.

All the above-mentioned methods are contact based. Since the British astronomer, Sir William Herschel, discovered the existence of IR radiation in 1800, major advances have taken place with IR imaging that do not need direct contact with the patient.

Infrared radiation occupies the region between visible and microwaves of the spectrum. All objects in the universe emit radiations in the IR region as a function of their temperature. As an object gets hotter, it gives off more intense IR radiation, and it radiates at a shorter wavelength (11). The human eye cannot detect IR rays, but they can be detected by using the IR cameras and detectors. Figure 1 illustrates the IR spectral band in finer scale. The boundaries between different IR spectral regions are not agreed upon and can vary. The boundaries that we adopt here are based on Refs. (14–18).

In general, IR radiation covers wavelengths that range from 0.75 to 1000 μm , among which the human body emissions that are traditionally measured for diagnostic purposes only occupy a narrow band at wavelengths of 8–12 μm (19). This region is also referred to as the long-wave

IR (LWIR) or body infrared rays. Another terminology that is widely used in medical IR imaging is thermal infrared (TIR), which, as shown in Fig. 1, covers wavelengths beyond $\sim 1.4 \mu\text{m}$. Within this region, the IR emission is primarily heat or thermal radiation, and hence the term thermography. The image generated by TIR imaging is referred to as the thermogram. The near infrared (NIR) region occupies wavelengths between 0.75 and 1.4 μm . The IR emission that we observe in this region is not thermal (17). Although the NIR and mid-wave IR (MWIR) regions are not traditionally used in human body screening, the new generation detectors have enabled the use of multi-spectral imaging in medicine, in which both NIR (20) and MWIR (21) are observed in different diagnostic cases.

In this article, we discuss IR imaging in medicine across the full IR spectral region with a focus on the thermal IR region, including the pathophysiological understanding of IR imaging, the development of new generation of IR imagers, and the advanced image processing algorithms of IR images.

PATHOPHYSIOLOGICAL-BASED UNDERSTANDING OF INFRARED IMAGING

Infrared imaging is a physiological test that measures the subtle physiological changes that might be caused by many conditions, for example, contusions, fractures, burns, carcinomas, lymphomas, melanomas, prostate cancer, dermatological diseases, rheumatoid arthritis, diabetes mellitus and associated pathology, deep venous thrombosis (DVT), liver disease, bacterial infections. These conditions are commonly associated with regional vasodilation, hyperthermia, hyperperfusion, hypermetabolism, and hypervascularization (19,22–27), which generate higher temperature heat source. Unlike imaging techniques, such as X-ray radiology and Computed Tomography (CT) that primarily provide information on the anatomical structures, IR imaging provides functional information not easily measured by other methods. Thus correct use of IR images requires in-depth physiological knowledge for its effective interpretation.

Human Thermal Models

The heat emanating on to the surface from the heat source and the surrounding blood flow can be quantified using the

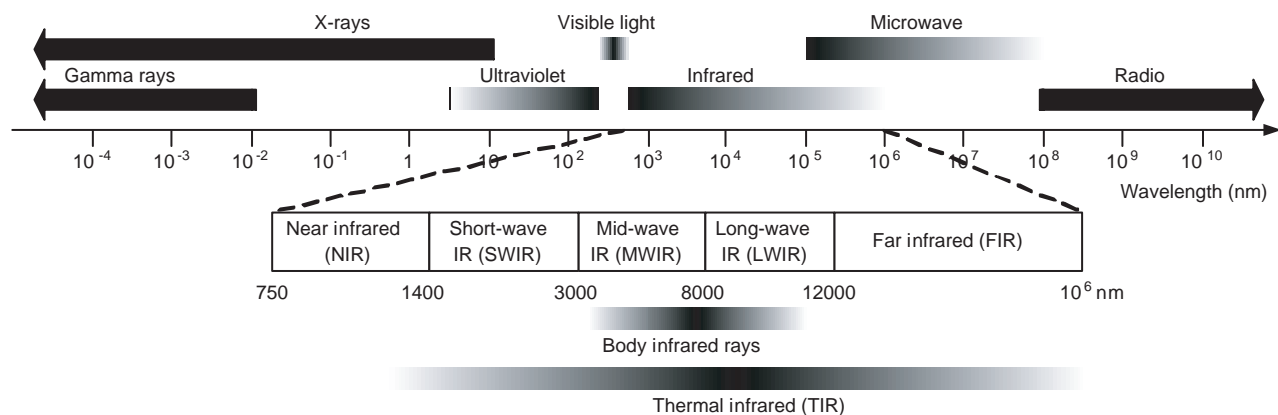


Figure 1. The electromagnetic spectrum and the IR region.

Pennes' bio-heat equation (28). This equation includes the heat transfer due to conduction through the tissue, the volumetric metabolic heat generation of the tissue, and the volumetric blood perfusion rate whose strength is considered to be the arteriovenous temperature difference (29). The equation is given as

$$k \Delta^2 T - c_b w_b (T - T_a) + q_m = 0 \tag{1}$$

where k is conductivity, q_m is volumetric metabolic rate of the tissue, $c_b w_b$ is the product of the specific heat capacity and the mass flow rate of blood per unit volume of tissue, T is the unknown tissue temperature, and T_a is the arterial temperature. In theory, given the heat emanating from the surface of the body measured by TIR imaging, by solving the inverse heat transfer problem, we can obtain the heat pattern of various internal elements of the body. Different methods of solving the bio-heat transfer equation have been presented in literature (30,31). Although it is possible to calculate the thermal radiation from a thermal body by thermodynamics, the complexity of the boundary conditions associated with the biological body makes this approach impractical.

One of the biggest hurdles in the diagnosis using thermograms is the various thermal environmental conditions that could affect detection and evaluation to a great extent. A computer model was presented in (32) that aims to simulate the heat-transfer phenomenon within the human body and predict the internal temperature as well as the skin surface temperature, providing a reference model that thermograms taken under different thermal conditions can be converted between each other. Figure 2 illustrates the 16-cylinder-segment model this work is based on as well as the simulated body temperature profile.

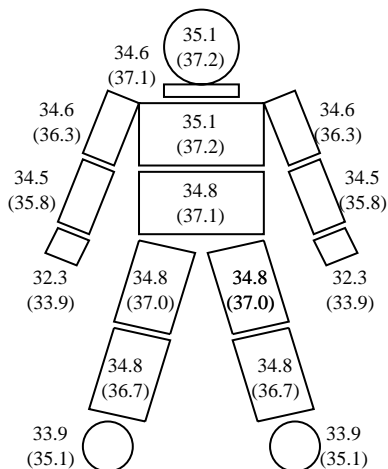


Figure 2. Simulated human body temperature profile based on the 16-cylinder-segment model, after the first 60 min. The air temperature and mean radian temperature were maintained at 30 °C for 60 min, then changed to 24 °C and maintained at that temperature for another 60 min. The value in the parentheses is the temperature at the center of the segment. (Redrawn from Ref. 32.)

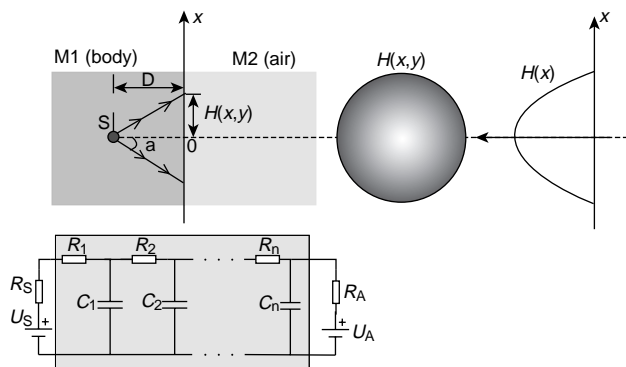


Figure 3. The thermal-electric analog.

The Thermal-Electric Analogue

Liu et al. (33,34) presented a new method for analyzing a thermal system based on an analogy to electric circuit theory; referred to as thermal-electric analog. This method does not require a direct solution to the inverse heat-transfer problem. Figure 3 illustrates the analogy between thermodynamics systems and the electrical circuit, where a battery with voltage U_S is used to simulate the heat source S inside the human body and the heat loss inside the heat source can be simulated as the heat loss on a resistor R_S . Hence, we can establish the correspondence between the temperature of the heat source and the voltage of the battery, as well as between the heat current and the circuit current. The set of R_i and C_i values correspond to the unit heat resistance and heat capacity along each radiation line. The circuit in Fig. 3 only shows the analogy for one radiation line. If the medium between the heat source (S) and the surface is homogeneous, then the radiation pattern sensed by the IR camera at the surface should have a distribution like Gaussian as shown in Fig. 3. If the medium is not homogeneous, then the surface radiation pattern can be represented as a linear combination of different Gaussain distributions.

This analogy can be used to estimate the depth of the heat source (34), and furthermore, help understand the metabolic activities undergoing within the human body, through a so-called slicing technique. The method has been used in early breast cancer detection and has achieved high sensitivity. It has also recently been used for the diagnosis of severe acute respiratory syndrome (SARS) patients, as reported in (35).

The Abnormal Thermogram Patterns

As mentioned before, one of the biggest hurdles in IR image interpretation is the lack of standardized image handling procedures. The human thermal model used in the section Human Thermal Models is one attempt in solving this problem. Fujimas also did some pioneer work (36) in 1998 by proposing eight thermophysiological expressions to identify abnormal thermogram patterns, referred to as the *thermatomes*.

Angiological *thermatomes*: Abnormal temperature regions caused by organic vascular abnormalities.

Functional angiological thermatomes: Abnormal temperature regions caused by vascular disfunctions.

Neurodermatomal thermatomes: Abnormal temperature bands caused by somatosensory neuronal disorders.

Myotomal thermatomes: Abnormal temperature regions suspected by abnormal muscular blood flow rate.

Metabolic thermatomes: Abnormal hot and/or cold spots caused by excessive and/or lower heat production and blood flow.

Dynamic thermatomes at environmental temperature stress: Regions with abnormal reactions when a patient received an applied thermal load.

Dynamic thermatomes at medication: Regions with abnormal reactions when a patient is given a medication.

Dynamic thermatomes at various kinds of stress: Regions with abnormal reactions when a patient receives a load (various in type).

IR Imaging in Early Breast Cancer Detection

Because IR imaging has been mainly used in breast cancer detection since its introduction to the medical field, in the following, we focus on the potential of IR imaging, especially TIR imaging, in *early* breast cancer detection.

Cancer cells are resulted from permanent genetic change in a normal cell triggered by some external physical agents such as chemical agents, X rays, ultraviolet (UV) rays, and so on. All types of cancer cells have an imbalanced metabolic activity that leads to the utilization of a large amount of blood glucose and the release of large amounts of lactate into blood. In addition, the high metabolic rate of cancer cells causes an increase in local temperature as compared to normal cells. These factors have enabled IR imaging as a viable technique to visualize the abnormality. The IR image provides more dynamic information of the tumor since the tumor can be small in size but can be fast growing making it appear as a high temperature spot in the IR image (37,38).

Many imaging modalities can be used for breast screening, including mammography using X-ray, IR, magnetic resonance imaging (MRI), CT, ultrasound, and positron emission tomography (PET) scans. Although mammography has been the baseline approach, it depends primarily on structural distinction and anatomical variation of the tumor from the surrounding breast tissue (7). Unless the tumor is beyond certain size, it cannot be imaged as X rays essentially pass through it unaffected. Other modalities like MRI and PET scan could provide valuable information to diagnosis, but they are not popularly adopted for various reasons including high cost, complexity and accessibility issues (12). Compared to mammography, MRI, CT, ultrasound, and PET scans that are also called the after-the-fact (a cancerous tumor is already there) detection technologies, IR imaging is able to detect breast cancers 8–10 years earlier than mammography (39,40). Keyserlingk reported in (7) that the average tumor size undetected by IR imaging is 1.28 cm versus 1.66 cm by mammography.

Samples of Other Advanced Interpretations

We would also like to mention two interesting work conducted recently although their influence on diagnosis is yet to be investigated. Alexjander and Deamer (41) propose to study the sound (rhythms and frequencies) made within the human body through the access of the IR frequencies of DNA bases. Imagine if we can “hear” the body, would a pleasing pattern to the ear indicate a healthy subject? Or would different patterns present a sign of a certain disease? Through nonlinear heat transfer modeling, Pavlidis and Levine (42) show that the periorbital blood flow in anxious states can be used to extract subtle facial temperature fluctuation patterns and thus assist in traditional polygraph examination. Perhaps if we go beyond imagination, more exciting applications of IR imaging can come into the light.

NEW GENERATION INFRARED TECHNOLOGIES

Infrared technology owes its origin to military research. Since IR imaging was first introduced to medical diagnosis in the 1960s, most of the IR equipment used has not been specifically designed for the medical market (23). Some of the problems associated with IR cameras at that time, for example, narrow field of view ($<20^\circ$) and low spatial resolution (~ 200 optical lines), although are not issues in military applications, they have affected the effectiveness and accuracy of diagnosis to a great extent. Some recent advances in IR sensor design expect to solve these problems and make IR sensors adequate for medical applications. In the following, we focus our discussion on the advances in the detector technologies, especially the uncooled camera development.

Cooled versus Uncooled Thermal Detectors

To some extent, the main factor that determines which wavelengths are included in which IR region is the type of detector technology used to capture IR radiation (17). The NIR radiations are observed in very similar way as the visible light, except that special IR detectors need to be used. On the other hand, TIR imaging generally requires the use of a cooling system in the form of a nitrogen or compressed air cooling bottle, which contains crystals like germanium whose electrical resistance is very sensitive to heat. Figure 4 shows the two main assemblies of the EYE-Z640 cooled FLIR (forward looking IR) camera from OPGAL Optronic Industries Ltd. (43). The detector is InSb cryogenically cooled and needs extra gadgets like the cooler and the dewar to support the cooling system. Compared to uncooled IR cameras, although cooled systems generally present better sensitivity, they consume more power, need a relatively longer cooling down time (e.g., a few minutes), and are more expensive. In addition, the average time that cooled IR cameras will function before failing is very limited (around a few thousand hours). Due to the size, weight, and complexity, these systems were limited to fixed deployment like tripod mounting.

The advance in solid state models has made a new class of sensors possible, the uncooled detector design. In the

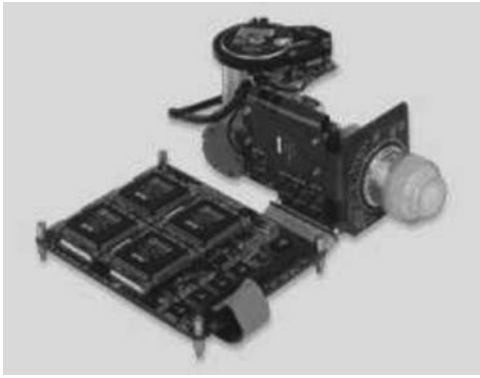


Figure 4. EYE-Z640 InSb cryogenically cooled FLIR camera from OPGAL (44) with the electronic card (lower left corner) and the DDCE (upper right corner: Detector, Dewar, Cooler, and Electronics).

1980s, the Department of Defense (DoD) sponsored companies like Honeywell and Texas Instruments (TI) with large classified contracts to develop uncooled IR detector technology (45). Honeywell’s microbolometer and TI’s pyroelectric sensors are both successful deliverables from these programs. In 1999, the Defense Advanced Research Projects Agency (DARPA) issued a Broad Agency Announcement (BAA) (46) that solicits proposals for increasing the performance of the uncooled IR sensors to their theoretical limit. The objective for the thermal sensitivity is set at <10 mk with the pixel size less than or equal to 25 μm. As far as the array size, high performance arrays for long-range systems can be as large as 960 × 1280 elements, while arrays for microsensors may be as small as 240 × 320 elements (46).

Compared to other uncooled IR detector technologies, like ferroelectric and pyroelectric, microbolometer sensors are less expensive, providing higher dynamic range, broader spectral range, and lower cross-talk. Therefore, this type of uncooled sensors are more popularly used, especially after DoD declassified the microbolometer technology in 1992.

The microbolometer technology, which is thermalelectric in nature, converts IR energy to a change in resistance. Each microbolometer detector consists of a silicon nitride microbridge that lies above a silicon substrate and is supported by silicon nitride legs, as shown in Fig. 5. A bolometer is a thermal detector that is deposited on the bridge. When heated by incoming radiation, the bolometer detector can result in a temperature rise that is sensed as a change in the element resistance.

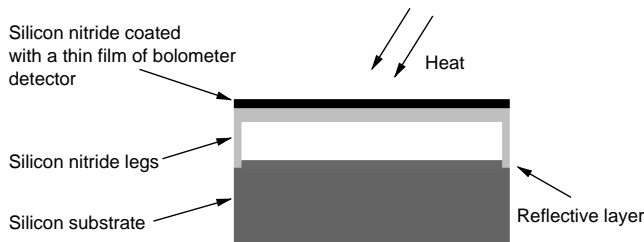


Figure 5. The microbolometer detector layout.



Figure 6. Bioyear’s PRISM 2000 thermal metabolic imaging system (Left: Gantry that hosts the FLIR camera. Middle: Console. Right: Bed). (Courtesy of Bioyear, Inc.)

Because the uncooled cameras do not require a cooling system, they are much lighter, smaller, more reliable and less expensive compared to the cooled cameras. Currently, the uncooled cameras are approaching to the thermal sensitivity of the cooled ones (0.05° or 0.02 °C of uncool vs. 0.01 °C of cool) and are very popular in breast imaging. Figure 6 shows the PRISM 2000 Thermal Metabolic Imaging system manufactured by Bioyear, Inc. (47), in which a FLIR microbolometer detector is installed on the gantry. The detector consists of 320 × 240 pixels with a sensitivity of 0.05 °C.

Although we have listed quite a few advantages of cooled IR cameras, the competition between uncooled and cooled detectors continues, especially with the recent development of the deeply cooled QWIP (quantum well photodetector) detectors. Wiecek conducted a brief comparison (48) between uncooled thermal detectors and QWIP and discussed the limits in both technologies.

Three Generations of Development

Since its first appearance, the development of thermal imagers has gone through three generations of evolution.

The first generation thermal imagers were fielded in the 1970s. They use a single detector or small-size linear array detectors. In order to generate the picture, two scanning mirrors are used. This generation imagers generally have the white out problem (or over saturation over high intensity sources). Although mechanical brightness controls are used later to address the problem, the images still lack clarity.

Second generation imagers appeared in the 1980s. They use a relatively larger linear array (~120 elements) or small two-dimensional (2D) focal plane array (FPA) (~64 × 64 elements) and the scanning mirrors are still used to generate the picture.

Third generation imagers upgrade the size of the 2D FPA a great deal, some of which contain as many elements as 1024 × 1024. In addition, the image processing capabilities are integrated on the FPA, hence the so-called on-chip image processing. According to Xenics’ definition (49), FPA is a matrix of detector cells that attached to a semiconductor chip. Each cell is responsive in IR wavelengths, in which it absorbs IR radiation, converts it into electrons, and sends a voltage signal in response to form an image. The FPA can capture multicolor images and brings great advantages to image capturing, including emissivity correction, lower

atmosphere influence on the temperature measurement, and so on. The third generation does not use mirrors that largely improves the image quality as the less moving part in the camera, the more reliable the system, and the less mechanical noise. Currently, the third generation FPA detectors can capture wavelength from 3 to 5 or 8–12 μm .

One of the most important features that distinguishes the third generation design is the employment of the time-delay integration (TDI) technique for image integration and enhancement. The TDI is a specialized detector readout mode. Instead of reading out the entire chip as a single large image, the image is read out continuously, line by line from the bottom of the detector chip. If the readout rate of the detector and the velocity of the object being imaged matches each other, then there will not be motion blurs.

Smart Image Processing Approaches to IR Images

Computer-aided diagnosis (CAD) has been playing an important role in the analysis of IR images, as human examination of images is often influenced by various factors like fatigue, being careless, and so on. The detection accuracy is also confined by the limitations of human visual system. On top of all these factors, a shortage of qualified radiologists also put an urgent demand on the development of CAD technologies. Currently, research on smart image processing algorithms on IR images tends to improve the detection accuracy from three perspectives: smart image enhancement and restoration algorithms, asymmetry analysis of the thermogram including automatic segmentation approaches, and feature extraction and classification.

Smart Image Enhancement and Restoration Algorithms

One of the problems with thermograms that has put IR imaging in a somewhat disadvantage situation is its lack of resolution due to blur compounded by rather high levels of noise. Snyder et al. (50) developed an algorithm to increase the effective resolution of thermograms by a 2:1 ratio while at the same time removing the noise and preserving edges in the image. This algorithm is based on a minimization strategy known as mean-field annealing, which takes into account processes of blur, noise, and image correlations, to make an optimal estimate of the missing pixels.

MIT's researchers attempt to enhance the resolution of IR images through another route. The Minimally Invasive Optical Biopsy System developed at MIT (51) uses IR light in conjunction with an intravenously injected dye and special computer software to create a clear, high contrast image that could easily allow physicians to detect breast masses and determine if they are benign or malignant.

Kaczmarek and Nowakowski (52) proposed the use of active dynamic thermography (ADT), commonly adopted in nondestructive testing of materials, to further enhance the image quality. ADT analyzes thermal transients after the application of external thermal excitation. Some preliminary results have shown the promise of this approach.

Asymmetry Analysis

Making comparisons between contralateral images are routinely done by radiologists. When the images are rela-

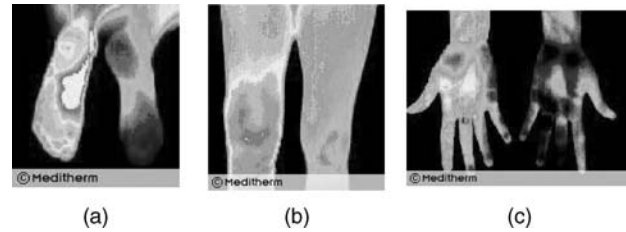


Figure 7. Case studies from Meditherm (53).

tively symmetrical, small asymmetries may indicate a suspicious region. From the human thermal model depicted in Fig. 2, the symmetry in temperature distribution between the left and the right part of the human body is very obvious. However, these small asymmetries might not be easy to detect and it is important to design an automatic approach to eliminate human factors. Figure 7 shows four case studies on how asymmetric thermal signatures have indicated different types of disease. These studies are originally conducted by Meditherm (53). Figure 7a shows a patient with complex regional pain syndrome (CRPS) in the right foot, which is developed after a fractured calcaneum 18 months previously. The thermogram indicates the right foot is 3.7 $^{\circ}\text{C}$ colder than the left foot. It is reported from the same study that some cases of CRPS are misdiagnosed as psychological or hysterical pain states but thermography is able to show characteristic changes. Figure 7b shows a patient with right knee surgery followed with a painful effusion in the early post operative period. In this case, thermography is able to confirm a significant inflammatory reaction and 30ml of blood-stained fluid was aspirated. Figure 7c shows a patient with the left wrist injured three years ago. Thermogram indicates obvious temperature change in the left wrist and hand.

Head et al. (54,55) recently analyzed the asymmetric abnormalities in IR images. In their approach, the image is segmented first by operator. Then breast quadrants are derived automatically based on unique point of reference, that is, the chin, the lowest, rightmost and leftmost points of the breast.

Qi and Head (56) developed an automatic approach to asymmetry analysis in IR images. It includes automatic segmentation and pattern classification. Hough transform is used to extract the four feature curves that can uniquely segment the left and right breasts. The feature curves include the left and the right body boundary curves, and the two parabolic curves indicating the lower boundaries of the breasts. Figure 8 shows the segmentation results of two patient images obtained using the Inframetrics 600M camera, with a thermal sensitivity of 0.05 K. The images are collected at Elliott Mastology Center. The results include the intermediate images from edge detection, feature curve extraction, to segmentation. From the figure, we can see that Hough transform can derive the parabola at the accurate location.

Mabuchi et al. (57) designed a computerized thermographic system, which would produce images of the distribution of temperature differences between the affected side and the contralateral healthy side. Because there is no standard skin surface temperature existed, the system

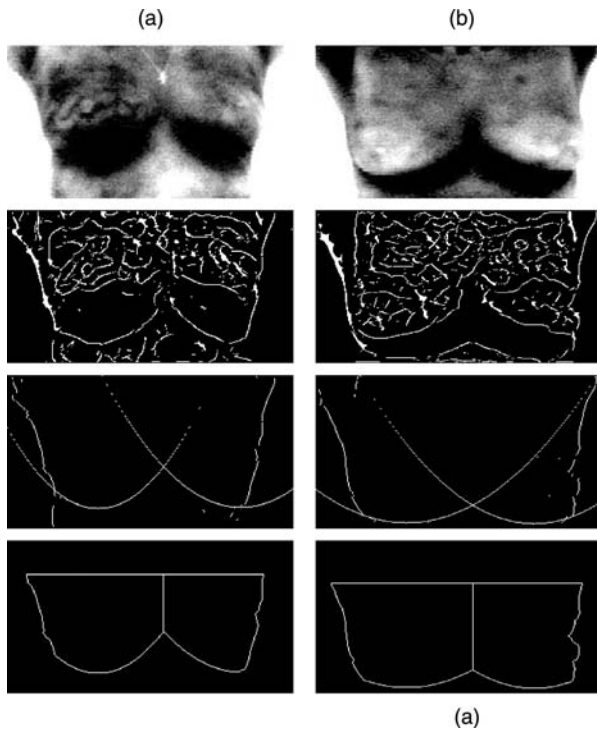


Figure 8. Segmentation results of two images. (a) results from cancerous patient. (b) Results from normal patient. From top to bottom: original image, edge image, four feature curves, segments.

measures the body-surface temperature of each pixel in the affected area and subtract from it the body-surface temperature of the corresponding pixel in the symmetrically located contralateral healthy area to generate the difference image.

Feature Extraction and Classification

Upon segmentation, different features can be extracted from the segments. Asymmetric abnormalities can then be identified based on mature pattern classification techniques. In this process, feature extraction is crucial to the success of computer-aided diagnosis (58) shows that the high order statistics (e.g., variance, skewness, and kurtosis) and joint entropy are the most effective features to measure the asymmetry, while low order statistics (e.g., mean) and entropy do not assist asymmetry detection. Jakubowska et al. (59) also addressed the importance of using statistical parameters (first and second order) in extracting thermal signatures for asymmetry analysis. From the figure, we observe that the high order statistics are the most effective features to measure the asymmetry, while low order statistics (mean) and entropy do not assist asymmetry detection. Figure 9 compares the effectiveness of the features used to analyze the symmetry of the segments derived in Fig. 8. The first data point along the x -axis indicates entropy, the second to the fifth points indicate the four statistical moments (means, variance, skewness, and kurtosis). The y axis shows the closeness metric we defined as the absolute difference between 1 and the ratio between the feature value from the left segment and that from the right segment. Hence, the smaller the difference, the more

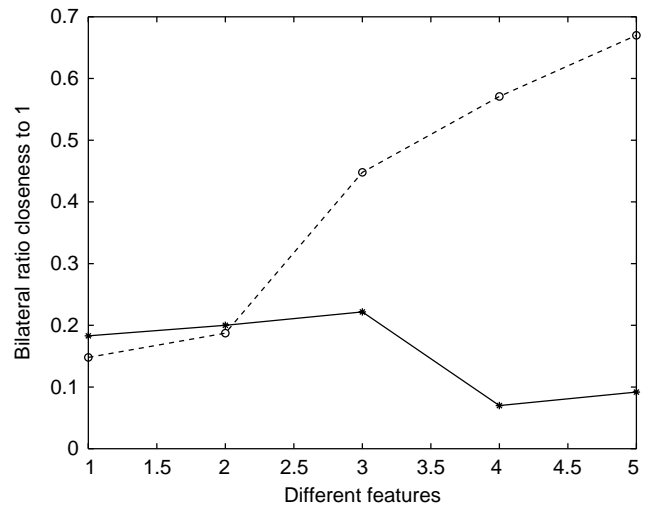


Figure 9. Performance evaluation of different feature elements. Solid line: noncancerous image; Dash line: cancerous image. The five data points along the x axis indicate (from left to right): entropy, mean, variance, skewness, kurtosis.

symmetric the two segments. We observe from the figure that the high order statistics are more effective features to measure the asymmetry than low order statistics (mean) or entropy.

Szu et al. (60) proposed a new paradigm shift that uses at least two dual-band (mid and long) IR imaging cameras operating simultaneously on the patient. This system enables a smart brain-like neural network algorithm, the Lagrange Constraint Neural Network (LCNN), to achieve submillimeter scaling of the close-up breast imaging for the vascular and angiogenesis effects as well as stage-zero detection of ductal carcinoma *in situ*.

The Dynamic Area Telethermometry Technique

To be able to yield objective clinical diagnosis, Anbar et al. (61) proposed the dynamic area telethermometry (DAT) technique. It has been demonstrated to be applicable to any quantitative patho-physiological assessment. The authors demonstrated that using classical fast Fourier transform (FFT) and elementary statistics, the large amount of sequential observations can be reduced to a single quantitative diagnostic parameter without the participation of human experts. Other related work also reported in (62,63).

The above mentioned techniques are just samples of activities reported in recent conferences, workshops, and symposia. Another trend of effort that is worth mentioning is the transition of automatic target recognition (ATR) algorithms developed for military application to medicine. "From tanks to tumors" (64,65) has been the theme of this transition and the rich collection of ATR algorithms that the military has sponsored will greatly improve the state-of-the-art of CAD development.

Concept Validation

Concept validation is an important procedure in the promotion of IR-based breast screening where blind diagnosis

and clinical evidence are necessary. Although there have been a lot of clinical trials conducted so far (66), there has not been a well-designed, standard database created for the purpose of concept validation.

The Advanced Concept Analysis, Inc. located at Falls Church, VA was awarded in 2000 to manage the creation of such a database. The project is sponsored by the Deputy Assistant Secretary of the Army for Installations and Environment (Environmental Safety and Occupational Health), the Office of the Deputy Undersecretary of Defense for Science and Technology (ODUSD/S&T), the Air Force Research Laboratory (AFRL), and the Office of Naval Research (ONR). An Internet-based Virtual Distributed Laboratory (VDL) at AFRL will house >8000 images from >2000 patients provided by E.H.H. Breast Cancer Research and Treatment Center, Baton Rouge, LA and Ville Marie Medical Center & Women's Health Center, Montreal, Canada. Each center will use the collaboration tools and evaluation procedures on the VDL to conduct blind diagnoses of the images provided by the other. Blind test results will be compared with actual clinical evidence stored with the imagery. VDL access may be applied for at (67).

INTERNATIONAL IR IMAGING ACTIVITIES IN MEDICINE

Before this article is summarized let's take a quick scan of activities reported worldwide on the usage of IR imaging in medicine.

United States of America and Canada: Infrared imaging is beginning to be reconsidered in the United States, largely due to the factors discussed early, that is new IR technology, advanced image processing, in-depth pathophysiological understanding of IR images. Currently, there are several academic institutions with research initiatives in IR imaging. Some of the most prominent include National Institute of Health (NIH), Johns Hopkins University (JHU), University of Houston, and University of Texas. The NIH has several ongoing programs, such as vascular disorders (diabetes, DVT), monitoring angiogenesis activity (Kaposi Sarcoma, pain reflex sympathetic dystrophy (68), monitoring the efficacy of radiation therapy, organ transplant) perfusion, multispectral imaging, and so on (69–71). The JHU does research in microcirculation, monitoring angiogenic activity in Kaposi Sarcoma and breast screening, laparoscopic IR images for renal disease. University of Houston just created an IR imaging laboratory to investigate with IR the facial thermal characteristics for such applications as lie detection and other behavioral issues (fatigue, anxiety, fear, etc.) (42). There are two medical centers specializing in breast cancer research and treatment which use infrared routinely as part of their first line detection system, which also includes mammography and clinical exam. These are: EHH Breast Cancer and Treatment Center, Baton Rouge, LA. and Ville Marie Oncology Research Center, Montreal, Canada

China: China has a long-standing interest in IR imaging. More recently, the novel method Thermal Texture Mapping (TTM) (34,35) has added increased specificity to static imaging. It is known that this method is widely used in this country. Introduction of TTM has been made to

NIH. They have been using this method successfully in detection and treatment in Kaposi Sarcoma (associated with AIDS patients). There are further possibilities for high-level research for this method in the United States and abroad. The TTM technology developed by Bioyear (47) has also been adopted in the EHH Center and the Ville Marie Center mentioned above.

Japan: Infrared imaging is widely accepted in Japan by the government and the medical community. More than 1500 hospitals and clinics use IR imaging routinely (70). The government sets the standards and reimburses clinical tests. Their focus is in the following areas: blood perfusion, breast cancer, dermatology, pain, neurology, surgery (open-heart, orthopedic, dental, cosmetic), sport medicine, oriental medicine. The main research is performed at the following universities: University of Tokyo—organ transplant; Tokyo Medical and Dental University (skin temperature characterization and thermal properties); Toho University—neurological operation); Cancer Institute Hospital (breast cancer). In addition, around forty other medical institutions are using infrared for breast cancer screening.

Korea: Began involvement in IR imaging during the early 1990s. More than 450 systems have been used in hospitals and medical centers. Primary clinical applications are neurology, back pain/treatment, surgery, oriental medicine. Yonsei College of Medicine is one of the leading institutions in medical IR imaging research along with three others.

United Kingdom: The University of Glamorgan is the center of IR imaging; the School of Computing has a thermal physiology laboratory which focuses in the following areas: medical IR research, standardization, training (university diploma), "SPORTI" Project funded by the European Union Organization. The objective of this effort is to develop a reference database of normal, thermal signatures from healthy subjects. The Royal National Hospital of Rheumatic Diseases specializes in rheumatic disorders, occupational health (Raynaud's Disease, Carpal Tunnel Syndrome and Sports Medicine). The Royal Free University College Medical School Hospital specializes in vascular disorders (diabetes, DVT, etc.), optimization of IR imaging techniques and Raynaud's Phenomenon).

Germany: University of Leipzig uses IR for open-heart surgery, perfusion, and microcirculation. There are several private clinics and other hospitals that use infrared imaging in various applications. EvoBus-Daimler Chrysler uses IR imaging for screening all their employees for wellness/health assessment (occupational health). InfraMedic, AG, conducts breast cancer screening of women from 20–85 years old for the government under a 2 year grant and IR is the sole modality used.

Austria: Ludwig Boltzmann Research Institute for Physical Diagnostics has done research in IR for many years and it publishes the *Thermology International* (a quarterly journal of IR clinical research and instrumentation). The General Hospital, University of Vienna, does research mainly in angiology (study of blood and lymph vessels) diabetic foot (pedobarography).

Poland: There has been a more recent rapid increase in the use of IR imaging for medicine in Poland since the

Polish market for IR cameras was opened up. There are >50 cameras being used in the following medical centers: Warsaw University, Technical University of Gdansk, Poznan University, Lodz University, Katowice University and the Military Clinical Hospital. The research activities are focused on the following areas: Active Infrared Imaging, open-heart surgery, quantitative assessment of skin burns, ophthalmology, dentistry, allergic diseases, neurological disorders, plastic surgery, thermal image database for healthy and pathological cases and multispectral imaging (IR, visual, X-ray, ultrasound) (52).

Italy: Much of the clinical use of IR imaging is done under the public health system, besides private clinics. The ongoing clinical work is in the following areas: dermatology (melanoma), neurology, rheumatology, anaesthesiology, reproductive medicine, and sports medicine (72). The University of G. d'Annunzio, Chieti, has an imaging laboratory purely for research on infrared applications.

SUMMARY

This article discussed recent research achievements in medical thermography. The objective is to show that due to the advances in IR technology, image processing techniques, and the pathophysiological-based understanding of thermograms, IR imaging is mature to be used as a first line supplement to both health monitoring and clinical diagnosis. We have established a website (73) to facilitate researchers working in the field of medical thermography to exchange research findings. We welcome contributions to enrich this list of collections.

BIBLIOGRAPHY

- Lawson RN. Implications of surface temperature in the diagnosis of breast cancer. *Can Med Assoc J* 1956;75:309–310.
- Handley RS. The temperature of breast tumors as a possible guide to prognosis. *Acta Unio Int Contra Cancrum* 1962; 18:822.
- Lawson RN, Chughtai MS. Breast cancer and body temperatures. *Can Med Assoc J*. 1963;88:68–70.
- Lloyd-Williams K, Handley RS. Infrared thermometry in the diagnosis of breast disease. *Lancet* 1961; (2):1378–1381.
- Gershen-Cohen J, Haberman J, Brueschke EE. Medical thermography: a summary of current status. *Radiol Clin N Am* 1965;3:403–431.
- Haberman J. The present status of mammary thermography. *Ca - A Cancer J Clin* 1968;18:314–321.
- Keyserlingk JR, et al. Functional infrared imaging of the breast. *IEEE Eng Med Bio Mag* May–June 2000; pp. 30–41.
- Anbar M. Quantitative and dynamic telethermometry—a fresh look at clinical thermology. *IEEE Eng Med Bio Mag* Jan.–Feb. 1995;14(1):15–16.
- Head JF, Elliott RL. Infrared imaging: making progress in fulfilling its medical promise. *IEEE Eng Med Bio Mag* Nov.–Dec. 2002;21(6):80–85.
- Head JF, Wang F, Lipari CA, Elliott RL. The important role of infrared imaging in breast cancer. *IEEE Eng Med Bio Mag* May/June 2000; 52–57.
- Jones BF. A reappraisal of the use of infrared thermal image analysis in medicine. *IEEE Trans Med Imaging* December 1998;17(6):1019–1027.
- Keyserlingk J. Time to reassess the value of infrared breast imaging? *Oncology News Int* 1997;6(9).
- Thermology. Available at <http://www.thermology.com/history.htm>.
- Electromagnetic spectrum. Available at <http://www.lbl.gov/MicroWorlds/ALSTool/EMSpec/EMSpec2.htm>. Last updated August 31, 2001.
- HyperPhysics. Available at <http://hyperphysics.phy-astr.gsu.edu/hbase/ems1.htmlc1>.
- Infrared. Available at <http://en.wikipedia.org/wiki/Infrared>.
- Near, mid and far-infrared. Available at <http://www.ipac.caltech.edu/Outreach/Edu/Regions/irregions.html>.
- Thermal imaging. Available at <http://www.ibd.nrc-cnrc.gc.ca/english/specethermal.htm>.
- Whale J. Radiometric thermal diagnostics and dielectric resonance management procedures. <http://www.positive-health.com>.
- Mansfield JR. Tissue viability by multispectral near infrared imaging: a fuzzy C-means clustering analysis. *IEEE Trans Med Imaging* December 1998;17(6):1011–1018.
- Office of Naval Research Press Release. Detecting breast cancer with a new algorithm and a multi-spectral infrared imaging system. Available at <http://www.onr.navy.mil/media/article.asp?ID=14>. September 2002.
- Anbar M. Clinical thermal imaging today. *IEEE Eng Med Bio Mag* July–Aug. 1998;17(4):25–33.
- Bale M. High-resolution infrared technology for soft-tissue injury detection. *IEEE Eng Med Bio Mag* July–Aug. 1998; 17(4):56–59.
- Harding JR. Investigating deep venous thrombosis with infrared imaging. *IEEE Eng Med Bio Mag* July–Aug. 1998;17(4): 43–46.
- Jones BF, Plassmann P. Digital infrared thermal imaging of human skin. *IEEE Eng Med Bio Mag* Nov.–Dec. 2002;21(6): 41–48.
- Ring EFJ. Progress in the measurement of human body temperature. *IEEE Eng Med Bio Mag* July–Aug. 1998;17(4): 19–24.
- Szabo T, et al. Cardiothermographic assessment of arterial and venous revascularization. *IEEE Eng Med Bio Mag* May–June 2000;19(3):77–82.
- Pennes HH. Analysis of tissue and arterial blood temperature in resting human forearm. *J Appl Physiol* 1948;2:93–122.
- Ng EYK, Sudarshan NM. Numerical computation as a tool to aid thermographic interpretation. *J Med Eng Technol* March/April 2001;25(2):53–60.
- Chan CL. Boundary element method analysis for the bioheat transfer equation. *ASME J Heat Transfer*, 1992;114:358–365.
- Hsu TR, Sun NS, Chen GG. Finite element formulation for two dimensional inverse heat conduction analysis. *ASME J Heat Transfer* 1992;114:553–557.
- Kakuta N, Yokoyama S, Mabuchi K. Human thermal models for evaluating infrared images. *IEEE Eng Med Bio Mag* Nov.–Dec. 2002;21(6):65–72.
- Liu Z, Wang C. Method and apparatus for thermal radiation imaging. U.S. Pat. 6,023,637 (2000).
- Qi H, Kuruganti PT, Liu Z. Early detection of breast cancer using thermal texture maps. *IEEE International Symposium on Biomedical Imaging: Macro to Nano*, Washington (DC): 2002. p 309–312.
- Liu Z, et al. Thermal texture maps (ttm): Concept, theory, and applications. In: Di-akides NA, editor. *Biomedical Engineering Handbook*. Vol. Infrared Imaging Section. New York: CRC Press; 2005.
- Fujimasa I. Pathophysiological expression and analysis of far infrared thermal images. *IEEE Eng Med Bio Mag* July–Aug. 1998;17(4):34–42.

37. Hay GA. *Medical Image: Formation, Perception and Measurement*. New York: America, The American Institute of Physics and Wiley; 1976.
38. Watmough DJ. The role of thermographic imaging in breast screening. discussion by C R Hill. *Medical Images: Formation, perception and measurement 7th L H Gray Conference: Medical Images*. 1976; pp. 142–158.
39. Gautherie M. *Atlas of breast thermography with specific guidelines for examination and interpretation*. Milan. Italy: PAPUSA; 1989.
40. Ng EYK, et al. Statistical analysis of healthy and malignant breast thermography. *J Med Eng Technol* November/December 2001;25(6):253–263.
41. Alexjander S, Deamer D. The infrared frequencies of DNA bases: Science and art. *IEEE Eng Med Bio Mag* March–April 1999;18(2):74–79.
42. Pavlidis I, Levine J. Thermal image analysis for polygraph testing. *IEEE Eng Med Bio Mag* Nov.–Dec. 2002;21(6):56–64.
43. Opgal Optronics Industries Ltd. FLIR cameras. Available at <http://www.opgal.com/tefir.htm>. 2001.
44. Opgal Optronics Industries Ltd. EYE Z640 cooled FLIR. Available at <http://www.opgal.com/z640.htm>, 2001.
45. Infrared Solutions. Historical facts of the microbolometer technology. Available at <http://www.infraredsolutions.com/html/technology/historicalFactsF.shtm>. 2005.
46. DARPA MTO Program. Baa99-30: Uncooled thermal imaging sensors. *Commerce Bus Daily*; July 12 1999.
47. Bioyear, Inc. Prism 2000 thermal metabolic imaging system. Available at <http://www.bioyear.net/English/prism.htm>. 2000.
48. Wiecek B. Advances in infrared technology—quantum well versus thermal detectors. *Proc Second Joint EMBS/BMES Conf*. Vol. 2. 2002; p 1135–1136.
49. Infrared glossary of terms—xenics. Available at <http://www.xenics.com/Products/Glossary.php>. 2004.
50. Snyder WE, et al. Increasing the effective resolution of thermal infrared images. *IEEE Eng Med Bio Mag* May–June 2000;19(3):63–70.
51. Braunstein M, Chan RW, Levine RY. Simulation of dye-enhanced near-ir transillumination imaging of tumors. *Proc 19th EMBS Annu Int Conf*, Vol. 2, Chicago: 1997; p 735–739.
52. Kaczmarek M, Nowakowski A. Analysis of transient thermal processes for improved visualization of breast cancer using ir imaging. *Proc 25th Annu Int Conf IEEE EMBS*. Vol. 2. 2003; p 1113–1116.
53. Meditherm—digital infrared thermal imaging. Available at <http://www.meditherm.com/thermpage9.htm>, Last Updated: September 3, 2003.
54. Head JF, Lipari CA, Elliott RL. Computerized image analysis of digitized infrared images of the breasts from a scanning infrared imaging system. *Proc 1998 Conf Infrared Tech Appl XXIV, Part I*, Vol. 3436. San Diego SPIE; 1998. p 290–294.
55. Lipari CA, Head JF. Advanced infrared image processing for breast cancer risk assessment. *Proc 19th Int Conf IEEE/EMBS*, Chicago: IEEE: Oct. 30–Nov. 2 1997. p 673–676.
56. Qi H, Head J. Asymmetry analysis using automatic segmentation and classification for breast cancer detection in thermograms. *Proc 23rd Annu Int Conf IEEE EMBS*. Vol. 3, Turkey: IEEE: October 2001. p 2866–2869.
57. Mabuchi K, et al. Evaluating asymmetrical thermal distributions through image processing. *IEEE Eng Med Bio Mag* March–April 1998;17(2):47–55.
58. Kuruganti PT, Qi H. Asymmetry analysis in breast cancer detection using thermal infrared images. In *Proc 2nd Joint EMBS-BMES Conf*. Vol. 2. October 2002; p 1129–1130.
59. Jakubowska T, Wiecek B, Wysocki M, Drews-Peszynski C. Thermal signatures for breast cancer screening comparative study. *Proc 25th Annu Int Conf IEEE EMBS*. Vol. 2. 2003; 1117–1120.
60. Szu H, et al. Lagrange constraint neural net de-mixing enabled multispectral breast imaging. *IEEE EMBS 2002*.
61. Anbar M, et al. Detection of cancerous breasts by dynamic area telethermometry. *IEEE Eng Med Bio Mag* Sept.–Oct. 2001;20(5):80–91.
62. Fujimasa I, Chinzei T, and Saito I. Converting far infrared image information to other physiological data. *IEEE Eng Med Bio Mag* May–June 2000;19(3):71–76.
63. Gulyaev V Yu, Markov GA, Koreneva GL, Zakharov PV. Dynamical infrared thermography in humans. *IEEE Eng Med Bio Mag* Nov.–Dec. 1995;14(6):766–771.
64. Irvine JM. Targeting breast cancer detection with military technology. *IEEE Eng Med Bio Mag* Nov.–Dec. 2002; 21(6):36–40.
65. Paul JL, Lupo JC. From tanks to tumors. *IEEE Eng Med Bio Mag* Nov.–Dec. 2002;21(6):34–35.
66. Amalu WC. A review of breast thermography. Available at <http://www.iact-org.org/articles/articles-review-btherm.html>.
67. AFRL. Medatr database using vdl. Available at <http://www.vdl.af.mil/access/>. 2003.
68. Hassan M, et al. Infrared thermographic imaging for the assessment of temperature asymmetries in reflex sympathetic dystrophy. *Proc 25th Annu Int Conf IEEE Eng Med Biol Soc*, Vol. 2, 2003; p 1102–1105.
69. Hassan M, et al. Non-invasive multi-modality technique to study angiogenesis associated with Kaposi's sarcoma. *Proc Second Joint EMBS/BMES Conf*. Vol. 2. 2002. p 1139–1140.
70. Harrison B, Mabuchi K. Biomedical infrared imaging in Japan. *IEEE Eng Med Bio Mag* July–Aug 1998;17(4):66–70.
71. Hassan M, et al. Noninvasive infrared imaging for quantitative assessment of tumor vasculature and response to therapy. *Proc 26th Annu Int Conf IEEE Eng Med Biol Soc*, Vol 2. 2004; p 1200–1202.
72. Merla A, et al. Infrared functional imaging applied to Raynaud's phenomenon. *IEEE Eng Med Bio Mag* Nov.–Dec. 2002;21(6):73–79.
73. Qi H. Thermal infrared imaging in early detection of breast cancer—a survey of recent research. Available at <http://aici-pe.ece.utk.edu/research/irsurvey.htm>, 2003.

See also MAMMOGRAPHY; THERMOMETRY; X-RAYS, INTERACTION WITH MATTER.

THERMOMETRY

THEODOROS SAMARAS
Aristotle University of
Thessaloniki Department of
Physics
Thessaloniki, Greece

INTRODUCTION

High body temperature has been an indication of illness since the time of Hippocrates, when only the hand could be used to detect heat or cold. The development of thermometers as we know them was a slow process; it was not before the late 19th century that thermometry became an important tool in clinical practice (1). Medical thermometry is routinely used nowadays across the spectrum of medical specialties and in all environments, ranging from the home to the critical care unit. New techniques have

appeared, which allow the measurement not only of the core body temperature, but also of local temperatures or regional temperature distributions, which can be used as health indicators.

Catheterization for thermometry measurements is used in the thermodilution technique for the assessment of the cardiac output and in intracoronary thermography for the prognosis of atherosclerotic plaques. Non-contact infrared thermometry has been used in the screening of travelers for the severe acute respiratory syndrome (SARS), but also in rheumatic diseases, vascular disorders, and the detection of breast cancer. The most extensive use of thermometry, however, is in thermal therapies (hyperthermia and tissue ablation). The physical method used (RF and microwave radiation, lasers or ultrasound), the treatment site, the target volume size, and the anticipated temperature rise determine the temperature measurement technique. In fact, several practical requirements in this area have resulted in the recent technological advancement of medical thermometry.

SPECIFICATIONS OF A TEMPERATURE MEASURING SYSTEM

To achieve a fair comparison among the various temperature measuring systems and to choose the appropriate one for each application, a set of parameters has to be considered. In the following article, these characteristic parameters are defined as closely as possible to the international vocabulary of basic and general terms in metrology (2).

Accuracy is the ability of a measuring system to provide a value of temperature close to its true value. The *resolution* of a temperature measuring system is the smallest change in the value of temperature that causes a perceptible change in the corresponding indication, whereas the *resolution of a displaying device* is the smallest difference between indications of a displaying device that can be meaningfully distinguished. *Stability* is the ability of a temperature probe to maintain its metrological characteristics constant with time.

The concept of *time constant* has been replaced by the *step-change response time of a measuring system*. As a consequence, the response time of a temperature probe is the duration between the instant when the temperature at the input of the system is subjected to a step change between two values and the instant when the corresponding indication settles within 63% of the step change. The factors affecting the response rate of a temperature probe are

1. The mass of the probe surrounding the active temperature sensitive point.
2. The thermal conductivity of materials used in manufacturing the probe (e.g., sheathing, protective, or insulating coating).
3. The mass and conductivity of the measured material.

TYPES OF TEMPERATURE MEASURING SYSTEMS

Liquid-in-Glass Thermometers

Thermal expansion of fluids is a physical phenomenon that can be used for accurate temperature measurements, as is

manifested by the adoption of gas thermometers in the lower ranges of the current international temperature scale, ITS-90 (3). The expansion of solids finds an application not only in thermostats but also in bimetallic thermometers, which are used in industry (furnaces, hot water pipes, vapor chambers). In medicine the liquid-in-glass thermometer has dominated medical thermometry in the last few centuries. However, the use of liquid-in-metal thermometers is not unknown to the chemical industry, as they can present robust and accurate low-cost solutions in a hostile environment.

A liquid-in-glass thermometer comprises a capillary tube sealed at both ends supported in a stem with a suitable scale. At the basis of the tube, a tiny reservoir (bulb) contains the liquid, which expands inside the tube and raises its height, when it gets hot, because glass expands considerably less compared with it. For the liquid to expand without difficulty, the remaining space of the tube is either empty (vacuum) or filled with a compressible gas. The scale on the stem of the thermometer is calibrated in such a way that temperature is proportional to the liquid height. This requirement dictates the use of liquids, for which the volume increases linearly with temperature. Moreover, it implies that the capillary bore has to be of the same diameter. If, due to manufacturing uncertainties, the inner diameter of the capillary tube changes, inaccuracies can occur in temperature measurement.

The choice of the liquid depends on the use of the thermometer and the temperature range of interest. Mercury, alcohol, and some synthetic oils are most commonly employed. Intended use and safety against breakage (e.g., mercury is a toxic liquid) also play a role. The main advantages of liquid-in-glass thermometers are their low cost, user-friendliness, and credibility (their accuracy can be as low as $\pm 0.01^\circ\text{C}$, or even half of this value for laboratory thermometers). Their drawbacks can be summarized in their low resolution, which strongly depends on the operator, their slow response, and their fragility, which calls for an environment free of vibrations. Another disadvantage is that they have to be read locally, because they have to be in good contact with the measured medium (sometimes even fully immersed in it) and, in some situations, their size. Medical applications fall within the temperature range of both mercury (from about -30 to 500°C) and alcohol thermometers (from about -80 to 70°C).

Further details on the use of liquid-in-glass thermometers can be found in Reference 4.

Electrical Resistance Thermometers

The change of resistance with temperature in conductors is related to changes in free electrons' motion and atomic lattice vibrations. In fact, any conductor could be used in principle to build a resistance temperature detector (RTD). However, manufacturing limitations have led to the choice of specific metals, like copper, gold, nickel, silver, and platinum. The increase of resistance with temperature in metals is often expressed in the form

$$R_T = R_0(1 + \alpha T) \quad (1)$$

where R_0 is the conductor resistance at temperature of 0°C , T_R is the conductor resistance at temperature of T (in degrees Centigrade) and $\alpha(^{\circ}\text{C}^{-1})$ is the *temperature coefficient of resistance*, which is characteristic for the metal.

The accuracy of RTDs is very high, especially when constructed with metals like platinum, for which a high degree of purity can be achieved. This is why platinum resistance thermometers are used in defining ITS-90 in the range between the triple point of hydrogen (13.8033 K) and the freezing point of silver (1234.93 K), which comprises the biological temperature range. However, the characteristic curve of resistance against temperature is usually modeled as a higher order polynomial for standardization purposes. Assuming a linear relationship, like above, results in an error smaller than 0.4°C at 50°C in the temperature range 0 to 100°C (5).

The temperature sensor of an RTD consists of a wire wound on a ceramic core or a thick film coated on a ceramic surface. The sensor is encapsulated within a ceramic casing to form the temperature probe. The two ends of the wire are connected to a Wheatstone bridge (Fig. 1). The value of resistance R_V is varied until the indication on the digital voltmeter is zeroed, i.e., until it matches the unknown resistance of the temperature sensor. Therefore, the bridge imbalance, which indicates that resistance changes can be readily calibrated to reflect temperature changes by assuming a linear relationship. One problem with the circuit of Fig. 1 is that lead resistance also changes with temperature. Therefore, other arrangements of the lead wires should be implemented, which directly measure and subtract this latter resistance from the sensor resistance. Inaccuracies in RTDs can occur due to self-heating, because the current, which must be passed through the sensor to measure its resistance, causes ohmic heating. This kind of error can be minimized by reducing the flowing current and ensuring a good thermal contact between the sensor and the surrounding medium. As mentioned, RTDs can be very accurate. However, they have a large time constant and are relatively large in size for biological implantation. Moreover, they are fragile and have a high cost.

Thermistors

The principle of resistance changes with temperature is also used in thermistors. These were introduced to over-

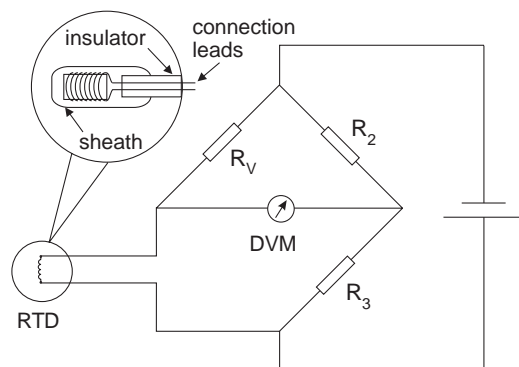


Figure 1. Schematic diagram of an RTD measuring circuit.

come the very small temperature coefficients of the cheaper metals used in RTDs. Thermistors are based on semiconductors, fabricated by mixing metal oxides (usually of manganese, nickel, chromium, and cobalt). The alteration of semiconductor resistance with temperature is founded on a different physical principle as compared with conductors. The exchange of electrons (and holes) from the bound valence band to the mobile conduction band depends exponentially on temperature, giving resistance changes of typically $1\ \Omega/0.01^\circ\text{C}$, whereas RTDs operate with changes on the order of $5\ \text{m}\Omega/^\circ\text{C}$.

Thermistors can present both reduction and increase of resistance with temperature. The change of resistance can be approximated by the relationship

$$R_T = R_0 e^{B(\frac{1}{T} - \frac{1}{T_0})} \quad (2)$$

where R_0 is the resistance at temperature T_0 and B is a constant, which characterizes the thermistor material. The values of B are in the range of 3000 to 5000 K. Thermistors are specified by their resistance at 25°C , which ranges from several ohms to some kilohms.

They are manufactured in the shape of beads, disks, or rods with conducting leads attached to them and encapsulated in epoxy resin or glass sheaths. Beads are used most of the time; they are available in very small sizes with a diameter less than 0.1 mm. They have response times in the range of some seconds and an accuracy of ± 0.1 to $\pm 0.5^\circ\text{C}$.

A noteworthy development in thermistor fabrication took place in the mid-1970s, when R. Bowman introduced an elegant design of a temperature probe, which was highly insensitive to radio-frequency electromagnetic radiation (6). The probe had an outer diameter of 1 mm, with a thermistor of 0.5 mm in size, and a response time of 0.2 s. It was based on high-resistance, plastic readout-leads with resistances of about $160\ \text{k}\Omega/\text{cm}$, which secured a heating error in the lines of less than 0.005°C for a heating rate of $1^\circ\text{C}/\text{min}$.

The use of standard photolithography has made possible the creation of miniaturized thermistor probes. For example, the evaporation of amorphous germanium (a-Ge) on a Pyrex glass substrate has resulted in a temperature sensing area of $50\ \mu\text{m} \times 100\ \mu\text{m} \times 0.25\ \mu\text{m}$ with large resistance ($4\ \text{M}\Omega$) that needs only a 20 nA measuring current (7). The accuracy of this probe is close to 0.05°C in the temperature range 0 to 60°C and its response time only 14 ms.

Self-heating is a source of inaccuracy for thermistors, as it is for RTDs. However, the high resistivity of thermistor materials can allow for reductions in the measuring current, down to values that ensure the desired resolution. The high resistivity offers another advantage, namely that of eliminating the need for complicated circuitry configurations, because variation of lead resistance with temperature becomes less important. Thermistors need frequent calibration and may show a drift in their characteristics due to changes of the semiconductor materials. Their main drawback is their nonlinearity.

Thermocouples

The operation of thermocouples is based on the *Seebeck effect* or *thermoelectricity*, or the ability of heat energy to

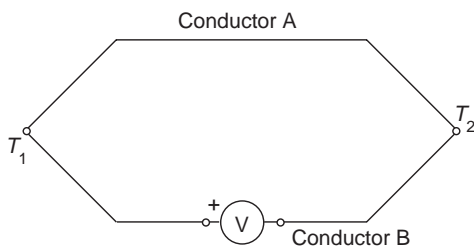


Figure 2. The Seebeck effect (thermoelectricity) is the physical principle for the operation of a thermocouple.

free electrons from a metal's surface into the free state. When a pair of dissimilar metals is connected into the form of a loop and the junctions are kept at different temperatures, an electromotive force (emf) develops (Fig. 2). If the two junctions have the same temperature, then no net emf will appear. Therefore, if one junction is kept at a constant temperature and the temperature of the other is varied, the measured emf will be proportional to the temperature difference of the two junctions.

Many thermocouple material combinations are listed by Kinzie (8). The combination determines the magnitude and the polarity of the resulting emf. The criteria for selecting a thermocouple can include cost, temperature range, chemical stability, physical properties of the measured medium, and duration of measurement. The most commonly used metals for the construction of thermocouples are rhodium, copper, iron, nickel, chromium, and aluminum, as well as some of their alloys. Some metal combinations are standardized and designated by a letter (*T, J, E, K, N, B, S, R*).

Thermocouples of the *T, J, E,* and *K* types operate in temperature ranges, which include the temperatures of biological and medical interest. The first three types are assembled with constantan (an alloy of nickel and copper) as one of the metals. Although they have lower stability than other thermocouple types, they are inexpensive and show good linearity and moderate and sensitivity (30 to 50 $\mu\text{V}/^\circ\text{C}$) in the biological temperature range. Their dimensions can be very small. Two metal wires of some microns in diameter (70 to 140 μm in practice) are insulated (in PVC, Teflon, or glass fiber) and connected at their distant end (Fig. 3) to form the measuring (primary) junction, which is in the size of the two wires combined. The small

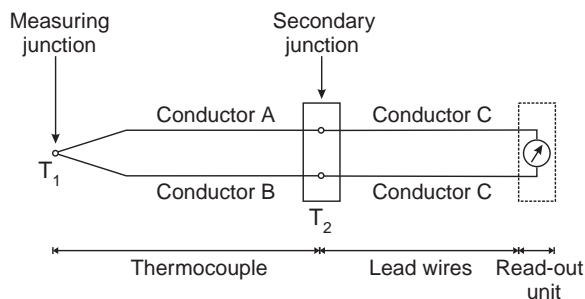


Figure 3. Schematic diagram of a thermocouple measuring circuit.

size of the wires and junctions allows for fast response times, especially when encapsulated in a thin sheath. Furthermore, it facilitates the introduction of thermocouples inside a hypodermic needle for interstitial temperature measurements more easily compared with thermistors and RTDs.

To operate a thermocouple, it is necessary to connect it to a voltage-measuring instrument. However, the connection to the readout module is usually achieved through the use of leads, which are made of a metal, different from the ones that make up the thermocouple. In the case of a type *T* thermocouple and copper leads, there is going to be one junction, namely that of thermocouple constantan to copper lead, which will introduce one more thermocouple junction. The advantage of the type *T* thermocouple is clear; for other types of thermocouples, for which none of the metals is copper, two new thermocouple junctions will be created at the connection of terminals to the readout module, whose emfs will be in series with the emf of the measuring junction. Therefore, the temperature of these secondary (Fig. 3) junctions (also called *reference* or *cold junctions*) must be stable and their emfs known to maintain the calibration of the primary junction. Historically, this used to be accomplished by immersing the reference junctions in an ice bath of 0°C , so that the total measured voltage was adjusted to zero when the probe temperature was also 0°C . A more convenient solution, however, is to provide an electronic bridge circuit between the secondary junction and the voltage measuring equipment. This circuit incorporates a resistance temperature device, whose voltage changes with temperature by the same amount as the reference junction, canceling out any variations of the latter (the temperature-sensitive device is connected in such a way that its voltage is subtracted from the emf appearing at the secondary junction).

Sources of uncertainty in temperature measurements with thermocouples include the spurious emfs from external electric and magnetic fields, temperature measurement of reference junctions, cable specifications, and drift and uncertainty of the voltage measuring instrument. Like in the case of RTDs and thermistors, thermal conduction along the metal readout wires gives another source of inaccuracy.

Interesting and practical information on thermocouple thermometry can be found in References 8 and 9.

Fiber-Optic Probes

The use of fiber-optic thermometers is necessary in situations, where electrical insulation for safety or electromagnetic immunity of the sensor is of concern. The most common medical applications for which these requirements are of paramount importance are cancer treatment with microwave or RF hyperthermia, temperature monitoring during Magnetic resonance imaging (MRI) and cardiac output measurement with the thermodilution technique. There are mainly three reasons for which conventional thermoelectric devices (RTDs, thermistors or thermocouples) should not be used inside electromagnetic fields:

1. The incident electromagnetic fields will be perturbed and scattered by the metal parts of the thermometer devices.
2. Currents will flow in the metal parts of the devices, resulting in their ohmic heating.
3. The currents in the devices can lead to spurious readings.

Fiber-optic sensors work on the principles of light absorption, reflection, scattering or interference, as well as with the effect of induced fluorescence. With respect to the implementation of the physical mechanisms, they either operate in the time domain or they involve intensity or wavelength modulation.

The simplest solution to temperature measurement with a fiber-optic probe is the use of a gallium arsenide (GaAs) crystal as the sensor. One implementation uses two optical fibers, a transmitting and a receiving one. The light transmitted by a light-emitting diode, after having been partially absorbed in the GaAs sensor at the tip of the probe, returns to the detecting module in the readout equipment. It is known that some of the light energy that gets absorbed in the crystal is used to raise electrons from the valence band to the conduction band. As the energy gap between the two bands is a known function of the crystal's temperature, the amount of absorbed power can be related to the temperature of the GaAs sensor. A second implementation with semiconductor sensors uses the same crystal and a dielectric mirror (Fig. 4) at the end of a single optical fiber and takes advantage of wavelength, instead of intensity, modulation due to temperature variations. The transmission spectrum of the crystal moves to larger wavelengths as its temperature rises (Fig. 5). This is known as the absorption/transmission shift and occurs because the energy of a photon is inversely proportional to its wavelength. When the temperature of the GaAs crystal increases, reducing the energy gap between the semiconductor's electron state bands, photons with less energy (longer wavelengths) are absorbed, making the transmission spectrum shift toward higher wavelengths. The advantages of this implementation include that the readout is independent of light intensity and, consequently, factors, which usually contribute to the attenuation in optical fibers, such as length, splices, and bending.

One of the first techniques that have been commercialized for fiber-optic probes is that of induced fluorescence. This technique makes use of the change in fluorescence decay time (lifetime) with temperature. At the tip of the probe, a thermosensitive phosphor sensor is located. This

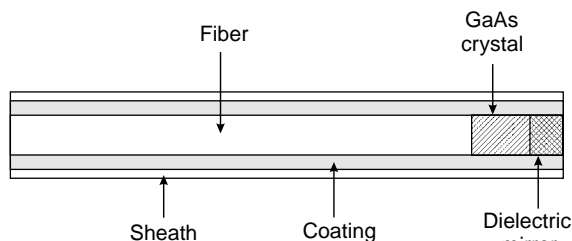


Figure 4. Diagram of a fiber-optic probe.

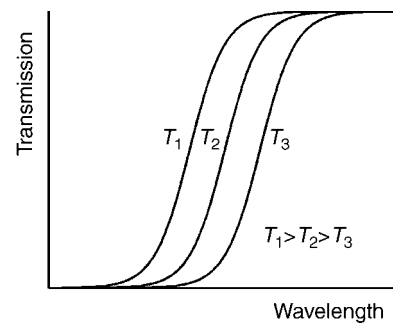


Figure 5. The absorption/transmission shift with changing temperature in semiconductors.

sensor is stimulated by the red light of a pulsed light-emitting diode and emits light over a broad spectrum in the near infrared region. The time needed for this fluorescence effect to decay depends on the temperature of the phosphor sensor; the higher the temperature, the shorter the decay time. The fluorescent signal continues to be transmitted for some milliseconds through the same fiber back to a detector in the readout equipment, even after the stimulating light from the led is off. The fluorescent decay time is then measured by a multipoint digital integration of the decay curve. The use of the decay-time technique eliminates the need to reference the output light intensity to that incident on the sensor, a significant problem experienced in intensity modulation schemes.

Another approach to temperature measurement with fiber-optic probes is Fabry–Perot interferometry, in which two parallel, partially reflecting surfaces are brought very close to each other to form an optical reflecting cavity, also known as etalon. If the distance between these surfaces (due to different coefficients of thermal expansion of the materials they are made of) or the refractive index of the material between them changes, the reflectance spectra and the interference fringes change accordingly.

Apart from electromagnetic immunity, other advantages of fiber-optic probes include minimal thermal conduction along the probe and small size, down to 0.5 mm in diameter. Their temperature range is very wide, their accuracy can reach 0.1 °C, and their response time is in the order of several hundred milliseconds.

MRI Thermometry

The use of minimally invasive surgery is very attractive due to low costs and high effectiveness. It is usually conducted in the guidance of ultrasound imaging or MRI. Thermal therapies combined with MRI have gained recognition in the recent years and include laser-induced thermotherapy (LITT), RF ablation, hyperthermia, and focused ultrasound. In these techniques, it is not recommended to use temperature sensors with metal parts, due to the reasons described above. Instead, the signals collected to reconstruct the image in the MRI devices can be used to create temperature maps inside the patients in three dimensions.

The effect of temperature on physical parameters measured by MRI devices has been known for a long time; the first study on temperature measurement with such a

device appeared in 1983 (10). This first report was based on changes of the longitudinal relaxation time (T_1). In a space free of magnetic fields, the magnetic orientations of atomic nuclei in a biological sample are directed randomly. Once placed into a magnetic field, these nuclei take two different preferred orientations, aligned with or against the external magnetic field, and the sample becomes magnetized. The transition from the random distribution of the unmagnetized sample to a magnetized state requires some exchange of quantized energy. The process of magnetization is exponential, and the rate at which a sample becomes magnetized is characterized by a quantity known as the *longitudinal relaxation time* (T_1):

$$\vec{M}_z = \vec{M}_0(1 - e^{-t/T_1}) \quad (3)$$

If the availability of the exact energy required to flip the nuclei between their two energy states is low, then T_1 will be long. The energy for changing the magnetic state of nuclei is obtained from molecular motion, and thus, it depends on temperature. At absolute zero, where there is no molecular motion, T_1 approaches an infinite value. In fact, for the hydrogen protons, which are in abundance in a biological sample, the spontaneous state change would take place once every about 10^{25} years. Smaller molecules, like water, exhibit a great deal more motion than larger ones, although one should keep in mind the difference between the freely moving bulk water in tissues and that bound at surfaces of proteins and membranes, which is less mobile. The change in T_1 with temperature can be described by

$$T_1 = T_1(\infty)e^{-E_a/kT} \quad (4)$$

where E_a is the activation energy of the longitudinal relaxation process, k is the Boltzmann constant, and T is the temperature (K).

Temperature measurement with T_1 , using conventional sequences in MRI, suffers from long acquisition times, which would render the technique insufficient for most thermal procedures. Consequently, other sequences are used, like RF-spoiled gradient echo imaging, where the changes in T_1 can be assessed faster, but with limited signal-to-noise ratio (SNR). Another major problem of T_1 temperature measurements is the errors that occur from movement, which call for high-quality image registration.

Meanwhile, new measurement techniques have emerged based on two other parameters known to be affected by temperature, namely the diffusion coefficient and the proton resonance frequency (PRF) of tissue water. It is known that the diffusion coefficient D depends exponentially on temperature, because it relates the random Brownian motion of molecules with the diffusion process:

$$D \approx e^{-E_a/kT} \quad (5)$$

where E_a is the activation energy for water diffusion in this case. The problem with diffusion temperature mapping is that the motion of water in tissues is strongly dependent on the existence of biological structures, e.g., membranes. The permeability of these structures is dependent on temperature, making the diffusion process

nonlinear. Moreover, biological structures of larger dimensions, e.g., muscle fibers or myelin sheaths, lead to the anisotropic diffusion of water, which leads in a tensor for the diffusion coefficient, requiring the determination of its nine elements and long acquisition times.

The most popular technique for MRI temperature measurements is based on the PRF of water. The local magnetic field B_l , perceived by the hydrogen protons in a biological sample, relates to the main magnetic field of the MRI device B_0 , with the following equation:

$$B_l = (1 - \sigma_t)B_0 + \delta B_0 \quad (6)$$

where σ_t is the total screening constant of the proton and δB_0 includes all local deviations from B_0 , which are not temperature dependent. As the total screening constant increases linearly with temperature and the phase distribution in an image volume (slice) depends on the local field distribution, it is possible to obtain temperature information by directly subtracting phase images. According to the Larmor equation, the phase within a volume at temperature T is given by

$$\varphi(T) = \gamma T_E[(1 - \sigma_t(T))B_0 + \delta B_0] \quad (7)$$

where γ is the gyromagnetic ratio and T_E is the echo time of the gradient echo pulse sequence, which is used for the acquisition of the phase images. Therefore, the phase difference between two images at two different temperatures is

$$\begin{aligned} \Delta\varphi &= \varphi(T') - \varphi(T) = \gamma T_E[\sigma_t(T') - \sigma_t(T)]B_0 \\ &= \gamma T_E \alpha \Delta T B_0 \end{aligned} \quad (8)$$

where α is the proportionality constant of linear temperature dependence of σ_t , the value of which has been measured at approximately 0.01 ppm/°C (11). The main advantage of the PRF method is its independence of tissue composition. A potential artifact may arise from the presence of lipids, because the PRF of lipid hydrogen protons is independent of temperature. Nevertheless, lipids can be suppressed in gradient echo imaging by frequency-selective slice excitation. It is clear from equations 7 and 8 that the method is sensitive to changes in δB_0 between image acquisitions, due to drift of the external field, movement of the measured object, or change of magnetic susceptibility.

Comparisons among the three methods of MRI thermometry have shown that the PRF method is the one with higher precision (12). An accuracy of 0.5 °C with an estimated resolution of 0.3 °C could be reached in a heterogeneous human phantom (13).

Radiation Thermometry

The random agitation of electrical loads or dipoles in matter at a temperature above absolute zero is associated with the generation of an electromagnetic wideband noise signal with a spectrum extending from RF waves to gamma rays. *Infrared thermography*, i.e., the recording of the temperature distribution of the body using the IR radiation emitted from its surface, and some millimeters beneath it, exploits the wavelengths between 0.8 and

20 μm . *Microwave radiometry* uses the microwave energy emitted at larger wavelengths ranging from 1 mm to 1 m; this means that temperature is collected from a depth of some centimeters inside the body, requiring a huge amount of data to solve the inverse problem formulated (14).

By definition, a *black body* absorbs all radiation incident on it. Yet, if it is at an absolute temperature T , it emits electromagnetic radiation, which is described by the Planck's radiation law. According to the latter, the power P per unit area emitted into solid angle $d\Omega$ within the frequency bandwidth Δf is given by

$$P = \frac{2h f^3 \Delta f d\Omega}{c^2 [\exp(\frac{hf}{kT}) - 1]} \quad (9)$$

where h is Planck's constant and c is the speed of light. The emissivity ϵ of any surface at a given direction and frequency is then defined as the ratio of the power emitted through the surface to the emissive power through the black-body surface at the same frequency.

The major problem, when measuring temperature with radiation thermometers, is the knowledge of the surface emissivity, which depends on temperature. The techniques for measuring this value are complex and expensive. Apart from the uncertainty in the emissivity value, other sources of inaccuracy include the attenuation of radiation between the target and the thermometer (taking into account humidity and distance), the background radiation present (some of it will be reflected by the measured object) and errors in the radiation detectors. The latter can be either thermal detectors, using the absorbed electromagnetic energy to increase their temperature and sense a change in a physical property, such as resistance or dielectric constant, or photon detectors, which measure the direct effect of incident photons in matter with the excitation of electrons. For example, in a quantum well IR photodetector (QUIP), the incident photons produce electron-hole pairs, which are carried away by an applied voltage, giving rise to a pulse of charge.

Commercial IR cameras have a resolution of about 0.1°C at the biological temperature range and an accuracy of 2% of the temperature reading (in degrees Centigrade). With the use of continuous calibration techniques, the accuracy can reach 0.04°C (15). The most popular and controversial class of radiation thermometers are *infrared ear thermometers* or *infrared tympanic thermometers* (ITTs), which have been in the middle of a debate on both their accuracy (16) and their calibration (17).

More details on radiation thermometry can be found in Reference 18.

Electrical Impedance Tomography

The electrical impedivity (impedance of a unit cube) of tissue decreases by about $1.7\% \text{ }^\circ\text{C}^{-1}$ with increasing temperature, due to changes in ionic mobility (19) inside the intra- and extracellular fluids. Electrical impedance tomography (EIT) gives pictures of the conductivity distribution inside the body. Therefore, it can be used, in principal, to determine any variations induced to conductivity through temperature changes. The results in phantoms and *in vitro* were encouraging (20), despite poor resolution in central

regions, nonuniform sensitivity, and equipment interference.

However, the conductivity change with temperature in the living tissue is more complicated, because it involves temperature-induced changes in the interstitial fluid volume and the cell membrane. As a consequence of thermoregulation, vasodilatation can lead to changes in tissue conductance in the same order of magnitude as the changes due to ionic mobility of electrolytes (21). Nevertheless, *in vivo*, the method correlated with direct temperature measurements within 1.5°C , although large errors ($>5^\circ\text{C}$) did exist (22). The low cost and fast response of EIT are two reasons for which it seems reasonable to continue the effort of improving its application in temperature measurement, in particular for thermotherapy.

Ultrasound

Three methods can be used to estimate with ultrasound temperature changes inside tissues. The first one exploits the time shift of received echoes due to changes in tissue thermal expansion and speed of sound, which result in actual and apparent displacements of scattering regions, respectively. These displacements can be related to changes in temperature $\Delta T(z)$ along the direction of propagation z according to (23)

$$\Delta T(z) = \frac{c_0}{2(a - \beta)} \frac{\delta t(z)}{\delta z} \quad (10)$$

where $t(z)$ is the estimated time-shift at depth z , c_0 is the speed of sound before heating, a is the linear coefficient of thermal expansion, and the coefficient $\beta = (1/c_0)(\delta c/\delta T)$ describes the change in the speed of sound with temperature. It is assumed that the speed of sound varies linearly with temperature up to about 45°C , whereas the term $(a - \beta)$ depends on tissue type. It is clear that the limitation of the above method is the requirement for prior knowledge of the speed of sound and thermal expansion coefficients.

The second method is based on the changes of ultrasound attenuation with temperature, which, however, are more pronounced at temperatures larger than 50°C . Therefore, this technique is a good candidate for temperature monitoring in thermal ablation. The third method makes use of the changes on backscattered energy, which could be as much as 5 dB over the temperature range from 37 to 50°C for individual scatterers (24).

Although temperature measurements by ultrasound are a convenient and inexpensive alternative to MRI, the performance of all of the above methods needs to be evaluated *in vivo*, where sophisticated motion tracking techniques have to be employed to correct for tissue movement.

BIBLIOGRAPHY

1. Pearce JMS. A brief history of the clinical thermometer. *QJM Mon J Assoc Phys* 2002;95:251–252.
2. ISO/DGuide 9999, International vocabulary of basic and general terms in metrology (VIM). 3rd ed. 2004
3. Preston-Thomas H. The International Temperature Scale of 1990 (ITS-90). *Metrologia* 1990;27:3–10.

4. Nicholas JV, Liquid-in-glass thermometers. In: Webster JG editor. *The Measurement, Instrumentation and Sensors Handbook*. Boca Raton, FL: CRC Press; 1999
5. Childs PRN, Greenwood JR, Long CA. Review of temperature measurement. *Rev Sci Instrum* 2000;71:2959–2978.
6. Bowman RR. A probe for measuring temperature in radio-frequency-heated material. *IEEE Trans Microwave Theory Techn* 1976;24:43–45.
7. Schuderer J, Schmid T, Urban G, Samaras T, Kuster N. Novel high-resolution temperature probe for radiofrequency dosimetry. *Phys Med Biol* 2004;49:N83–N92.
8. Kinzie PA. *Thermocouple Temperature Measurement*. New York: Wiley; 1973.
9. Kerlin TW. *Practical Thermocouple Thermometry*. Research Triangle Park, NC: ISA International; 1998.
10. Parker DL, Smith V, Sheldon P, Crooks LE, Fussell L. Temperature distribution measurements in two-dimensional NMR imaging. *Med Phys* 1983;10:321–325.
11. Peters RD, Hinks RS, Henkelman RM. Ex vivo tissue-type independence in proton-resonance frequency shift MR thermometry. *Magn Reson Med* 1998;40:454–459.
12. Wlodarczyk W, Hentschel M, Wust P, Noeske R, Hosten N, Rinneberg H, Felix R. Comparison of four magnetic resonance methods for mapping small temperature changes. *Phys Med Biol* 1999;44:607–624.
13. Gellermann J, Wlodarczyk W, Ganter H, Nadobny J, Föhlhling H, Seebass M, Felix R, Wust P. A practical approach to thermography in a hyperthermia/magnetic resonance hybrid system: Validation in a heterogeneous phantom. *Int J Radiat Oncol Biol Phys* 2005;61:267–277.
14. Leroy Y, Bocquet B, Mamouni A. Non-invasive microwave radiometry thermometry. *Physiol Meas* 1998;19:127–148.
15. Baker JM, Noman JM, Kano A. A new approach to infrared thermometry. *Agr Forest Meteorol* 2001;108:281–292.
16. Craig JV, Lancaster GA, Taylor S, Williamson PR, Smyth RL. Infrared ear thermometry compared with rectal thermometry in children: A systematic review. *Lancet* 2002;360:603–609.
17. Pušnik I, van der Ham E, Drnovšek J. IR ear thermometers: What do they measure and how do they comply with the EU technical regulation? *Physiol Meas* 2004;25:699–708. Erratum in *Physiol Meas* 2004;25:1337.
18. DeWitt DP, Nutter GD, editors. *Theory and Practice of Radiation Thermometry*. New York: Wiley; 1988.
19. Brown BH. Electrical impedance tomography (EIT): A review. *J Med Eng Technol* 2003;27:97–108.
20. Conway J, Hawley M, Mangnall Y, Amasha H, van Rhooon GC. Experimental assessment of electrical impedance imaging for hyperthermia monitoring. *Clin Phys Physiol Meas* 1992; 13(Suppl A):185–189.
21. Gersing E. Monitoring temperature-induced changes in tissue during hyperthermia by impedance methods. *Ann N Y Acad Sci* 1999;873:13–20.
22. Paulsen KD, Moskowitz MJ, Ryan TP, Mitchell SE, Hoopes PJ. Initial in vivo experience with EIT as a thermal estimator during hyperthermia. *Int J Hyperthermia* 1996;12:573–591. discussion 593–594.
23. Arthur RM, Straube WL, Trobauch JW, Moros EG. Non-invasive estimation of hyperthermia temperatures with ultrasound. *Int J Hyperthermia* 2005;21:589–600.
24. Straube WL, Arthur RM. Theoretical estimation of the temperature dependence of backscattered ultrasonic power for noninvasive thermometry. *Ultrasound Med Biol* 1994;20: 915–922.

See also HYPERTHERMIA, INTERSTITIAL; HYPERTHERMIA, SYSTEMIC; HYPERTHERMIA, ULTRASONIC; NONIONIZING RADIATION, BIOLOGICAL EFFECTS OF; THERMOCOUPLE; THERMISTOR.

THERMOREGULATION. See BIOHEAT TRANSFER.

THERMOTHERAPY. See HEAT AND COLD, THERAPEUTIC.

TISSUE ABLATION

DIETER HAEMMERICH
Medical University of
South Carolina,
Charleston, South Carolina

INTRODUCTION

Tissue ablation (literal translation “removal”) is the destruction of diseased (*pathologic*) body tissue, with the aim to cure a disease. Tissue destruction is achieved by thermal methods, or by application of chemical substances (e.g., ethanol). Thermal methods cause either local heating or cooling of the tissue to lethal temperatures (typically below -40°C , or above 50°C). A number of different physical principles are employed for heating and cooling tissue, such as radio frequency (rf) electric current, microwaves, laser, ultrasound, and cryogenic cooling.

Current clinical applications include treatment of heart arrhythmia, cancer (liver, lung, brain, kidney, bone, prostate), uterine bleeding, varicose veins, and enlarged prostate (benign prostate hyperplasia), with other emerging applications (see Table 1 for an overview). Typically, an applicator is introduced under imaging guidance [ultrasound imaging, fluoroscopy, computerized tomography (CT), magnet resonance imaging (MRI)] into the body, to the treatment site. Then the tissue region around the applicator is ablated, destroying the diseased tissue area.

This article describes physical principles, clinical devices, and applications of the different ablation modalities.

PHYSICAL PRINCIPLES OF TISSUE ABLATION

Thermal Ablation Methods

All thermal ablation methods rely on thermal conduction to some extent to heat or cool a region of tissue near the applicator. The problem of thermal ablation can mathematically be described by following heat-transfer equation:

$$\rho c \frac{\partial T}{\partial t} = \nabla \cdot k \nabla T + Q_A - Q_P \quad (1)$$

Energy Q_A ($\text{W}\cdot\text{m}^{-3}$) is applied to the tissue by the applicator, resulting in heating (or cooling) of the tissue. Some energy Q_P is carried away by blood perfusion; depending on tissue type, perfusion may be negligible (e.g., heart), or may have major impact (e.g., liver). The left-hand side of equation 1 describes how local tissue temperature T changes, depending on tissue density ρ and tissue specific heat c . The first term on the right-hand side describes how thermal energy is conducted through the tissue, depending on thermal conductivity k . Different models are available in the literature that approximate tissue perfusion, the

Table 1. Frequency of Use of Different Ablation Modalities in Clinical Applications^a

Ablation Modalities	rf	Cryo	Microwave	Ultrasound	Laser	Chemical
Cardiac ablation	<i>b</i>	<i>c</i>	<i>d</i>	<i>d</i>	N/A ^e	N/A ^e
Tumor ablation (liver, lung, kidney)	<i>b</i>	<i>c</i>	<i>d</i>	N/A ^e	<i>d</i>	N/A ^e
Endometrial	<i>c</i>	<i>c</i>	<i>d</i>	N/A ^e	N/A ^e	N/A ^e
Prostate (cancer, enlarged prostate)	<i>d</i>	<i>c</i>	<i>c</i>	<i>d</i>	N/A ^e	N/A ^e
Intervertebral disk	<i>d</i>	N/A ^e	N/A ^e	N/A ^e	<i>d</i>	N/A ^e
Endovascular	<i>c</i>	N/A ^e	N/A ^e	N/A ^e	<i>c</i>	N/A ^e
Cornea	<i>c</i>	N/A ^e	N/A ^e	N/A ^e	<i>b</i>	N/A ^e

^aAdapted from Jie Zhang, University of Wisconsin-Madison.

^bUsed frequently.

^cUsed sometimes.

^dUsed rarely/under investigation.

^enot applicable, not used clinically.

most widely used being Pennes' Bioheat equation (1). In the Bioheat equation, blood perfusion is modeled as a distributed heat sink term, according to

$$Q_p = \rho_{bl} c_{bl} w_{bl} (T - T_{bl}) \quad (2)$$

where ρ_{bl} ($\text{kg}\cdot\text{m}^{-3}$), c_{bl} [$\text{J}(\text{kg}\cdot\text{K})^{-1}$] and T_{bl} are density, specific heat, and temperature of blood, respectively. The parameter T is the tissue temperature, and w_{bl} is the blood perfusion ($1\cdot\text{s}^{-1}$).

Figure 1 shows the heat transfer problem on the example of cardiac radio frequency (rf) ablation.

For thermal ablation methods the resulting zone of tissue death is usually called *thermal lesion*, or *coagulation zone* (for heat-based methods). In cardiac ablation literature, *lesion* is the accepted term. The term lesion should be avoided in tumor ablation applications (though it is used, especially in earlier literature), since lesion is a general medical term referring to a pathological part of tissue (e.g., often tumors are called lesions).

Following each of the different ablation principles will be described. A comparison between thermal ablation methods is given in Table 2.

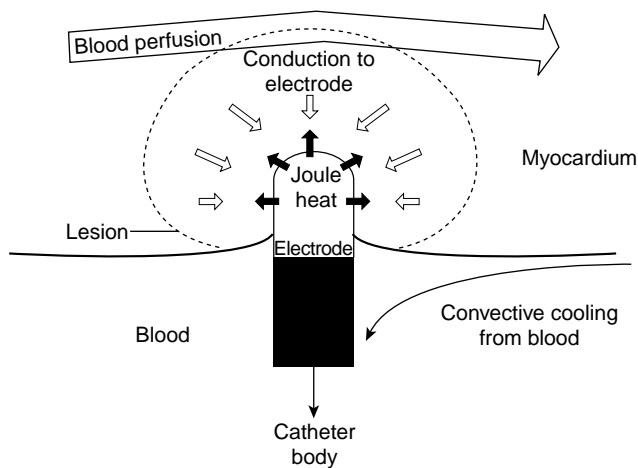


Figure 1. Heat transfer during cardiac rf ablation. Tissue close to the electrode is heated by resistive heating from rf current. Heat is then conducted thermally into the tissue. Heat is lost due to blood perfusion in the myocardium, due to thermal conduction through the electrode, and from convective cooling due to blood flow in the heart chamber. (From Ref. 2.)

Radio Frequency Ablation. Radio frequency ablation is the currently most widely used ablation method. The rf current in the frequency range of typically 450–500 kHz is applied to the tissue via the inserted electrode (catheter); a dispersive electrode (grounding pad) is required to serve as current return path. Inside the tissue electric current is carried by ions (mainly sodium, potassium, and chloride), that is, the ions oscillate due to the alternating electric field. The ion movement results in resistive heating (joule heating) of the tissue due to ionic friction, with most heat generated at the location of maximum rf current density (typically closest to the catheter). One parameter that is often used to characterize electrode performance is the specific absorption rate (SAR), expressed in units of $\text{W}\cdot\text{kg}^{-1}$. The SAR determines how much mass related power is deposited at a certain location in the tissue. To determine the SAR resulting from a specific rf electrode, we first have to solve the electric field problem to determine the electric field strength E in the tissue. The SAR can then be calculated from the local version of Ohm's law according to

$$\text{SAR} = \frac{\sigma}{\rho} |E|^2 = \frac{1}{\sigma \cdot \rho} |J|^2 \quad (3)$$

where σ is the electrical tissue conductivity, ρ is the tissue density, and J is the electric current density.

Figure 2 shows the tissue temperature distribution around a cardiac rf catheter after 45 s.

Maximum tissue temperature during rf ablation is limited to $\sim 110^\circ\text{C}$; above that temperature tissue water vaporizes, forming an electrically insulating barrier preventing further energy deposition. If too much rf power is applied, tissue around the electrode can char; this area of carbonization is electrically insulating and irreversible. Applied rf power therefore has to be controlled to prevent tissue charring (carbonization). In cardiac applications, maximum temperature is further limited to prevent tissue cavitation from rapidly expanding vapor.

There are three control methods currently used in clinical rf devices:

Power Control. Applied rf power is kept constant throughout the ablation procedure.

Temperature Control. The ablation catheter has one or more thermal sensors (thermocouples or thermistors) embedded in the electrode tip, or at a specified distance

Table 2. Comparison of Heat-Based Ablation Modalities

Modality	Advantages	Disadvantages
Radio frequency	Simple applicator design	Limited by tissue charring Dispersive electrode (ground pad) required Not usable under MRI (rf interference)
Microwave	High tissue temperatures Short application times Constructive interference of microwaves from multiple applicators	
Ultrasound	Directional applicators possible Can be used noninvasively	
Laser	Simple applicator design	Limited by tissue charring
Cryo	Iceball visible under ultrasound imaging Reversible tissue damage (for short application times)	

from the catheter. Applied rf power is controlled to keep the measured temperature constant.

Impedance Control. Applied rf power is controlled depending on tissue impedance, as measured between the active electrode and the grounding pad. Initially, impedance drops since electrical tissue conductivity increases with temperature due to higher ion mobility. As tissue vaporizes, an increase in impedance results. When impedance exceeds a certain threshold, rf power is temporarily shut down to allow vapor to settle, and then reapplied at a lower level. Figure 3 shows a typical time course of impedance during an impedance-controlled ablation procedure.

Microwave Ablation. Microwaves (MW) are electromagnetic (EM) waves in the frequency spectrum from 300 MHz to 300 GHz. During MW ablation, a MW antenna is inserted into the tissue, radiating microwaves into the tissue. These EM waves cause polar water molecules in the tissue to align with the applied alternating electric field. The resulting rotating water molecules cause frictional heating of the tissue. For MW ablation, microwaves

at frequencies of 915 MHz or 2.45 GHz are used due to Federal Communications Commission (FCC) restrictions, with wavelengths in the centimeter range inside the tissue. To determine the SAR during MW ablation, first the Maxwell equations have to be solved to determine the electric field distribution in the tissue. Then the SAR can be calculated using equation 3; note that the tissue conductivity σ is strongly dependent on frequency. Since the propagation of microwaves is not hindered by vapor or charred tissue, much higher tissue temperatures compared to rf (up to 150 °C) can be obtained (3).

Different antenna types have been proposed. The dipole antenna has been commonly used (see Fig. 4); other common antenna types are slot antennas and monopoles (4,5). Note that the SAR of MW antennas is dependent on insertion depth of the antenna, and it is significantly different at smaller insertion depths.

Contrary to other ablation methods, there has not been use of any advanced control methods to adjust applied power during MW ablation. So far, constant power (typically 40–100 W, depending on application) has been used. For MW ablation, impedance match between antenna and tissue is important. If there is mismatch in impedance, significant amounts of power are reflected, resulting in

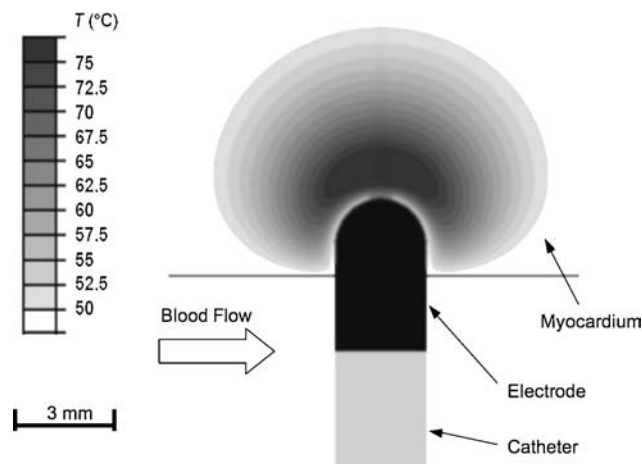


Figure 2. Tissue temperature (computer simulation) after 45 s of cardiac rf ablation. Catheter is 2.3 mm in diameter, with 5 mm electrode length, and inserted 2 mm into tissue. Note that location of maximum temperature is ~ 1 mm distant from the electrode due to electrode cooling from blood flow in the heart chamber. Outer-most boundary (50°C) marks the thermal lesion boundary.

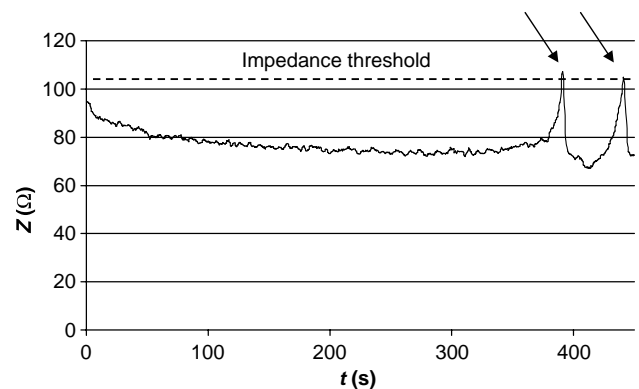


Figure 3. Typical time course of impedance Z (measured between electrode and grounding pads) during rf ablation. Initially, impedance decreases as electrical tissue conductivity increases from heating. When tissue starts to vaporize around the electrode (arrows), impedance rises. When impedance exceeds threshold (dotted line), power is shut down for 15 s, and then reapplied.

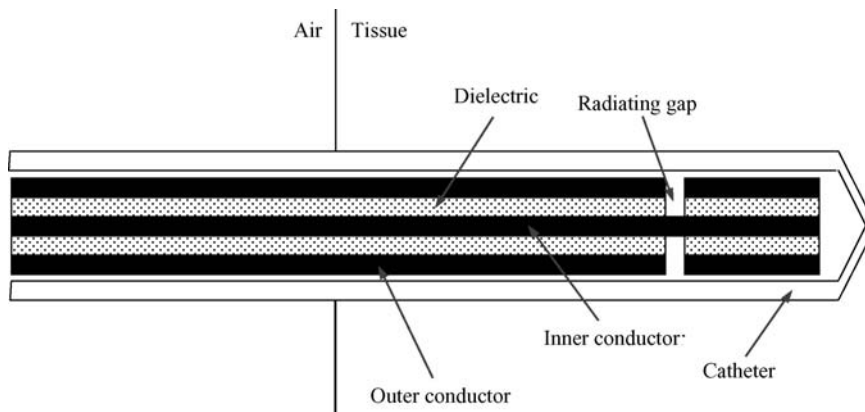


Figure 4. Cross-section of a dipole antenna for microwave ablation. Microwaves are radiated from the gap into the tissue.

undesirable heating of the antenna shaft and cable. Proper matching is complicated by the fact that tissue properties change significantly during heating.

Antenna Arrays. For all thermal ablation methods, the SAR is only significant very close to the applicator; most of the tissue heating is caused by thermal conduction. This can be a disadvantage when large tissue volumes need to be heated, or tissue close to large blood vessels is heated. Microwaves have an advantage in that regard; microwaves from multiple sources can produce constructive interference when the sources (i.e., antennas) are placed accordingly (6). Thereby, large SAR at far distances from the antennas can be achieved. Figure 5 shows the SAR of a square array of four microwave antennas, both for 915 MHz and 2.45 GHz microwaves. Depending on distance of the antennas and wavelength (i.e., frequency), different interference patterns result. An additional way to adjust the SAR pattern is to modify the phase angle at which the microwaves are supplied to the antennas in the array (7).

Ultrasound Ablation. Ultrasound as used in medical applications is typically in the frequency range of 0.5–20

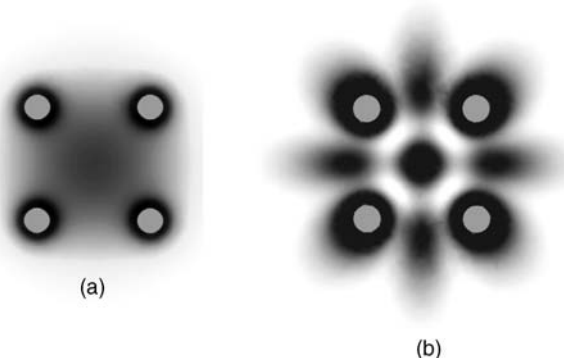


Figure 5. Four dipole antennas are placed in an array 2 cm apart, and driven in-phase at frequencies of 915 MHz (a) and 2.45 GHz (b). The SAR is shown in linear scale (black = maximum). Note constructive interference at array center (*left*), and complex interference pattern with constructive and destructive interference (*right*) due to shorter wavelength at 2.45 GHz. (Images provided courtesy of Deshan Yang, University of Wisconsin-Madison.)

MHz. At high enough intensities, the absorbed mechanical energy results in tissue heating. There are two fundamentally different ways of delivering ultrasound energy to the tissue. Ultrasound can be applied noninvasively from the outside of the body, and focused at the treatment site (High Intensity Focused Ultrasound, HIFU). In addition catheters with integrated ultrasound transducers can be inserted into the tissue, similar to other ablative techniques (Interstitial Ultrasound Thermal Therapy, or Direct Ultrasound Ablation).

High Intensity Focused Ultrasound. Ultrasound has a high penetration depth, and can therefore be applied from outside the body and still reach deep tissue sites. Typically, an array of ultrasound transducers is placed outside the body, and the waves are coupled into the tissue by a gel. The ultrasound waves are focused at the desired location, resulting in high intensity and heating near the focal spot (Fig. 6). The location and depth of the focal spot can be adjusted by modifying the phase difference between the transducer elements. A large area can be ablated by moving the focal spot and “painting” the desired area. The transducer may be attached to a computer controlled mechanical positioning system (8).

Interstitial Ultrasound Thermal Therapy (Direct Ultrasound Ablation). One or more transducer elements are placed inside a catheter. Figure 7 shows the schematics of such a catheter, and Fig. 8 shows achieved coagulation zones at different power levels. The catheter is inserted invasively into the application site, and tissue close to the catheter is heated by ultrasound. Sector transducers that emit ultrasound at angles between 30 and 270° (10) can be used, allowing for directional ablation which is important in certain applications like prostate treatment; on the other hand rf, MW, and laser applicators provide typically uniform heating (axial symmetric) around the applicator. Another potential advantage of direct ultrasound ablation is the use of the transducer elements also for imaging; an ultrasound image from inside the treatment zone can be obtained allowing the monitoring of tissue heating.

Laser Ablation (Laser Interstitial Thermal Therapy, LITT). When high intensity laser light is applied to tissue, the light is absorbed, resulting in tissue heating. The penetra-

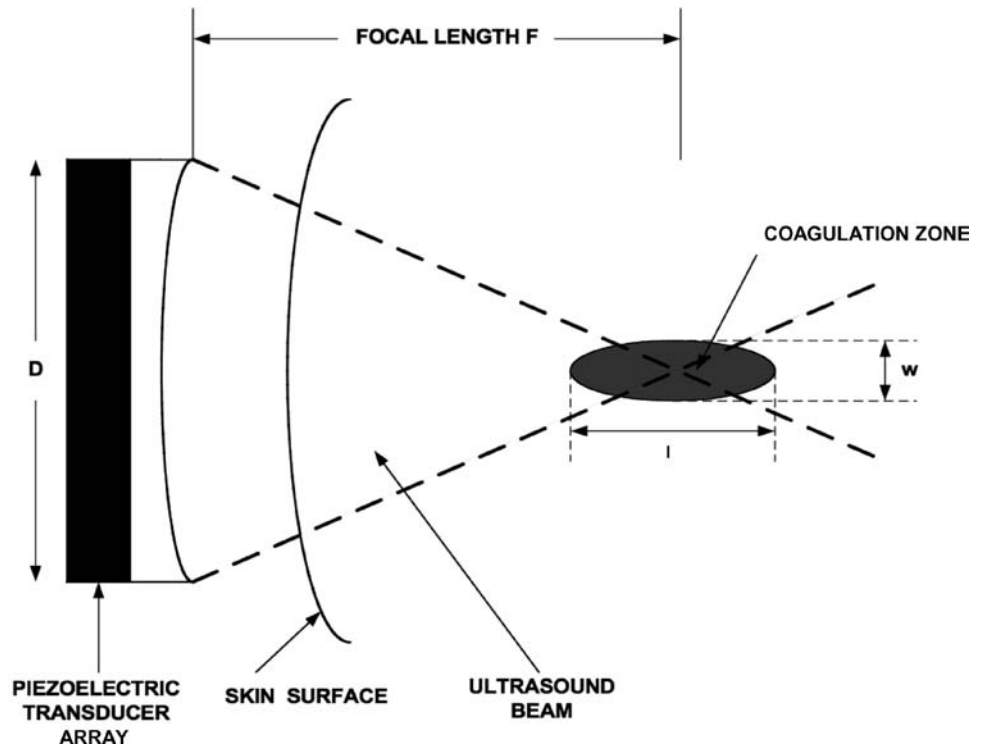


Figure 6. Schematic Diagram showing the principle of HIFU. An array of ultrasound transducers emits ultrasound waves into the tissue, which are focused and result in tissue coagulation around the focal point. The location of the focal point can be shifted by varying the phase between the array elements. (Image provided courtesy of Abhijit Sathaye, University of Wisconsin-Madison.)

tion depth is dependent on wavelength, where penetration increases with wavelength. The most widely used laser type, due to large wavelength and penetration is the Nd:YAG laser with 1064 nm wavelength (near-infrared, IR). This laser has a penetration depth of 3–4 mm. A quartz fiber with diffuser element at the tip is used to introduce the laser light to the treatment site. The transmission of the light into tissue changes during tissue heating. After tissue coagulation occurs, transmission is reduced to 69% of normal and further decreases to 15% of normal with onset of carbonization. Control of applied power to avoid carbonization is therefore essential. Typically, either constant power, or temperature feedback control where applied power is controlled so that the fiber tip is kept at constant temperature is used. Unlike most other ablation modalities, laser ablation has the advantage that the fibers are MRI compatible, allowing use of MR imaging (11).

Applicator Cooling. In the first heat-based ablation devices, the size of the achieved coagulation zone was insufficient for many applications like tumor ablation. Even today, there is a trend toward larger coagulation

zones to enable treatment of larger volumes. One successful method that has been applied to virtually all heat-based methods is cooling the applicator (e.g., catheter). In all catheter-based ablation methods, the highest SAR and subsequently highest temperature is obtained close to the catheter. In extreme cases, this can lead to tissue carbonization, which should be avoided as discussed earlier. Applicator cooling is achieved by circulating cooled water inside the catheter. Thereby, tissue close to the catheter is cooled and carbonization is prevented. Furthermore, the location of highest tissue temperature is more distant from the applicator (see Fig. 9). Since maximum tissue temperature is limited (e.g., to ~110 °C for rf ablation), this shift of maximum temperature results in an increase of the size of coagulation zone. The radial dimension of the coagulation zone can be increased by up to a factor of two when applicator cooling is used. Several commercial devices employ applicator cooling.

Cryoablation (Cryotherapy, Cryosurgery). Cryoablation is historically older than other ablative methods, and was introduced in the early 1960s (12). However, it

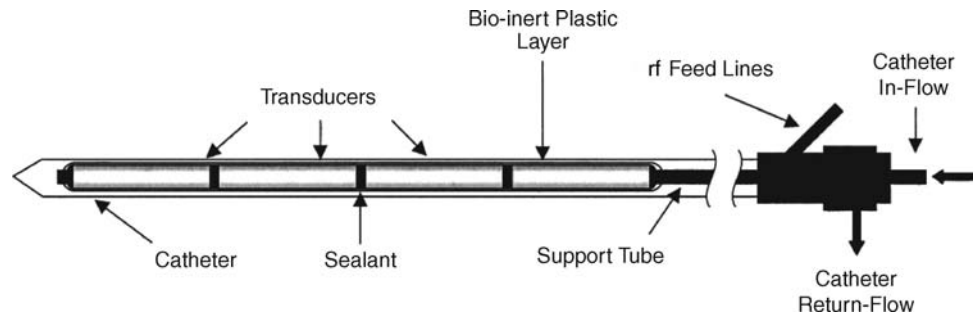


Figure 7. Catheter for Interstitial Ultrasound Thermal Therapy. An array of four transducers is used, and the catheter is cooled by circulating water. (From Ref. 9.)

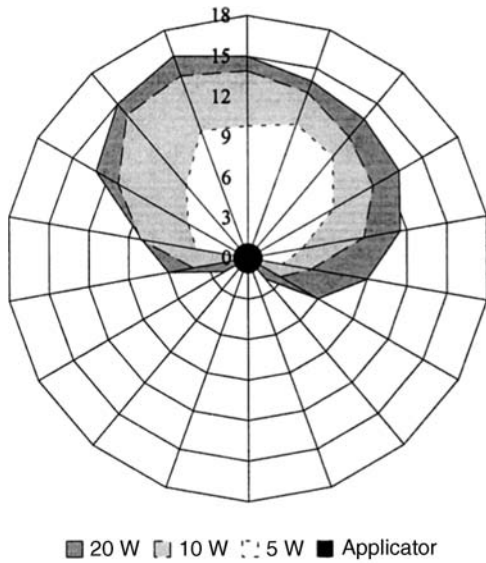


Figure 8. Thermal lesion sizes obtained with cooled ultrasound applicator at different power levels. Directional ultrasound arrays are used, which allows for better control of the coagulation zone. (From Ref. 9.)

did not receive considerable interest until the 1990s, when the development of interoperative ultrasound imaging allowed guidance of probe placement and the monitoring of the procedure.

Cryoablation relies on cooling (freezing) tissue to cause injury. Contrary to most heat-based methods, cryoablation relies solely on thermal conduction. A cryoprobe (see Fig. 10) is introduced into the tissue, and cooled by circulating a cryogen inside. Cryogens used for cryoablation include liquified gases, such as nitrogen, argon, helium, and nitrous oxide. Liquid nitrogen has a boiling temperature of $-196\text{ }^{\circ}\text{C}$. Temperatures reached at the cryoprobe are down to $-160\text{ }^{\circ}\text{C}$, resulting in formation of an iceball around the cryo probe. Nitrous oxide has a higher boiling temperature of $-88\text{ }^{\circ}\text{C}$, but has the advantage of being safer in case of leakage into the body compared to other gases.

Cooling with argon and helium is based on the Joule–Thompson effect, which involves expansion of a gas

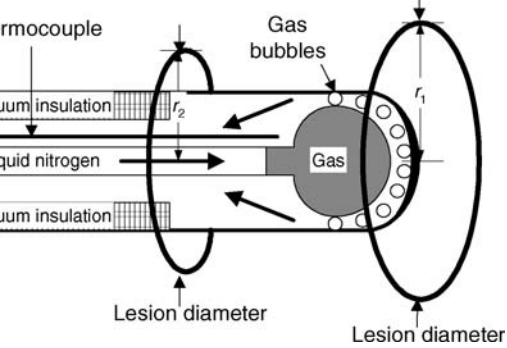
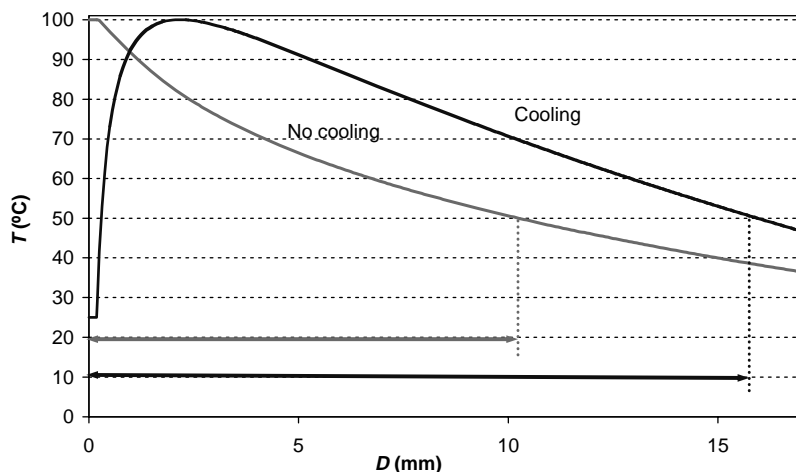


Figure 10. Internal structure of a typical cryoprobe based on liquid nitrogen (LN_2). The probe has vacuum insulation to prevent freezing of the probe shaft and subsequent destruction of normal tissue. The LN_2 changes phase when it hits the warm metal surface of the probe tip. Thus, a thin film of gas bubbles is formed on the metal surface resulting in lowest temperatures and largest ice ball near the tip. (From Ref. 13.)

through an orifice. The Joule–Thompson effect can produce either heating or cooling, depending on the type of gas, temperature, and pressure before expansion. For cryoablation, argon is decompressed from ~ 3000 psi (21,000 kPa) to ~ 150 psi (1000 kPa), resulting in cooling down to $-186\text{ }^{\circ}\text{C}$ (the boiling temperature of argon). Figure 11 shows a typical temperature distribution in tissue surrounding a cryoprobe.

One advantage of cryoablation over heat based methods is the visibility of the ice ball using ultrasound imaging (14). The interface between the ice ball and surrounding tissue is evident in ultrasound imaging (see Figure 12), and allows real-time monitoring of the ablation procedure.

Chemical Ablation. During chemical ablation a cytotoxic agent is injected into the tissue site to be treated, and diffuses into the tissue from the injection site. The most commonly used agents are ethanol and acetic acid. When ethanol is injected (ethanol ablation or percutaneous ethanol injection) into tissue, it causes cell death by cell dehydration, protein denaturation, and thrombosis of small vessels. Acetic acid has the ability to dissolve lipids

Figure 9. Tissue temperature (T) as a function of distance (D) from applicator. With catheter cooling, location of maximum tissue temperature is moved farther away from the applicator, resulting in an increase in diameter of the coagulation zone. Ablation zone dimensions are indicated by arrows as the regions with temperatures $> 50\text{ }^{\circ}\text{C}$. This image shows tissue temperature during rf ablation (from computer model), but the principle of cooling is applicable to all heat-based ablation methods.

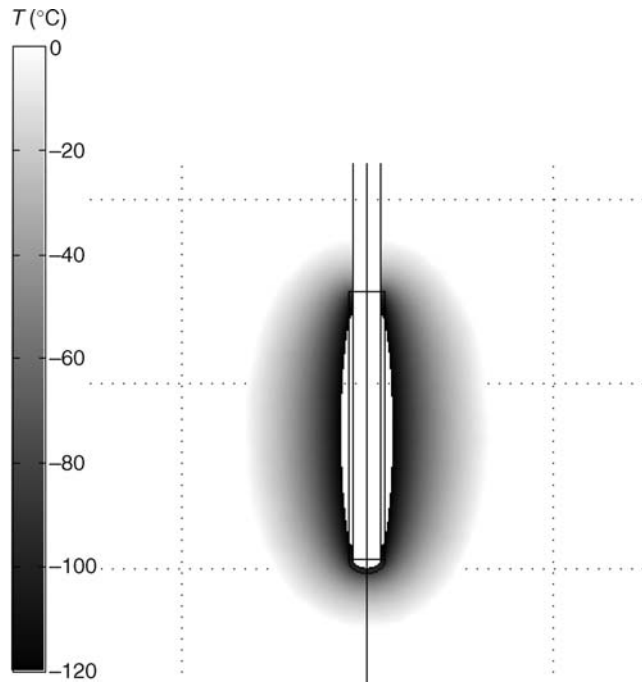


Figure 11. Tissue temperature next to a cryoprobe after 12 min. Outer gray boundary represents the border of the ice ball. Dotted lines are 20 mm apart. (Image provided courtesy of Cheolkyun Kim, University of Wisconsin-Madison.)

(e.g., cell membrane) and possesses a higher toxicity than ethanol.

PRINCIPLES OF THERMAL TISSUE INJURY

Tissue Injury from Heating

The normal range of human body temperature is between 36 and 38 °C; with fever, body temperature can rise up to

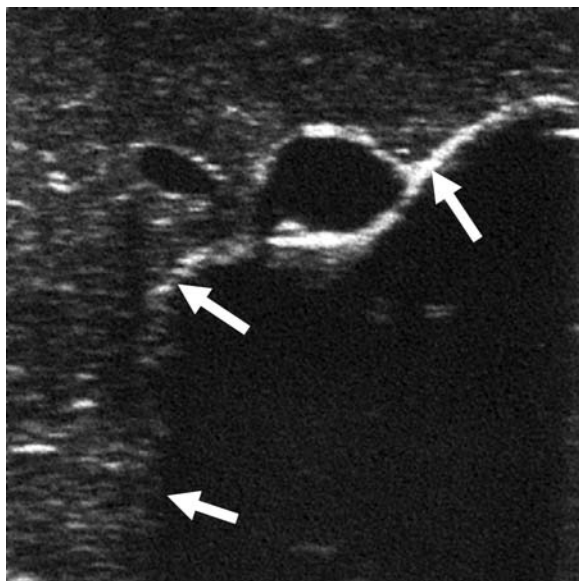


Figure 12. Ice ball forming during cryoablation is visible under ultrasound imaging (arrows). (From Ref. 14.)

Table 3. At Elevated Tissue Temperatures Different Effects Take Place Depending on Temperature and Time

Temperature, °C	Tissue Effects
36–38	Normal physiological range
38–42	Fever
>42	Elevated rates of enzyme activity, cell death possible
45	Protein coagulation (after 1–2 h)
50	Protein coagulation (after 2–3 min)
60–100	Protein coagulation (instantaneous)
>100	Tissue vaporization, carbonization

42 °C, which is the maximum temperature a person can typically sustain. Significant cell damage occurs above 42 °C (hyperthermia) with possible cell death, depending on time duration for which temperature is increased (15).

There are several different cell responses to elevated temperature. At low temperatures (42–45 °C), a number of subtle changes in metabolic activity, pH, blood flow, and vascular permeability occur. From ~45 °C protein coagulation occurs after 1–2 h of elevated temperature. Above ~60 °C instantaneous protein coagulation occurs, and above 100 °C tissue vaporization takes place, with possible carbonization at even higher temperatures. Table 3 lists the different temperature ranges and effects on tissue.

Since very high temperatures in excess of 60 °C are obtained during ablative therapies, cell death (necrosis) due to coagulation (i.e., coagulative necrosis) is the most important mechanism of tissue damage; the region of cell death is practically the region where tissue coagulation occurs. From Table 3, we see that we can roughly define the region of cell death (the coagulation zone) as the region where tissue reaches temperatures >50 °C, since ablative therapies have application times of typically between 1 and 30 min. Figure 13 shows a typical coagulation zone in liver tissue, cut right after rf ablation.

To determine exactly whether cell death results at a certain location, the time history of temperature has to be taken into account. It has been shown in many types of tissue that there is an exponential relationship between cell death, temperature, and time. Above 43 °C, the time required to cause cell death is cut in half with each degree centigrade of temperature increase (18). Even though the time may vary between different tissue types, this exponential time–temperature relationship is the same for all tissues. Figure 14 shows the time required for cell death for different types of tissue, plotted on a double-logarithmic scale. The exponential time–temperature relationship can be modeled mathematically by the Arrhenius model. The time–temperature relationship can then be expressed by an isoeffect equation:

$$t_1 = t_2 \times R^{(T_1 - T_2)} \tag{4}$$

where t_1 and t_2 are the treatment durations at treatment temperatures T_1 and T_2 , respectively. The parameter R can be assumed a constant with a value of 0.5 above 43 °C, and 0.25 below 43 °C. Because the onset of appreciable tissue damage occurs at ~43 °C, it has been suggested by Sapareto and Dewey (18) to quantify tissue damage by a

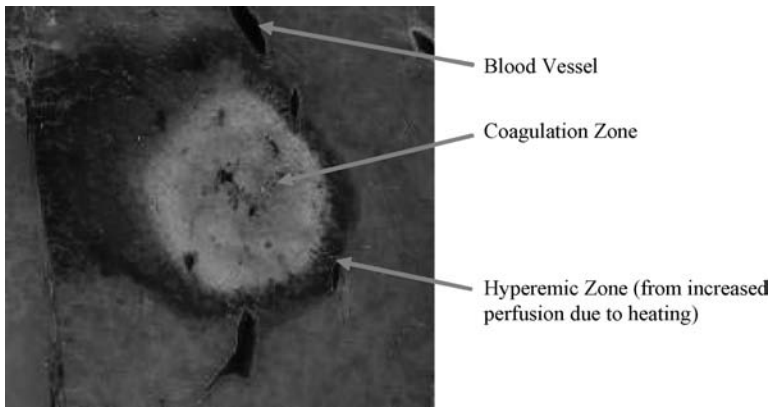


Figure 13. Gross pathology of a coagulation zone created by rf ablation *In vivo* in porcine liver. Liver was sliced right after the ablation procedure. The hyperemic zone contains viable cells.

thermal dose at 43 °C. This thermal dose is expressed as cumulative equivalent minutes at 43 °C (CEM_{43}); that is, a certain thermal treatment has the same effect as keeping the tissue at 43 °C for CEM_{43} minutes. If we set T_1 to 43 °C in equation 4 and allow temperature to be changing with time (as it does during ablative treatments), we obtain

$$CEM_{43} = \int R^{[43-T(t)]} dt \quad (5)$$

Once CEM_{43} exceeds a certain critical value, the tissue can be considered to be destroyed (i.e., ablated). The critical value of CEM_{43} has been measured for many tissues, and is ~340 min for liver; that is, tissue with $t_{43} > 340$ min can be considered to be destroyed. Even though the relationship stated in equation 5 may show inaccuracies in certain cases (e.g., long application times), it is an accurate approximation for the time durations and temperatures that occur during thermal ablation procedures.

Tissue Injury from Freezing

There are two basic mechanisms of tissue injury due to very low temperatures, dehydration and intracellular ice for-

mation (19). In general, extracellular water will freeze before intracellular water. The salinity of the remaining extracellular water increases, resulting in water transport from intra- to extracellular space due to difference in solute concentration. This water transport causes *dehydration* of the cells, and may result in cell death. While mainly dehydration happens at lower cooling rates ($<50 \text{ }^\circ\text{C}\cdot\text{min}^{-1}$), *intracellular ice formation* occurs at higher cooling rates. Formation of ice crystals inside the cells results in damage of cell membranes, organelles, and ultimately cell death. Generally, it is assumed that a minimum temperature of $-40 \text{ }^\circ\text{C}$ is required to ensure cell death.

Other mechanisms assumed to contribute to cell death are loss of blood supply and bursting of cells due to water pouring back into the cells during thawing. Multiple freeze-thaw cycles are typically used during cryoablation to accentuate these effects. An animal study has shown that the iceball boundary visible under ultrasound correlates with the boundary of cell death within 1 mm (20).

CLINICAL APPLICATIONS AND DEVICES

In the following section, different applications of tissue ablation are discussed. The areas where ablative therapies are used most widely clinically are treatment of cardiac arrhythmia and cancer. For each application, we will briefly discuss the clinical background, describe devices, and review imaging modalities used for guidance and monitoring.

Cardiac Catheter Ablation

Cardiac ablation is now a standard treatment method for different types of *cardiac arrhythmia* (i.e., abnormal heart beats) (21). Even though other ablative methods have been investigated, rf ablation is the most widely clinically used technique. Recently, cryoablation systems have become commercially available and are clinically used. Microwave systems have also become available. Other modalities like laser, ultrasound, and chemical ablation are mainly found in the research literature with no commercial devices currently available.

One major difference between cardiac ablation and ablation at other sites is that the catheter electrode is in direct contact with the blood pool inside the heart chamber. High temperatures can result in blood clot formation,

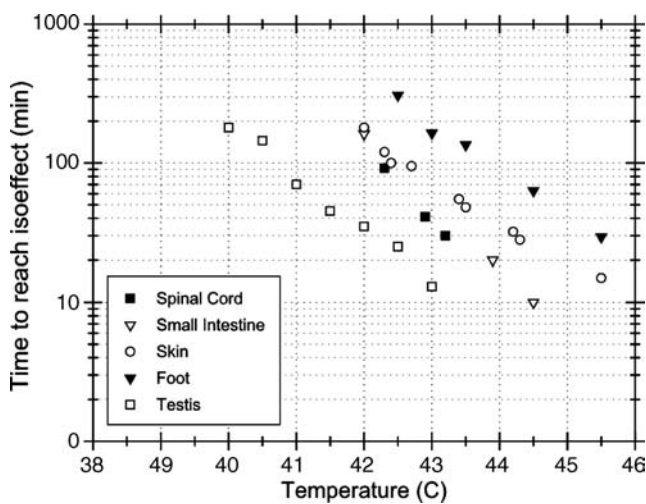


Figure 14. Time-temperature relationship of cell death for different cell types. Note that all cell types exhibit the same exponential relationship (i.e., parallel lines), even though they have different sensitivity (i.e., different times to cell death). (From Ref. 17.)

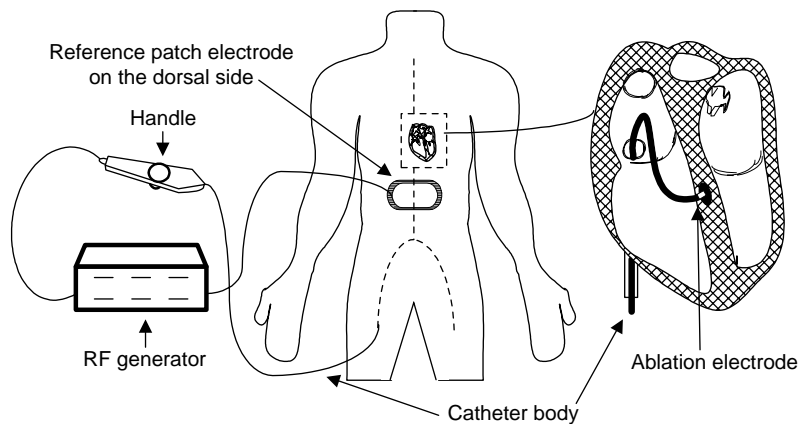


Figure 15. Overview of rf ablation system for treatment of cardiac arrhythmia. An ablation catheter is inserted through a leg vein, and steered into the heart to the treatment site. The procedure is guided by X-ray fluoroscopy. (From Ref. 22.)

which has to be avoided. Blood flow inside the heart results in cooling of the electrode (see Fig. 1), resulting in varying sizes of thermal lesion depending on blood velocity at the specific location inside the heart.

Clinical Background. Cardiac arrhythmias result from abnormalities in the conduction pathways in the cardiac tissue. The types of cardiac arrhythmias treated by ablative methods can be broadly classified into two categories.

Regular Tachycardias with a Discrete Mechanism. This is the condition of the heart (atrium or ventricle) beating too rapidly, typically at a rate >150 beats·min⁻¹ at rest. In a normal heart, the excitation (and associated contraction) starts at the sinoatrial (SA) node, the heart's pacemaker. Tachycardia results from excitation originating from locations other than the SA node, or from circular conduction involving abnormal pathways. Tachycardia can further be divided depending on where it originates, into supraventricular [above the ventricle, meaning the atria, the atrioventricular (AV) node and the bundle of His], and ventricular tachycardia (within the ventricle or purkinje system). These tachycardias can often be treated successfully with ablation at a single point or small area of the heart.

Atrial Fibrillation. Atrial fibrillation (AF) is the result of disorganized excitation of the atria, resulting in irregular contraction of the ventricles. The blood flow through the atria is hampered, possibly forming blood pools in the atria. Eventually, a blood clot may form that could lead to a stroke or heart attack (myocardial infarct). Atrial fibrillation is the most common type of arrhythmia, with ~2.2 million people affected in the United States, and 160,000 new cases diagnosed each year. Atrial fibrillation appears to often originate with abnormal excitations from the entry points of the heart, most commonly the pulmonary (lung) veins and less often the systemic veins from the body. Continuation of AF may depend on large areas of the atria. Because of the complexity of both the initiation and maintenance of AF, ablation typically requires a much broader treatment area than that for regular tachycardias. Since its introduction in the 1980s, cardiac catheter ablation has become a standard treatment for many of these types of arrhythmias.

Procedure. The ablation and associated electrophysiology study are performed in a specially equipped laboratory. The patient is either anesthetized or awake with sedation to reduce discomfort and facilitate relaxation. Electrodes and sensors attached to the patient report blood pressure, blood oxygen saturation, and electrocardiogram (ECG). Next, the locations where multiple diagnostic catheters and the ablation catheter will be inserted (typically groin and/or neck) are locally anesthetized. The catheters are inserted into a blood vessel, and guided into the heart (see Fig. 15). The physician performs mapping as described below to determine the mechanism of the arrhythmia, and site of ablation. The primary imaging modality used for guidance is fluoroscopy, but other techniques are now used to assist in both guidance and mapping, as described below. The patient is typically discharged a few hours after the procedure, but may be monitored in the hospital overnight.

Devices

Radio Frequency Ablation. Figure 16 shows a typical rf ablation catheter tip with ablation electrode, and mapping electrodes for measurement of biopotentials from the heart. Most rf devices use temperature control, where power is controlled to keep the temperature measured by a thermocouple or thermistor embedded in the catheter tip constant (typically 60–80 °C). Typical application times are 45–60 s.

For treatment of atrial fibrillation, linear (i.e., elongated) thermal lesions are required. Linear lesions can be created by dragging or sequentially moving a standard catheter (Fig. 16), but some special multielectrode catheters can be

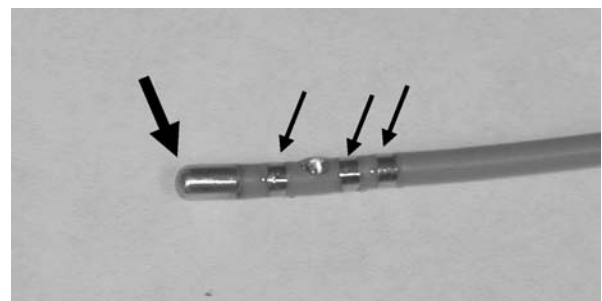


Figure 16. Cardiac rf ablation catheter (7F = 2.3 mm diameter) with ablation electrode (large arrow), and mapping electrodes (small arrows).

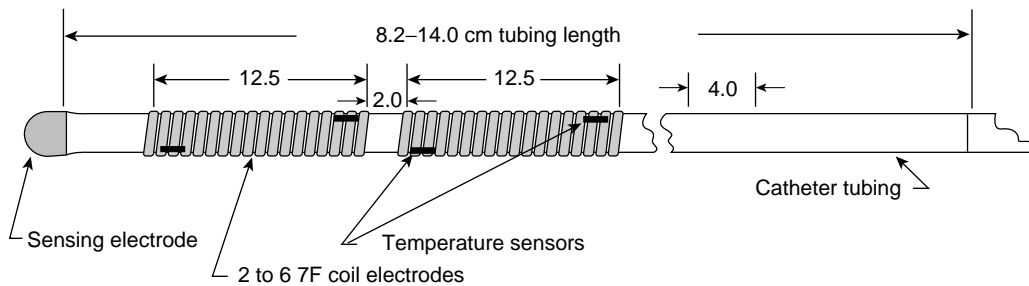


Figure 17. Cardiac multielectrode catheter for creation of elongated (linear) thermal lesions. Each electrode has two temperature sensors located near the edges for independent control. (From Ref. 23.)

more effective (Fig. 17). To keep all electrodes in the array at ideal temperatures, each electrode is controlled separately with multiple thermocouples placed at the edges of the electrodes.

Cryoablation. Cryoablation has the advantage of potential reversibility over heat based methods like rf ablation. For moderate hypothermic temperatures, tissue function can be restored after thawing. Thus, the correct location of a catheter can be confirmed by freezing at moderate temperatures (cryomapping). If the correct site is found (i.e., the arrhythmia stops or conduction is blocked), a longer freezing cycle with lower temperature is performed to destroy the tissue at that location. Once the tissue starts to freeze, the catheter sticks to the tissue (cryoadhesion), whereas with heat-based methods there is a risk of catheter movement. Other advantages include reduced complication rates compared to rf ablation. There should be minimal risk of tissue perforation, neighboring vessels like the coronary arteries seem to be preserved, and there is minimal risk of thrombus formation. Cryothermal lesion sizes are typically smaller than rf, and application times are longer (~4 min).

Microwave Ablation. Recently, microwave ablation catheters have become available for creating linear lesions (Microwave Ablation System Flex, Guidant, Indianapolis, IN), and are currently in clinical trials for treatment of atrial fibrillation. Microwave ablation has the potential benefits of directional heating, and deeper penetration compared to rf heating.

Cardiac Mapping and Imaging. Before ablation can be attempted, the treatment site in the heart has to be located, which is done by a procedure termed cardiac mapping. During the mapping procedure, local biopotentials known as electrograms are recorded typically from multiple catheters placed at different sites within the heart (Fig. 18). The ablation catheter itself has multiple mapping electrodes to record the electrograms (Fig. 16). This mapping procedure is performed under fluoroscopy to visualize the location of each catheter in the heart. These catheters can also be used for pacing the atria and/or the ventricles. From the temporal relationship between the electrograms, and from the relationship between these potentials and the ECG, the physician determines the mechanism of the arrhythmia.

During the mapping procedure, the mapping–ablation catheter is moved to locate the exact site that causes the arrhythmia, where ablation is then performed.

Multiple Electrode Mapping. Mapping using the above procedure can be long and cumbersome. Mapping with multielectrode catheters can speed up the mapping procedure by acquiring simultaneous electrograms, and enabling pacing from multiple locations. Fig. 19 shows the Constellation catheter (Boston Scientific), a basket-type catheter that is expanded within the heart, and can pace or record signals from 64 electrodes.

Another multielectrode catheter, the Ensite Array (St. Jude Medical), also employs 64 electrodes, but by using an inverse solution it estimates ~3000 signals on the interior (endocardial) surface of the heart chamber. It is deployed in without direct contact with the heart tissue (Figure 20). This has the advantage of not disturbing the beating heart and providing a highly detailed map, but is prone to errors since signals are extrapolated without direct tissue contact. Since fluoroscopy is not required during mapping, the radiation dose for the patient is highly reduced.

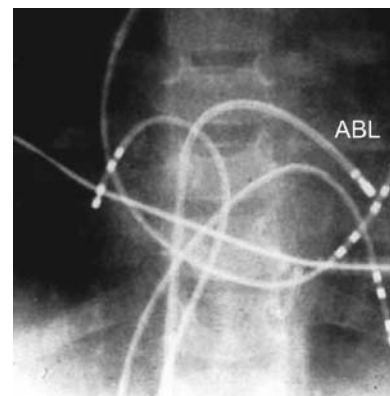


Figure 18. Mapping procedure using multiple catheters. This is an X-ray image of the heart showing the ablation catheter (ABL), and additional catheters for recording electrograms inside the heart. From the relationship between the electrograms at the different sites, the physician can determine the mechanism of the arrhythmia, and site of ablation. (Image provided courtesy of J. Phil Saul, Medical University of South Carolina.)

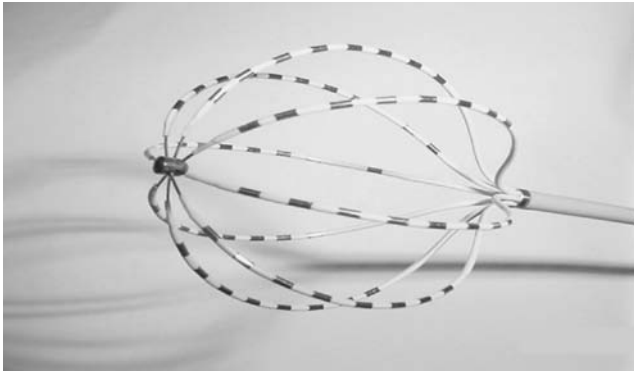


Figure 19. Constellation catheter has a total of 64 electrodes for cardiac mapping. The catheter is expanded inside the heart chamber. (From Ref. 24.)

Electroanatomic Mapping System. The CARTO system (Biosense-Webster) uses electromagnetic methods to determine the exact location of a specially designed catheter (25). A magnetic emitter employing three orthogonal coils is placed under the operating table, below the patient's chest. The reference catheter has a set of antennas located at the tip, which picks up the three magnetic fields. From the strength of the magnetic fields, the distance of the catheter from the magnetic emitter, the exact orientation of the tip (and subsequently the location in the heart) is determined. At each location, the reference catheter measures electrograms within the heart. By moving the catheter within the heart, a three-dimensional (3D) surface depicting the spread of activation through the heart (activation map) is created. From the activation map, the physician can determine the mechanism of the cardiac arrhythmia, and design a treatment plan. As with the Ensite Array, fluoroscopy dose can be markedly reduced with the CARTO system.



Figure 20. Ensite array catheter has 64 electrodes located on the surface of an expandable balloon catheter. The catheter records signals while floating inside the heart without contact. (Image provided courtesy of Endocardial Solutions.)

Impedance Mapping. The impedance signal between each catheter electrode and a set of reference electrodes can also be used to localize catheter positions. However, the accuracy of the location will depend somewhat on the variations in the impedance of the tissues through which the signal traverses. Two commercial systems, LocaLisa (Medtronic) and NAVEX (St. Jude Medical) use this technique to track catheter locations in real time, reducing mapping complexity and fluoroscopy time.

Tumor Ablation

After cardiac catheter ablation became clinically accepted, ablative methods were investigated for cancer treatment starting in the early 1990s. An estimated 70,000 clinical tumor ablation procedures had been carried out in 2004, with numbers still rising.

Clinical Background. The cancer type where tumor ablation was applied first is liver cancer. There was a need for a new treatment modality for liver cancer because surgery, the standard treatment, is only possible in ~15–20% of the cases, and chemotherapy does not work well for liver cancer.

As for many patients there was no viable treatment option, tumor ablation quickly became a standard treatment for unresectable (i.e., not treatable by surgery) liver cancer. More recently, tumor ablation techniques have been applied to other sites such as kidney, lung, and bone.

Tumor ablation can be performed during open surgery, using laparoscopy, or through a small incision in the skin (percutaneously). The treating physician is either a surgeon (open surgery, or laparoscopically) (26), or a radiologist (percutaneous approach) (27). Figure 21 shows a typical patient setup. The patient is either under conscious sedation, or light general anesthesia. The applicator is inserted under imaging guidance [typically ultrasound or computed tomography (CT)] into the tumor (in this case liver). Progress of ablation is monitored usually by ultrasound. Successful ablation is typically confirmed by CT. If

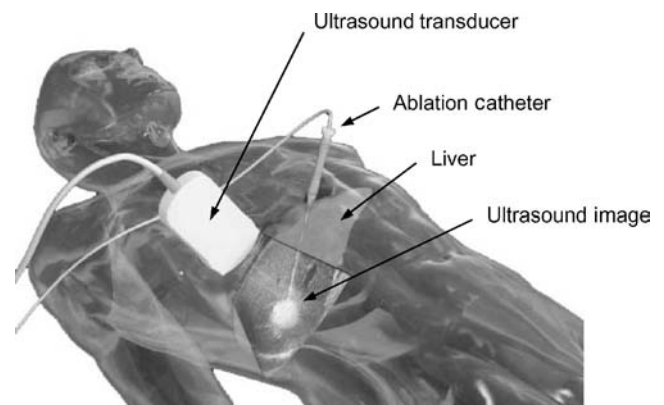


Figure 21. Patient setup for liver tumor ablation. The ablation catheter is inserted into the liver tumor through a small incision in the skin. Ultrasound imaging is used for guiding the electrode into the tumor, and monitoring the ablation. The white region in the overlaid ultrasound image represents areas of gas bubbles due to high temperatures. (From Ref. 28.)

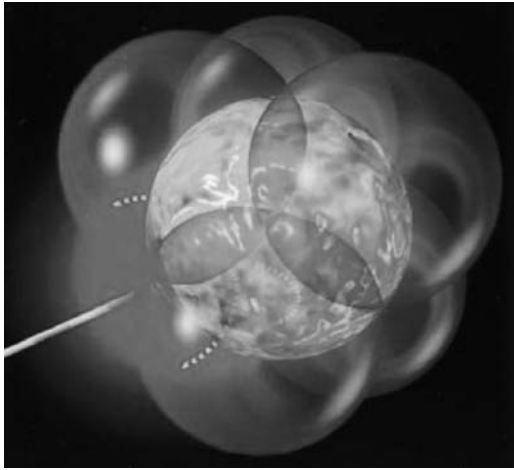


Figure 22. Multiple overlapping ablation zones are required for large tumors to ablate tumor and a 1 cm rim of surrounding tissue. (From Ref. 28.)

the procedure is performed minimally invasively, the patient can leave the hospital the same or the next day.

To successfully treat a tumor, the whole tumor including a 1 cm surrounding zone of normal tissue has to be ablated. It has been shown that the chance of recurrence (i.e., regrowth of the tumor due to incomplete treatment) increases when <1 cm of surrounding tissue is ablated. A single ablation can only treat a limited volume of tissue. Therefore, for larger tumors, multiple overlapping ablations have to be performed (see Fig. 22). This can be done with multiple applicators, but for devices that only support a single applicator (like most rf devices), multiple sequential ablations have to be performed.

Currently, long-term results are only available for liver cancer. For small (<3 cm) primary liver tumors (hepatocellular carcinoma), complete tumor necrosis is achieved in typically 80–90% of the cases, with 3 year survival rates of $\sim 75\%$. Results are less favorable for metastatic cancer, where complete necrosis in 52–67% of small tumors is achieved, with 3 year survival of 40% (28). Treatment results are significantly worse with tumors >3 cm, where multiple sequential ablations are required.

Devices. Cryoablation was historically the first ablation method applied to cancer treatment. However, cryoa-

blation has been so far mainly limited to use during open surgery. Bleeding from the insertion site could result in internal bleeding if cryoablation were done minimally invasively. The first thermal ablation method that was applied minimally invasively through a small incision in the skin (percutaneously) was rf ablation (Chemical ablation was used even earlier). Today, rf ablation is the most widely used tumor ablation method, with competing device technologies, like MW ablation, emerging on the market.

Radio Frequency Ablation. There are currently three rf devices for tumor ablation commercially available on the U.S. market (29). They use applied power between 200 and 250 W, application times of 12–45 min, and create coagulation zones of 3–6 cm in diameter. All manufacturers employ different electrodes and power control algorithms. Figure 23 shows the different electrode types. Grounding pads (2–4) for rf current dissipation are typically located on the patient's thighs, equidistant from the rf electrode. One disadvantage of most current rf devices compared to other methods (MW, cryo, laser) is that only one electrode can be used at a time. This prolongs treatment of larger tumors (>3 cm diameter), with up to several hours procedural time. Below we compare the three most widely used systems.

RITA Medical. This system employs a tree shaped multiprong electrode (Fig. 23a). The prongs are stepwise expanded during the ablation procedure. Five of the prongs have thermocouples located at the prong tips that monitor tissue temperature. The rf power is regulated to keep tip temperatures ~ 100 °C. Different electrodes are available depending on tumor size. In one electrode design, saline is infused into the tissue during the ablation procedure. Local cooling and increase in thermal and electrical conductivity from the saline results in increased coagulation zones, up to 6 cm in diameter with 45 mins application time.

Boston Scientific. This system employs an umbrella shaped multiprong electrode (Fig. 23b). The prongs are completely expanded once the catheter is placed in the tumor. Power is controlled depending on impedance, which is measured between the electrode and the ground pad. If impedance exceeds a certain threshold resulting from tissue vaporization near the electrode, power is turned off for 15 s, and then applied at 70% of previous power level.

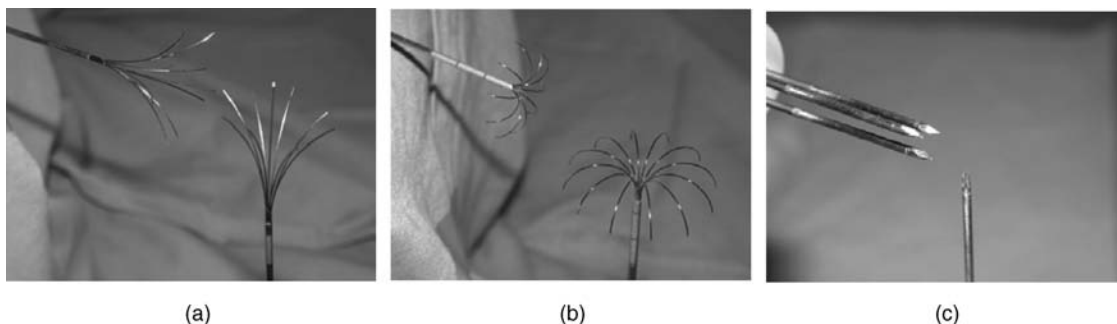


Figure 23. The rf electrodes of three manufacturers currently commercially available in the United States, and in clinical use. Multiprong electrodes by RITA medical (a), and Boston Scientific (b). Cooled needle electrode (single and triple cluster) by ValleyLab (c).

Valleylab. This system uses needle electrodes, which are cooled internally by circulating chilled water. Two types of electrodes are available (Fig. 23c): a single needle electrode and a triple-cluster electrode that achieve 3 and 4 cm coagulation zone diameters, respectively. This system employs impedance control, similarly to the Boston Scientific system. Treatment time is 12 min. Another recently introduced system originally developed at the University of Wisconsin, allows the use of up to three electrodes simultaneously, allowing rapid treatment of large tumors (30).

Cryoablation. Different cryoablation systems that employ either argon or liquid nitrogen cooling are clinically available. Cryoprobes in different sizes are available, though the general shape is often similar (Fig. 10). Cryoablation is mainly carried out during open surgery, though recently, small cryoprobes (17 gauge = 1.2 mm diameter) for minimally invasive treatment have become commercially available. Cryoablation has the advantage that the iceball is visible under ultrasound imaging, allowing real-time monitoring of the procedure (20).

Microwave Ablation. Microwave (MW) ablation has been used for small tumors with generic antennas in the Asian region for several years. Only recently has a MW ablation device for ablation of larger tumors become commercially available in the US (Vivant Medical, Mountain View, CA). Potential advantages over rf ablation are shorter treatment times and higher tissue temperatures in addition the possibility of use of multiple antennas takes advantage of constructive interference. Preliminary studies have indicated that MW ablation may show superior performance close to large vessels compared to rf ablation.

Imaging. Imaging serves different purposes during ablative treatment. Initially, the presence of a tumor is identified, typically by CT or MRI (though frequently additional tumors are found during interoperative ultrasound imaging, requiring change of the treatment plan). Figure 24 shows a contrast-enhance CT image of a patient with a liver tumor. Placement of the ablation applicator is typically guided by ultrasound imaging, and often confirmed by CT. Figure 25a shows an ultrasound image of



Figure 24. Tumor (arrow) can be identified from this contrast-enhanced CT image. Tumors are visible because they are typically hypervascular (i.e., have higher density of blood vessels than normal), and take up more of the contrast agent. (Image provided courtesy of Bradford J. Wood, NIH.)

a tumor with inserted rf electrode. Real-time monitoring of the ablation procedure is done by ultrasound imaging. Figure 25b shows the bright (hyperechoic) region resulting from microbubbles due to tissue heating during rf ablation; however, this bright area does not correspond to the zone of tissue destruction. Figure 12 shows the dark region (hypoechoic) depicting the boundary between frozen and unfrozen tissue during cryoablation. This boundary corresponds well with the boundary of cell death (20), which is one of the advantages over heat-based methods. After the ablation procedure, destruction of tumor and sufficient margin of normal tissue is confirmed by another CT scan or MRI.

Contrast agents for ultrasound imaging have recently become available in Europe (31), and should be available in the United States within the next years. These contrast agents employ microbubbles that are visible under ultrasound to visualize vasculature. Tumors are typically hypervascular, and show up using these contrast agents. As the tissue coagulates during ablation and blood perfusion stops, the contrast agent cannot penetrate the

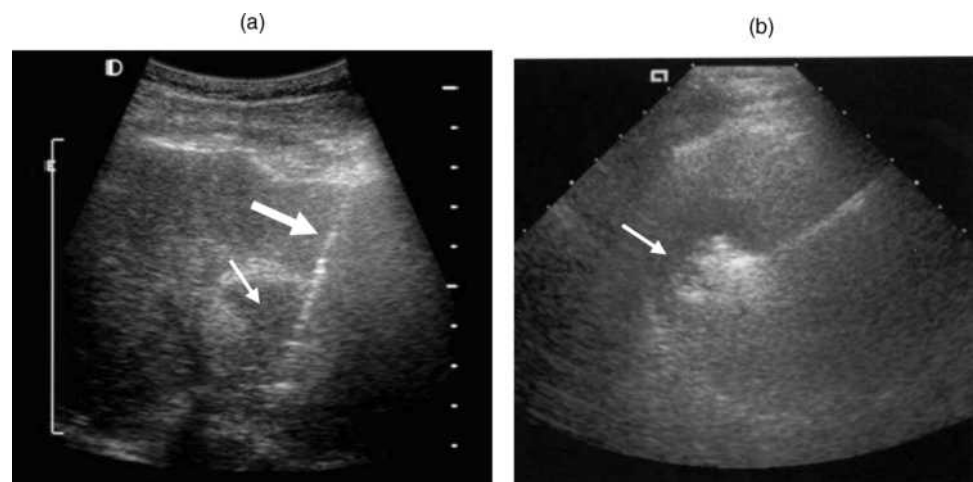


Figure 25. Ultrasound imaging is typically used to guide and monitor the ablation procedure. (a) Shows the rf needle electrode (large arrow) inserted into the tumor (small arrow). (b) Shows bright area (arrow) due to formation of microbubbles as a result of tissue heating. (Images provided courtesy of Bradford J. Wood, NIH.)

coagulated areas anymore. Ultrasound imaging using contrast agents allows real-time monitoring of the ablation procedure contrary to CT and MRI, which can typically only be performed after the procedure.

Future Directions. A continuing trend in tumor ablation is toward larger coagulation zones and shorter treatment times. High tumor recurrence rates close to large vessels are still problematic, with improved applicators attempting to address this issue. Adjuvant therapies (e.g., radiation, chemotherapy agents) can kill cells at lower elevated temperatures (above ~ 45 °C) than heat alone (above ~ 50 °C). Thereby, the zone of cell death may be extended (cf. Fig. 13), which may be beneficial close to large vessels where it is difficult to achieve sufficient heating.

Unavailability of adequate monitoring of the extent of the coagulation zone is a major problem for all heat-based methods, especially when damage to nearby structures is to be avoided. This problem is partially addressed by ultrasound contrast agents as discussed above. Other potential imaging solutions include thermometry, using MRI or ultrasound imaging, and ultrasound elastography, which allows imaging of the coagulation zone due to change in elasticity after coagulation. Some companies are working on computer-assisted treatment planning systems that guide the physician in optimal applicator placement, and estimate coagulation zones.

Endometrial Ablation

Clinical Background. Ablation has become a standard procedure for women with dysfunctional uterine bleeding, with $\sim 30,000$ annual procedures. Endometrial Ablation (endometrium = lining of the uterus) is indicated for patients that do not respond to standard treatments like drugs and curettage (scraping tissue from the endometrium with a spoon-shaped instrument). Initially either laser or rf energy applied by a rollerball electrode was used; the physician had to manually direct the catheter to ablate the whole *endometrium*, which was time consuming. A number of newer U.S. Food and Drug Administration (FDA) approved devices that are more effective with shorter treatment times are now available (32). These second generation devices treat the whole lining at once, and provide success rates of 67–80%.

Devices. There are currently five second generation devices available on the U.S. market. Two devices use heated fluid to ablate the endometrium. In one device, heated saline (80–90 °C) is circulated within the uterine cavity for 8 min (Hydro Thermablator, Boston Scientific). Another device employs a balloon catheter, inside which heated Dextrose solution (5%) at 87 °C is circulated for 10 min (Thermachoice UBT, Gynecare). One device uses cryoablation with a freeze–thaw cycle (10 min freezing) at two locations in the uterine cavity (Her Choice, AMS). The shortest treatment time is achieved by a device that uses bipolar rf ablation employing mesh electrodes (Novasure, Cytac Corp.; see Fig. 26). The rf energy is applied for 90–120 s, and controlled by tissue impedance. The last device uses microwave ablation at 9.2 GHz, and 42 W power for 3 min (MEA, Microsulis).



Figure 26. NovaSure catheter used for endometrial ablation. This catheter has an electrode consisting of multiple electrically isolated meshes, with rf energy applied bipolar between different meshes. Image provided courtesy of Cytac Corporation and affiliates.

Prostate

Clinical Background. Two types of prostate diseases are prevalent in men. Prostate enlargement (Benign prostate hyperplasia) is common in men >50 years of age. The enlarged prostate imparts pressure on the urethra resulting in restriction, at which point it has to be treated to reduce prostate size. Cancer is the second common disease affecting the prostate, and is the most common form of cancer affecting men.

Current standard treatment options for enlarged prostate include medication, surgery where part of the prostate is removed, and minimally invasive ablative methods. For prostate cancer, treatment options are surgical removal (radical prostatectomy) and external radiotherapy. Ablative methods are currently only used in cases when conventional treatment fails, or in cases with advanced disease to reduce tumor volume. Ablative treatment of prostate cancer requires more precise control of the ablation zone than for enlarged prostate, since the cancer is usually located in the periphery of the prostate (away from the urethra) making treatment from inside the urethra more difficult (33).

Commercial devices are available that employ cryo, rf, microwave, and ultrasound. Depending on the device, the prostate is treated by inserting the ablation catheter into the urethra, rectum, or through the skin in the region between the scrotum and anus (perineum). Urethra and rectum have to be protected from damage by cooling when heat-based devices are used, and by heating when cryoablation is used.

Devices for Enlarged Prostate Treatment

Microwave Ablation. Several devices are available that use microwave antennas inserted through the urethra.

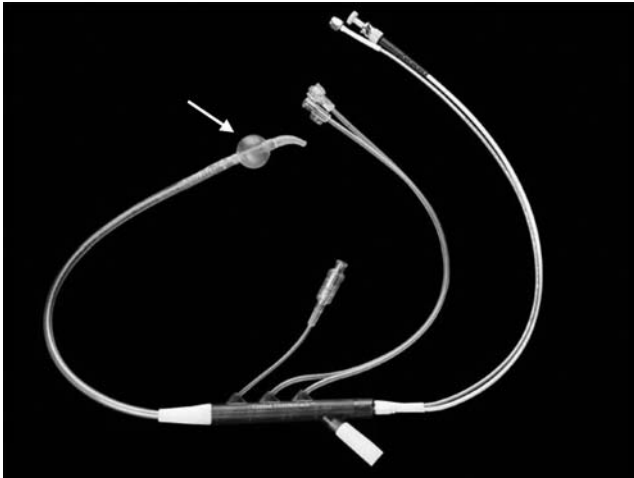


Figure 27. Targis catheter for treatment of enlarged prostate. The different ports are used for inflating the balloon (arrow) inside the bladder, and for perfusing the catheter with cooling water. (Image provided courtesy of Urologix.)

Two devices (Targis and Prostatron, Urologix) use cooled antennas to avoid damage to the urethra. Figure 27 shows the Targis device. The balloon is inflated inside the bladder to facilitate proper positioning. The Thermatrix device (AMS) uses low energy and does not require cooling. The CoreTherm device (ACMI) employs temperature sensors to monitor tissue damage and control applied power. Treatment time is typically 30–60 min, and tissue temperatures are in the range of 45–50 °C.

Radio Frequency Ablation. The transurethral needle ablation device (TUNA, Medtronic) uses a catheter, from which two needle electrodes project from the urethra into the prostate. Maximum power of 30 W is applied for 4 min, so that tissue temperature as measured by sensors located within the needles tips reaches 100 °C. Since each needle ablates only a small tissue volume, multiple insertions are required.

Devices for Prostate Cancer Treatment

Cryoablation. While the rf and MW devices described above are used for treatment of enlarged prostate, cryoablation is used for treatment of prostate cancer. The probes are typically inserted into the prostate through the skin (percutaneously) in the region between the scrotum and anus (perineum). The diameters of current probes are between 1.22 (SeedNet, Galil Medical) and 3 mm (Cryocare CS, Endocare) for the two systems currently available in the United States. A warming catheter perfused with 40–42 °C saline is placed in the urethra to protect it from damage.

Ultrasound Ablation. Two devices that employ high intensity focused ultrasound are commercially available in Europe and Asia, though not yet FDA approved in the United States. Both systems (Ablatherm, EDAP; Sonablate, Focus Surgery) use catheters inserted into the rectum, and catheters have both imaging and treatment transducers. When the treatment transducer is activated, it creates a

single cylindrical thermal lesion ~20 mm long and 1–2 mm in diameter. The focal point is gradually moved according to a treatment plan to ablate the treatment region.

Radio Frequency Ablation. The same devices described under “Devices section for Tumor Ablation” are also under investigation for treatment of prostate cancer.

Endovascular Ablation

Clinical Background. Varicose veins are visibly dilated and twisted veins near the skin surface, most often affecting legs and thighs. Insufficiencies of the venous valves result in blood pooling and enlargement of the veins. Varicose veins affect ~10% of the population, mostly between ages 30 and 60.

Several treatment options are available, all aiming to close the affected veins. Veins can be treated using surgical stripping, sclerotherapy (injection of an agent that causes vein swelling, and closure), and ablation. During treatment, a catheter is introduced into the vein and the vessel wall is heated. The heat results in shrinkage of the collagen in the wall, eventually closing off the vessel.

Devices

Radio Frequency Ablation. A bipolar, multielectrode device (ClosurePlus, VNUS Medical) is commercially available (see Fig. 28) in two sizes (2 and 2.7 mm diameter). The catheter is inserted into a vein and advanced to the treatment location. The rf energy is applied between two sets of electrodes (2.7 mm catheter), or between outer electrodes and inner ball electrode (2 mm catheter). Power is controlled so that vein wall temperature, as measured by a thermocouple (located in the outer electrodes tips), reaches 85 °C. Heparinized saline is infused through the central lumen of the catheter to prevent blood coagulation. Once the target temperature is reached, the operator moves the catheter at a rate of 2–3 cm per minute while keeping

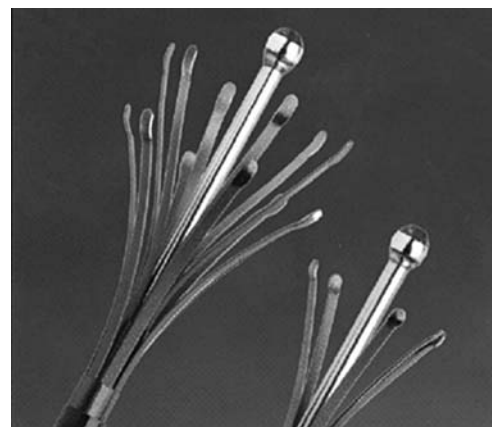


Figure 28. ClosurePLUS endovascular rf catheters for treatment of varicose veins in two sizes. The catheter is introduced with retracted prongs, and expanded at the treatment site. The rf current is passed between the center electrode, and surrounding prongs (left, smaller 2 mm catheter), or between two sets of prongs (right, larger 2.7 mm catheter), while catheter is pulled along the vein to shrink and close a section of the vein. (Image provided courtesy of VNUS Medical Technologies, Inc.)

Table 4. List of Manufacturers of Ablation Devices

Application	Company	Device Name (type)
Cardiac	Biosense-Webster (Diamond Bar, CA) www.biosensewebster.com Tel. 800-729-9010	Stockert 70 (rf ablation) Carto XP (cardiac mapping system)
	Boston Scientific (Natick, MA) www.bsci.com Tel. 888-272-1001	Cobra (rf ablation) Chilli (rf ablation) EPT-1000XP (rf ablation) RPM (cardiac mapping system)
	Cardima (Fremont, CA) www.cardima.com Tel. 800-354-0102	Revelation (rf ablation)
	Cryocath (Montreal, Canada) www.cryocath.com Tel. 877-694-1212	Freezor (Cryoablation) SurgiFrost (Cryoablation)-FrostByte (Cryoablation)
	Cryocor (San Diego, CA) www.cryocor.com Tel. 858-909-2200	CryoBlator (Cryoablation)
	Medtronic (Minneapolis, MN) www.medtronic.com Tel. 763-514-4000	Atakr (rf ablation) LocaLisa (cardiac mapping system)
	St. Jude Medical (St. Paul, MN) www.sjm.com Tel. 800-328-9634	Livewire (rf ablation) Ensite Array (cardiac mapping system) NAVEX (cardiac mapping system)
		RF 3000 (rf ablation)
		CelonSurgical (Bipolar rf ablation)
		Accuprobe (Cryoablation)
Tumor	Boston Scientific (Natick, MA) www.bsci.com Tel. 888-272-1001	Seednet (Cryoablation) CryoHit (Cryoablation)
	Celon (Teltow, Germany) www.celon.com Tel. + 49-3328-3519-0	Model 1500X (rf ablation)
	Endocare (Irvine, CA) www.endocare.com Tel. 800-418-4677	Cool-Tip (rf ablation)
	Galil Medical (Westbury, NY) www.galilmedical.com Tel. 516-794-7020	VivaWave (Microwave ablation)
	Rita Medical (Mountain View, CA) www.ritammedical.com Tel. 650-314-3400	
	Valleylab (Boulder, CO) www.valleylab.com Tel. 800-255-8522	
	Vivant (Mountain View, CA) www.vivantmedical.com Tel. 650-694-2900	
Endometrial	AMS (Minnetonka, MN) www.americanmedicalsyste.ms.com Tel: 800-328-3881	Her Choice (Cryoablation)
	Boston Scientific (Natick, MA) www.bsci.com Tel. 888-272-1001	Hydro Thermablator (heated saline)
	Cytec (Marlborough, MA) www.cytec.com Tel. 800-442-9892	Novasure (Bipolar rf ablation)
	Gynecare (Somerville, NJ) www.gynecare.com Tel. 888-496-2273	ThermaChoice UBT (Balloon ablation, heated fluid)
	Microsulis Medical (Hants, UK) www.microsulis.co.uk Tel. +44-23-9224-0011	MEA (Microwave ablation)
	ACMI (Southborough, MA) www.acmicorp.com Tel. 888-524-7266	
Prostate	AMS (Minnetonka, MN) www.americanmedicalsyste.ms.com Tel: 800-328-3881	CoreTherm (Microwave ablation)
	EDAP TMS (Lyon, France) www.edap-tms.com Tel. +33-472-1531-50	Thermatrix (Microwave ablation)
	Endocare (Irvine, CA) www.endocare.com Tel. 800-418-4677	Ablatherm (Focused ultrasound)
	Ethicon Endo-Surgery (Cincinnati, OH) www.ethiconendo.com 800-873-3636	Cryocare CS (Cryoablation)
	Focus Surgery (Indianapolis, IN) www.focus-surgery.com Tel. 317-541-1580	Indigo Laser System (Laser ablation)
	Galil Medical (Westbury, NY) www.galilmedical.com Tel. 516-794-7020	Sonablate (Focused ultrasound)
	Medtronic (Minneapolis, MN) www.medtronic.com Tel. 800-328-2518	SeedNet (Cryoablation)
	Urologix (Minneapolis, MN) www.urologix.com Tel. 800-475-1403	TUNA (rf ablation)
		Targis and Prostatron Systems (Microwave ablation)
		EVLT (Laser ablation)
	Endovascular	Diomed (Andover, MA) www.diomedinc.com Tel. 987-475-7771
VNUS Medical Technologies (Sam dose, CA) www.vnus.com Tel. 888-797-8346		

Table 4. (Continued)

Application	Company	Device Name (type)
Cornea	Alcon Laboratories (Fort Worth, TX) www.alconlabs.com Tel. 800-757-9195	LADARVision (Laser ablation)
	Bausch & Lomb (Rochester, NY) www.bausch.com Tel. 800-553-5340	Technolas 217A (Laser ablation)
	Nidek (Fremont, CA) www.nidek.com Tel. 510-226-5700	NAVEX (Laser ablation)
	Refractec (Irvine, CA) www.refractec.com Tel. 800-752-9544	Viewpoint CK (rf ablation) NearVision CK (rf ablation)
	VISX (Santa Clara, CA) www.visx.com Tel. 408-773-7321	STAR S4 (Laser ablation)
Intervertebral	Smith & Nephew (Andover, MA) endo.smith-nephew.com Tel. 800-343-5717	SpineCath (rf ablation)
	Valleylab (Boulder, CO) www.valleylab.com Tel. 800-255-8522	discTRODE (rf ablation)

^aSorted by application type and alphabetically.

the temperature within 3 °C of the 85 °C target. Treatment success is confirmed via ultrasound imaging, where no flow should be present.

Laser Ablation. The second commercially available device (EVLTL, Diomed) uses laser to produce heating, shrinkage of the vein. A laser fiber is inserted into a vein and advanced to the treatment location. Laser light of 810 nm wavelength at 14 W power is applied, while pressure on the vein ensures contact between fiber and vein. The fiber is advanced at a rate of 2–3 mm per second.

A study comparing the ClosurePlus and EVLT devices showed much higher temperatures when the laser device was used, resulting in vein perforations and reduced performance compared to the rf device (34).

Cornea Ablation

Clinical Background. Many vision disorders are a result of imperfections in the shape of the cornea. Conditions treated include astigmatism (cornea has oblong shape), myopia (nearsightedness), hyperopia (farsightedness), and presbyopia (blurred vision at close range, due to age-related loss of elasticity).

Treatment options are surgical treatment (part of the cornea is surgically removed), and ablative methods.

Devices

Laser Ablation (Laser Refractive Surgery). Laser is the most widely used ablative treatment method for corneal ablation. For most applications the excimer laser is used, which is an Argon laser operating in the ultraviolet (UV) range (193 nm wavelength); the laser is used in pulsed mode with 10–60 pulses (each a few ns) per second. The excimer laser does not cause tissue damage as described in the “Tissue injury from Heating” section, but causes the corneal tissue to vaporize. Part of the cornea is thereby removed, resulting in the desired change of shape.

In some applications, the Holmium:YAG laser is used, which operates in the IR region (2100 nm wavelength). This laser causes heating, and shrinkage of the collagen in the cornea, resulting in change of shape.

Radio Frequency Ablation (Conductive Keratoplasty). More recently, rf devices have become available for the treatment of hyperopia and presbyopia. RF energy is applied by fine electrode (90 μm diameter, 450 μm long) in pulsed fashion (exponentially damped rf pulses, pulse rate ~8 kHz), with power levels of ~1 W, and application times in second range. The rf heating results in collagen shrinkage, and change of cornea shape.

Intervertebral Disk Ablation

Degenerative diseases of the intervertebral disks are a major cause of lower back pain. Ablative techniques have been introduced in the late 1990s, and are now used in certain patient populations (35). Even though exact mechanisms of pain alleviation are not known, it is assumed that two mechanisms are responsible: shrinkage of the disk reduces pressure on nerve fibers, and destruction of sensitive nerve fibers. Currently, two devices are commercially available, both of which use rf ablation to heat tissue.

Other Applications

Ablative techniques are investigated for a number of other applications. Treatment of different types of cancer other than the ones discussed above such as in the breast and esophagus, are being investigated. In the brain, treatment of deep-seated tumors and other disorders, like Parkinson’s disease and Epilepsy, is examined. Ablation is investigated for treatment of chronic pain by ablating responsible nerve fibers. In dentistry, laser ablation is investigated as a potential replacement for mechanical drills.

DEVICE MANUFACTURERS

See Table 4.

BIBLIOGRAPHY

1. Pennes HH. Analysis of tissue and arterial blood temperatures in the resting human forearm. *J Appl Physiol* 1948;1:93–122.

2. Lai YC, et al. Lesion size estimator of cardiac radiofrequency ablation at different common locations with different tip temperatures. *IEEE Trans Biomed Eng* 2004;51:1859–1864.
3. Wright AS, Lee FT Jr, Mahvi DM. Hepatic microwave ablation with multiple antennae results in synergistically larger zones of coagulation necrosis. *Ann Surg Oncol* 2003;10:275–283.
4. Labonte S, et al. Monopole antennas for microwave catheter ablation. *IEEE Trans Microw Theory* 1996;44:1832–1840.
5. Saito K, et al. Heating characteristics of array applicator composed of two coaxial-slot antennas for microwave coagulation therapy. *IEEE Trans Microw Theory* 2000;48:1800–1806.
6. Camart JC, et al. 915 MHz microwave interstitial hyperthermia. Part ii: Array of phase-monitored antennas. *Int J Hyperthermia* 1993;9:445–454.
7. Tremblay BS, et al. Effect of phase modulation on the temperature distribution of a microwave hyperthermia antenna array in vivo. *Int J Hyperthermia* 1995;10:691–705.
8. Hynynen K, McDannold N. Mri guided and monitored focused ultrasound thermal ablation methods: A review of progress. *Int J Hyperthermia* 2004;20:725–737.
9. Nau WH, Diederich CJ, Burdette EC. Evaluation of multi-element catheter-cooled interstitial ultrasound applicators for high-temperature thermal therapy. *Med Phys* 2001;28:1525–1534.
10. Diederich CJ, et al. Catheter-based ultrasound applicators for selective thermal ablation: Progress towards mri-guided applications in prostate. *Int J Hyperthermia* 2004;20:739–756.
11. Vogl TJ, et al. Mr-guided laser-induced thermotherapy (litt) of liver tumours: Experimental and clinical data. *Int J Hyperthermia* 2004;20:713–724.
12. Gage AA. History of cryosurgery. *Semin Surg Oncol* 1998;14:99–109.
13. Webster JG, editor. *Minimally Invasive Medical Technology*. Bristol, UK: IOP Publishing; 2001.
14. Lee FT Jr, Mahvi DM, Chosy SG, Onik GM, Wong WS, Littrup PJ, Scanlan KA. Hepatic cryosurgery with intraoperative us guidance. *Radiology* 1997;202:624–632.
15. Miller MW, Ziskin MC. Biological consequences of hyperthermia. *Ultrasound Med Biol* 1989;15:707–722.
16. Dewhurst MW, et al. Basic principles of thermal dosimetry and thermal thresholds for tissue damage from hyperthermia. *Int J Hyperthermia* 2003;19:267–294.
17. Sapareto SA, Dewey WC. Thermal dose determination in cancer therapy. *Int J Radiat Oncol Biol Phys* 1984;10:787–800.
18. Bischof JC, et al. Cryosurgery of dunning at-1 rat prostate tumor: Thermal, biophysical, and viability response at the cellular and tissue level. *Cryobiology* 1997;34:42–69.
19. Weber SM, et al. Hepatic cryoablation: US monitoring of extent of necrosis in normal pig liver. *Radiology* 1998;207:73–77.
20. Zipes DP, Haissaguerre M. *Catheter ablation of arrhythmias*. 2nd ed. Armonk, NY: Futura Publishing; 2001.
21. Panescu D, Wayne JG, Fleischman SD, Mirotznik MS, Swanson DK, Webster JG. Three-dimensional finite element analysis of current density and temperature distributions during radio-frequency ablation. *IEEE Trans Biomed Eng* 1995;42:879–890.
22. Panescu D, Fleischman SD, Wayne JG, Swanson DK, Mirotznik MS, McRury I, Haines DE. Radiofrequency multielectrode catheter ablation in the atrium. *Phys Med Biol* 1999;44:899–915.
23. Panescu D. Intraventricular electrogram mapping and radiofrequency cardiac ablation for ventricular tachycardia. *Physiol Meas* 1997;18:1–38.
24. Shpun S, et al. Guidance of radiofrequency endocardial ablation with real-time three-dimensional magnetic navigation system. *Circulation* 1997;96:2016–2021.
25. Poon RT, et al. Locoregional therapies for hepatocellular carcinoma: A critical review from the surgeon's perspective. *Ann Surg* 2002;235:466–486.
26. McGahan JP, Dodd GD. Radiofrequency ablation of the liver: Current status. *Am J Roentgenol* 2001;176:3–16.
27. Dodd GD, et al. Minimally invasive treatment of malignant hepatic tumors: At the threshold of a major breakthrough. *Radiographics* 2000;20:9–27.
28. Pereira PL, et al. Radiofrequency ablation: In vivo comparison of four commercially available devices in pig livers. *Radiology* 2004;232:482–490.
29. Haemmerich D, et al. Large-volume radiofrequency ablation of ex vivo bovine liver with multiple cooled cluster electrodes. *Radiology* 2005;234:563–568.
30. Solbiati L, et al. Guidance and monitoring of radiofrequency liver tumor ablation with contrast-enhanced ultrasound. *Eur J Radiol* 2004;51 (Suppl): S19–23.
31. Cooper J, Gimpelson RJ. Summary of safety and effectiveness data from fda: A valuable source of information on the performance of global endometrial ablation devices. *J Reprod Med* 2004;49:267–273.
32. Shinohara K. Thermal ablation of prostate diseases: Advantages and limitations. *Int J Hyperthermia* 2004;20:679–697.
33. Weiss RA. Comparison of endovenous radiofrequency versus 810 nm diode laser occlusion of large veins in an animal model. *Dermatol Surg* 2002;28:56–61.
34. Bass EC, et al. Heat-induced changes in porcine annulus fibrosus biomechanics. *J Biomech* 2004;37:233–240.

See also CRYOSURGERY; ELECTROSURGICAL UNIT (ESU); STEREOTACTIC SURGERY.

TISSUE ENGINEERING

KRISTYN S. MASTERS
WILLIAM L. MURPHY
University of Wisconsin
Madison, Wisconsin

INTRODUCTION (1–4)

Definition

Tissue engineering represents a unique convergence of work from the worlds of clinical medicine, engineering, and basic science. The most commonly cited definition of tissue engineering originates from an influential 1993 paper by Langer and Vacanti (1): “Tissue engineering is an interdisciplinary field that applies the principles of engineering and the life sciences toward the development of biological substitutes that restore, maintain, or improve tissue function.”

Although the term tissue engineering had existed for several years prior to this 1993 publication, and the concept of tissue engineering had existed for several decades, it is Langer and Vacanti's paper that is ultimately credited with stimulating broad awareness and acceptance of this description. This definition also served to unify seemingly diverse lines of research, as it encompasses three general strategies for the creation of new tissue, namely, the use of isolated cells or cell substitutes; tissue-inducing substances; or cells placed on or within matrices.

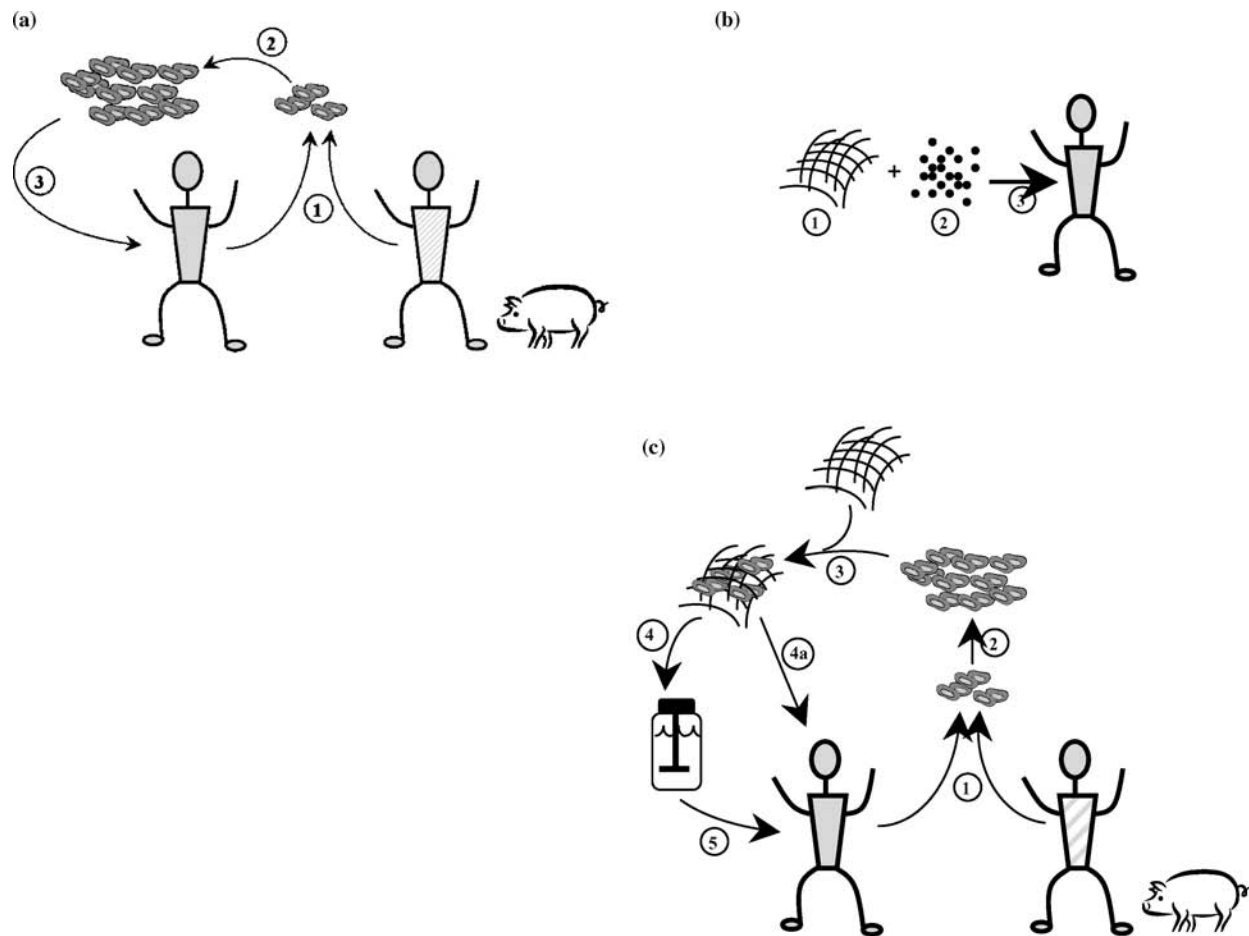


Figure 1. Depiction of three main tissue engineering strategies. In (a) cell transplantation, cells are obtained from a donor (1), who is either the patient themselves, another person, or an animal, then expanded *in vitro* (2), and reimplanted into the diseased/injured site (3). Regeneration via scaffold alone is represented by (b), where a degradable, and possibly bioactive scaffold (1) may be combined with inductive agents (2), and then implanted (3). Lastly, (c) depicts the implantation of cell-seeded scaffolds, as cells are isolated from a donor (1), expanded *in vitro* (2), and combined with a biomaterial scaffold (3), followed by either immediate implantation (4a) or further culture in an *in vitro* bioreactor (4b) prior to implantation (5).

Thus, tissue engineering includes a wide variety of therapies, including cell transplantation, the implantation of biomaterials alone as scaffolds for *in situ* cell growth, and the *in vitro* or *in vivo* development of cell-seeded materials; these strategies are depicted in Fig. 1. The choice of tissue engineering approach is dependent on the specific type of tissue to be repaired; factors, such as cell availability, tissue regeneration potential, and mechanical stresses in the tissue environment, are all important considerations in the selection of a tissue engineering strategy. Ultimately, the goal of tissue engineering is the formation of neotissues that are mechanically and biologically integrated in the patient's body.

Motivation

Therapeutic options for patients with diseased or dysfunctional tissues or organs are currently limited. While organ transplantation has become a successful means of replacing diseased organs, this option continues to be seriously

limited by problems with immune rejection and organ availability. Historically, the number of organs available for transplantation has been exceeded by the number of patients requiring an organ transplant, and this gap continues to widen. The number of needed organs continues to grow at an annual rate that is over twice the rate of increase of donated organs. Over 4000 people die each year while on the UNOS (United Network for Organ Sharing) transplantation waitlist, while 100,000 others die from organ failure without even qualifying for the waitlist.

Certain structural elements of the body (e.g., bone, blood vessels, and skin) may also be replaced via autologous tissue grafts. Such grafting procedures transfer the patient's own healthy tissue to a location that requires assistance with repair, thus circumventing immunogenicity problems associated with receiving tissues from other sources. However, tissue grafting is also accompanied by significant drawbacks and challenges, including limited availability of healthy donor tissue, donor site morbidity, and the need for multiple surgical sites. Grafted or reconstructed tissues are often

functionally inferior and less durable than the natural organs that they replace. Moreover, tissue grafting is appropriate for only select types of tissue.

Lastly, artificial prostheses and permanent implants have been developed to take the place of diseased or defective tissues, particularly in the areas of orthopedics and cardiovascular medicine. While these implants are capable of improving both the patient's lifespan and quality of life, tissue replacement with a permanent, artificial structure results in the loss of that tissue's natural biological functions. Most notably, artificial prostheses are unable to grow or remodel, meaning that they are often unsuitable for pediatric patients and are incapable of responding to changes in the body's needs or environment. Furthermore, significant issues arise from the permanent implantation of synthetic structures, and the types of tissues that can be mimicked by current prosthesis technology are very limited.

The field of tissue engineering holds the potential to overcome the aforementioned challenges associated with organ and tissue availability, immunogenicity, and retention of tissue function. Tissue engineering promises to provide a means of regenerating or replacing diseased or dysfunctional tissues and organs while leaving no permanent implant. While organ regeneration may be the most explicable motivation for the creation of engineered tissues, the objectives of tissue engineering as a field can be quite broad, as tissue engineering can be used to offer an alternative to drug therapy, gene therapy, and whole-organ transplantation; inspire or control the normal processes of tissue repair and healing; replace cells that are missing within an otherwise functional tissue or organ; use cellular control mechanisms to enhance drug delivery; lead to new models of human physiology.

The principal variables in the process of tissue engineering are (1) cell source; (2) scaffold type and properties; (3) method of tissue development (bioreactor design); (4) inclusion of inductive factors (i.e., cytokine delivery, gene therapy). This article will introduce and provide background on these key components of tissue engineering, and then discuss how tissue engineers have manipulated these variables in the development of two different types of engineered tissue: bone and cardiovascular tissue.

History

The National Science Foundation (NSF) (2) defines 1987 as the year that tissue engineering became formalized as a field; it is in this year that the NSF became consciously involved in tissue engineering. Although a defined label did not exist to describe their work, several pioneering researchers did perform tissue engineering research in the decades preceding the field's "official" emergence. In the vascular area, the concept of a resorbable vascular graft was introduced in the 1960s, with the first fully resorbable graft reported in 1979. In 1978, researchers observed improved healing after seeding synthetic vascular grafts with endothelial cells, and the first attempt to tissue engineer a vascular structure *in vitro* using collagen and cultured vascular cells was described by Burkel et al. in 1982 (see Ref. 3 for details). The origins of dermal tissue

engineering reach into the 1950s, when Billingham and Reynolds demonstrated that cultured epidermal cells could be applied to a graft bed to reconstitute an epidermis. By the 1970s, the ability to culture cells *in vitro* had significantly advanced, enabling the formation of multilayered epidermal sheets that could be transferred intact to a wound bed. The first living skin equivalent (LSE) was created using fibroblasts seeded on a collagen matrix and described in 1981 in a notable paper by Bell et al. (see Ref. 2 for details). In the area of orthopedics, the development of nondegradable biomaterial implants and the discovery of osteo- and chondrogenic cytokines and growth factors constituted the majority of pre-1987 tissue engineering work. Lastly, early (and current) tissue engineering of organs (e.g., the kidney, pancreas, and liver) was (and continues to be) hindered by the functional complexity of these structures and the difficulty of cellular expansion *in vitro*. Several types of cell-seeded, nonimplantable bioreactors were first developed during the 1970s to replace critical metabolic functions provided by these organs; however, there remains debate regarding whether such bioartificial devices fall under the definition of tissue engineering, as they are never incorporated into the body's reparative and homeostatic mechanisms.

In the years since these early forays into tissue engineering, the field has expanded substantially and has experienced some clinical and commercial successes. In 1988, Langer and Vacanti described a method of seeding cells on a resorbable polymer matrix for cell transplantation, which ultimately became the most important enabling technology for advancement, expansion, and recognition of the field of tissue engineering. This specific technique of seeding cells on a three-dimensional (3D), porous, biodegradable scaffold catalyzed an explosion of tissue engineering research in the late-1980s–mid-1990s. While the exploration of fundamental concepts underlying tissue engineering's viability generally took a backseat to the practice of tinkering with various combinations of cells plus materials, these investigations did significantly advance tissue engineering by enabling identification and a better understanding of the obstacles facing tissue engineers. By the year 2000, mainstream media outlets were touting the promise of this new field, with *Time* magazine proclaiming Tissue Engineer as the number one projected career in the twenty-first century.

Today's tissue engineering research sees the collaboration of engineers with clinicians, biologists, chemists, and many other scientists to effect the creation of numerous types of engineered tissues. Present tissue engineering ventures have addressed the regeneration or replacement of components found in every system of the body (cardiovascular, musculoskeletal, neural, endocrine, digestive, reproductive, and respiratory). While significant progress has been made toward the recreation of many complex tissues, the clinical and commercial success stories of tissue engineering represent less complicated structures (e.g., skin, cartilage, and the bladder). Yet, through the incorporation of emerging technologies, (e.g., the use of embryonic stem cells) the field of tissue engineering continues to evolve and progress in order to meet its promise of revolutionizing regenerative medicine.

COMPONENTS OF TISSUE ENGINEERING

Cell Sources (1,2,4–7)

As discussed in earlier sections, tissue engineering strategies may include the transplantation of cells alone or the seeding of cells upon implantable scaffolds. Cell sourcing is a formidable challenge in current tissue engineering techniques, as it is essential to select cells of appropriate origin and maturity for each tissue engineering application. In this section, cell sources for tissue engineering are arranged and discussed with respect to cell maturity, as the regeneration potential of a cell source is a crucial consideration in designing a tissue engineering strategy.

Mature Cell Sources. Mature cells are differentiated cells that are committed to performing given cell type-specific functions. These cells are generally obtained via primary cultures originating from small pieces of donor tissue. For example, a small skin biopsy may be performed in order to isolate keratinocytes and dermal fibroblasts, the two main cell types in skin. Once isolated, these cells may be cultured and expanded *in vitro* to yield the number of cells required for creation of tissue-engineered skin. However, the expansion potential of these cells is not without limit; over time, mature primary cells drift from their original phenotype, and will either die or become genetically unstable after a certain number of population doublings. One advantage of using mature cells in tissue engineering strategies is that it enables tissue–organ regeneration using the patient’s own cells, thus eliminating concerns of immunogenicity. Additionally, decades of research have concentrated on characterizing many mature cell types and their responses to various biological factors, thus facilitating the tissue engineer’s ability to predict and control cell behavior and tissue formation. Thus, in situations where a tissue sample of proliferative cells can be obtained using minimally invasive means, the use of primary cell lines as a component of the tissue engineering platform represents a viable strategy. The regeneration of cartilage via Carticel (Genzyme, Inc.) is an excellent example of the successful use of mature autologous cells in tissue engineering. In the Carticel process, healthy, mature chondrocytes are isolated from a patient with large articular cartilage lesions, propagated *in vitro*, and then reimplanted into the patient. Carticel is approved by the U.S. Food and Drug Administration (FDA) and has been clinically used on > 10,000 patients. Note, however, that isolation via primary culture and subsequent subculturing–expansion is not a feasible option for all cell types. For example, while cells from the liver (hepatocytes) readily regenerate *in vivo*, their *in vitro* growth is difficult and results in rapid loss of hepatocyte-specific functions. Furthermore, not all mature cell types are capable of regeneration. The cells of the adult myocardium (cardiomyocytes) are terminally differentiated and incapable of proliferation, even in the *in vivo* environment.

Immature Cell Sources. Historically, successful tissue engineering schemes have used mature, adult cells isolated from a specified tissue type as tools to produce their tissue

of origin. A common example, described above, involves use of dermal fibroblasts and keratinocytes to engineer skin tissue. This general strategy has gained a measure of success in cases where cells are accessible and expandable with minimal patient trauma, as in the case of skin and cartilage regeneration. However, many tissues do not contain readily accessible mature cells that can grow outside the body and remain capable of generating their corresponding tissue type. This limitation has led investigators to search for other cell types that are capable of generating functional tissues.

Various “stem cell” types have emerged as a potentially important cell source in tissue engineering (5). These cell types possess several properties that are ideal for regenerative applications, including: (1) an ability to make copies of themselves, or self-renew, which may allow for creation of an endless cell source; (2) the ability to differentiate into multiple mature cell types; and (3) the ability to differentiate in response to environmental cues. These capabilities, in principle, could allow for production of highly complex tissues and organs from a renewable stem cell source. However, the capabilities of stem cells are largely based on observation of natural stem cell activities *in vivo*, and recreation of these activities in engineered systems has been a significant challenge. The following paragraphs delineate the stem cell types in current use in tissue engineering and highlight important challenges in stem cell-based tissue engineering.

Adult Stem Cells. Several investigators are working toward generation of functional tissues using stem cells isolated from a multitude of adult tissues, including skin, skeletal muscle, retina, adipose tissue, dental pulp, blood vessels, and bone marrow. These cell types are often termed tissue-specific stem cells, to distinguish them from the more primitive and pluripotent embryonic stem cells (described in the next section). The bone marrow has been a particularly fruitful cell source, yielding two multipotent cell types: hematopoietic stem cells (HSCs), and bone marrow stromal cells. HSCs were initially defined by their ability to generate all of the mature blood cell types. However, recent studies indicate that these cells can also be coaxed to differentiate into nerve cells or liver cells when delivered to the central nervous system or the liver, respectively. Therefore, HSCs appear to have applications in engineering of several key tissue types. The marrow stromal tissue has also been an intriguing source of stem cell types. Cells from this source have been termed mesenchymal stem cells, bone marrow stromal cells, or marrow-derived mesenchymal stem cells, depending on their isolation and selection procedures. Marrow-derived cells have been used to generate bone, skeletal muscle, cartilage, adipose, and vascular tissues. It is important to note that similar mesenchymal stem cells have also recently been isolated from synovial joints, adipose tissue, and umbilical cord tissue, and have been driven to differentiate into mature cell types, including bone and cartilage cells. A subpopulation of the bone marrow-derived mesenchymal stem cells, which are termed multipotent adult progenitor cells (MAPCs), have recently demonstrated an ability to differentiate into an even more extended range of adult cell

types, including liver, nerve, blood, and lung cells. The identification of these cell types suggests that tissue-specific stem cells, which have historically been considered limited in their differentiation potential, may be capable of transforming into a wider range of mature cell types than anticipated. However, no adult stem cell type has been shown to be pluripotent, and therefore able to give rise to all types of specialized adult cells, and adult cells are more limited in their ability to self-renew in culture (< 60 population doublings) when compared with the more primitive embryonic stem cells (hundreds of population doublings), described in the next section.

Embryonic Stem Cells. Another stem cell type, the human embryonic stem cell (ESC), has also generated a great deal of excitement due to its potential use in tissue engineering applications. Human ESCs are pluripotent cells derived from the inner-cell mass of blastocysts generated via *in vitro* fertilization. These ESCs have also been isolated from embryos created *in vitro* via somatic cell nuclear transfer, which is often termed therapeutic cloning. In addition, similar cell types called embryonic germ cells have been isolated from the fetal gonadal ridge. Each of these cell sources produces cells that are capable of self-renewing for extended periods in culture without differentiating, and they are considered capable of generating all of the mature cell types in the body. It is therefore possible, in principle, to use ESCs as a renewable cell source to engineer any human tissue. Although this potential is intriguing, the use of ESCs in tissue engineering applications has not been extensive. These cells are primitive and often require complex signaling environments to direct them to become a specified mature cell type. Furthermore, the complex signals required to generate specified cell types are incompletely understood. These challenges, along with the social and ethical questions associated with ESC isolation and use, present significant roadblocks to ESC-based tissue engineering. However, recent studies have successfully utilized human ESCs combined with material carriers to generate a variety of adult tissue types. In addition, direct injection of human ESC-derived cell types into pathological sites is an active area of study. These efforts and others suggest that ESCs may be an important component of emerging tissue regeneration approaches.

Challenges. Each of the stem cell types described are being explored for stem cell based tissue engineering approaches, which have included (1) direct transplantation of the cells into a pathological location; (2) transplantation of cells upon or within a biomaterial carrier; or (3) differentiation of the cells in a cell culture bioreactor prior to implantation with or without a biomaterial carrier (Fig. 2) (6). A particular challenge in these stem cell based approaches is delivery of signals to stem cells to direct their differentiation and, in turn, new tissue formation. Stem cells are primitive by nature, and they therefore require instructions, which are provided in the natural stem cell micro-environment (or niche) (7). Many of the emerging approaches to stem cell based tissue regeneration attempt to adopt aspects of the natural stem cell niche to imitate natural tissue development. Design of new biomaterials and new inductive approaches, coupled with new insights about the signals that direct stem cell based tissue formation, may allow for a higher level of control over tissue regeneration in the future.

Biomaterials (8–12)

Tissue engineering strategies include the implantation of material scaffolds alone or in combination with cells. A material that can be used for tissue engineering applications must meet a number of requirements; namely, materials should be biocompatible; biodegradable to nontoxic products within appropriate time frame for application; easily processed to form complex shapes with appropriate porosity; able to support cell growth and proliferation; mechanically suited to an application. There exist numerous classes of biocompatible materials, and design and synthesis of new materials remain active areas of investigation. Several considerations and concerns are common to the design of all biomaterials, regardless of whether the material is derived from natural or synthetic sources.

Naturally Derived Biomaterials (8,10,11). Natural extracellular matrices (ECMs), which are protein-based matrices that surround most cell types in the body, can be considered the quintessential biomaterials. During natural tissue development and repair processes these natural ECMs serve many of the functions that are important for successful tissue engineering, including (1) provision of a space-filling scaffold for infiltration of cells and synthesis of

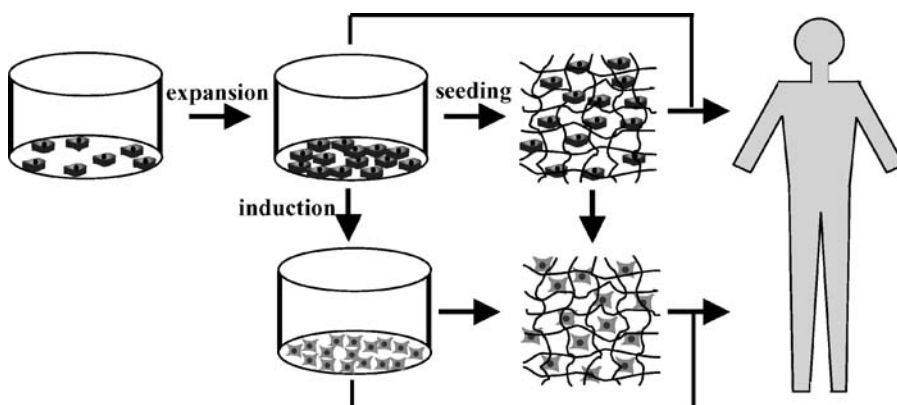


Figure 2. Schematic representation of current approaches to stem cell based tissue engineering. Stem cells are isolated from a variety of tissues, expanded in culture, and implanted into a tissue defect with or without a biomaterial carrier. In some cases, the cells are induced in culture to differentiate into a particular mature cell type prior to implantation.

a new tissue; (2) delivery of signals that influence cell activity during new tissue formation; and (3) the ability to degrade and remodel while a new tissue forms. It is therefore not surprising that a large fraction of biomaterials used in tissue engineering are derived directly from the natural ECM (8). Most of the natural biomaterials used currently are derived from connective tissues, including ligaments, tendons, blood vessels, skin, and bone. More specifically, the most common natural materials in general clinical and scientific practice include collagen (both native and chemically modified), fibrin, hyaluronic acid, and alginate. The following paragraphs provide a brief description of each of these biomaterials.

Collagen. Collagens are a class of vertebrate proteins (with >25 variations) that fold into a characteristic triple helical structure. In skin, tendon, and bone tissues the predominant type of collagen is Type I, and it is this type that is used extensively in tissue engineering applications. Type I collagen can be formed into a cross-linked hydrogel network, a fiber mesh, or a porous sponge, depending on the processing method. Therefore, it is a tremendously flexible base material, and applications have included skin, bone, cartilage, nerve, and liver tissue engineering. Collagen-based materials can also be readily mixed with cells, proteins, or plasmid DNA during processing, which has led to the use of collagen scaffolds in cell-based and inductive tissue engineering strategies.

Fibrin. Fibrin is an integral protein component of clots that form during blood coagulation. It is in this context that fibrin provides a temporary matrix during natural wound healing. Therefore, the natural function of fibrin demonstrates its potential use in wound healing and tissue engineering applications. Fibrin is capable of forming a hydrogel upon protein cross-linking, and these hydrogels adhere strongly to connective tissues, promote cell attachment and wound healing, and degrade in response to protease activity. Therefore, fibrin-based hydrogels serve as excellent matrices for blood vessel ingrowth and healing of connective tissues (e.g., skin, bone). In addition, synthetic biomaterials have been designed to mimic the biological properties of fibrin, including cross-linking, cell adhesion, and protease degradation. These biomimetic approaches highlight the importance of natural biomaterials as templates for design of synthetic biomaterials.

Hyaluronic Acid. A critical component of the extracellular matrix in many tissues, hyaluronic acid (HA) (10) is a relatively simple, yet unusual, high molecular weight polysaccharide. Hyaluronic acid is present in all mammals, playing a vital role in embryonic development, extracellular matrix homeostasis, wound healing, and tissue regeneration. This acid possesses unique biological and mechanical properties due to its hydrophilic, polyanionic composition, and the influence of HA on cell function is highly dependent on its molecular weight. Degradation of HA may occur via several cell-secreted enzymes, or nonenzymatically via free radicals or other various treatments. The intrinsic physicochemical and biological properties of HA have generated widespread use of this molecule in

clinical therapies, and many HA-based products have been approved by the FDA for clinical use in osteoarthritis, ophthalmology, wound healing, gastroenterology, and prevention of postsurgical adhesions. Due to the integral involvement of HA in tissue regeneration and the ease of chemically modifying HA to form numerous derivatives, the use of HA has also been extensively investigated in a wide array of tissue engineering applications. Moreover, as HA plays a crucial role during the morphogenesis of many organs, it may be capable of providing specific signals to cells to initiate tissue or organ regeneration. Derivatization or chemical cross-linking of HA yields HA-based materials with a wide range of mechanical, chemical, and biological properties. These scaffolds can take numerous physical forms including fibrous meshes, sponges, microspheres, and hydrogels.

Alginate. Alginate is a natural polysaccharide isolated from seaweed. In the presence of divalent cations (e.g., calcium) or multivalent polymers (e.g., polylysine) alginate chains become cross-linked into a network hydrogel with intriguing properties. The utility of alginate hydrogels in biomaterials and tissue engineering applications is largely a result of their ability to repel proteins. Proteins do not intrinsically interact with alginate hydrogels and, in turn, cells do not bind to these gels. Therefore, alginate hydrogels provide useful platforms for presentation of specific molecules to cells, allowing these gels to be used as controllable synthetic matrices. In addition, alginate hydrogels do not bring about a substantial immune response, so they are useful in applications designed to shield cells or tissues from the immune system (e.g., pancreatic islet delivery). Another attractive property of these hydrogels is their ability to gel *in situ* without the use of harsh cross-linking agents. This allows for minimally invasive implantation of a material, and may enable minimally invasive cell, protein, and gene delivery. Alginate hydrogels have been used in applications ranging from cell encapsulation to bone tissue engineering.

Other Naturally Derived Materials. Numerous other naturally derived materials have demonstrated promise for tissue engineering applications, including chondroitin sulfate, gelatin, agarose, chitosan, and dextran. A more detailed discussion of the role of naturally derived materials in tissue engineering is available in several reviews (8,10–12).

Limitations. Clearly, natural polymers can be used in a wide variety of tissue engineering applications, as they are generally biocompatible, biodegradable, and easy to process. However, there are limitations to the use of these materials. There is often substantial batch-to-batch variation in the properties of these polymers, which can be dependent on the species and tissue of origin or the harvesting procedure. In addition, because these materials are protein or polysaccharide based they are typically not amenable to standard polymer processing schemes that involve high temperatures or harsh organic solvents. In addition, the nature of protein-based natural materials makes them vulnerable to immune responses *in vivo*,

particularly in cases where the proteins are implanted into a host species that differs from their source. These concerns and others provide an impetus to develop synthetic biomaterials that exploit the advantages of natural materials and limit their disadvantages.

Synthetic Materials (9,11,12). Although synthetic materials lack the inherent bioactivity and biologic recognition possessed by natural substances, their advantages are numerous. In a broad sense, the primary advantage associated with synthetic materials is the researcher's ability to accurately control material characteristics. Controlled alterations in material chemistry, polymerization, or method of scaffold formation can produce changes in the type and time of material degradation, porosity, stiffness–elasticity, texture, shape, hydrophilicity, and bioactive ligand presentation. The ability to construct synthetic materials from scratch enables researchers to tailor material properties to the desired specifications required for a specific application. While many such manipulations are also possible with natural materials, it is widely accepted that synthetic materials are more amenable to modification. This section will discuss several of the most prevalent synthetic materials used in tissue engineering applications.

Poly(α -esters). Poly(glycolic acid) (PGA) and poly(lactic acid) (PLA) remain the most widely used synthetic materials for tissue engineering. These polymers are produced via a ring-opening polymerization of glycolide ($R = H$) or lactide ($R = CH_3$), respectively (Fig. 3).

Poly(glycolic acid) is a highly crystalline, hydrophilic poly(α -ester) that was used as early as the 1960s as the first biodegradable suture material. Simple, random hydrolysis of the ester bonds leads to bulk degradation of PGA, with the degradation rate dependent on the degree of PGA crystallinity. The degradation product of PGA is glycolic acid, which is processed through normal metabolic pathways in the body and ultimately eliminated via the respiratory system as carbon dioxide. Because the degradation products of PGA act as components of the body's natural metabolic pathways, this polymer may be classified as a bioresorbable material. For tissue engineering applications, PGA is most often processed via extrusion to form a 3D, porous mesh structure composed of uniform PGA strands or fibers. Cells seeded on these scaffolds readily adhere to the PGA fibers, then spread, proliferate, and produce extracellular matrix proteins. The PGA has proven to be an excellent carrier for numerous cell types and other biological agents. Due to its rapid degradation and poor mechanical properties, however, use of PGA alone is not suitable for many applications.

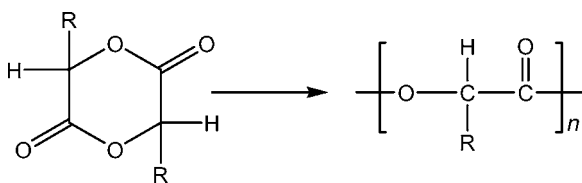


Figure 3. Ring-opening reaction of glycolide ($R = H$) or lactide ($R = CH_3$) to form PGA or PLA, respectively.

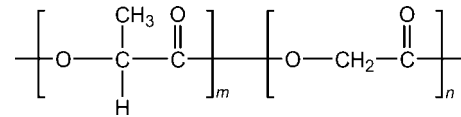


Figure 4. Copolymers of PGA with PLA can be synthesized to form poly(lactic-co-glycolic acid) (PLGA), whose characteristics represent a combination of the two original polymers.

Similar to PGA, PLA is a resorbable poly(α -ester) that degrades via hydrolysis to generate a product (lactic acid) that is readily metabolized by the body. However, PLA is more crystalline and hydrophobic than PGA, resulting in significantly slower degradation characteristics. The PLA is exceptionally strong for a biodegradable material and its superior strength makes it particularly well suited for regeneration of load-bearing tissues (e.g., bone and cartilage).

The synthesis of PLA–PGA copolymers (PLGA, Fig. 4) has yielded tissue engineering scaffolds with a wide range of mechanical and degradation characteristics. Because of their excellent biocompatibility and versatility, PLGA scaffolds have been widely used in tissue engineering applications with much success. The polymers PGA, PLA, and PLGA are all highly processable; scaffolds may be fabricated via extrusion, injection molding, compression molding, or solvent casting. They can be formed into complex structures, and there exist numerous methods for introducing pores. Nanofibrous scaffolds have also been fabricated from PLGA and PLA. These scaffolds mark an interesting advancement in the design of matrices for tissue engineering, as their physical structure (an array of fibers 50–500 nm in diameter) mimics the structure of fibrillar collagen, a component of the native extracellular matrix. A further reason that PGA and PLA remain the materials of choice for tissue engineers is that they are among the few synthetic scaffold materials approved by the FDA. The primary drawbacks of the poly(α -ester) family of polymers are their brittleness and tendency to crumble, lack of chemical functionalities other than end-groups, and production of acidic degradation products that often cause inflammation.

Poly(Anhydrides). Poly(anhydrides) are notable as a scaffold material for tissue engineering in that they undergo surface degradation via hydrolysis of the anhydride group, as opposed to the bulk degradation experienced by poly(α -esters). Poly(anhydrides) are also the first new synthetic degradable material approved by the FDA in >20 years. Poly(sebacic acid) (SA) and poly(*p*-carboxyphenoxyhexane) (CPH) are two commonly synthesized poly(anhydrides) for tissue engineering and drug delivery applications (Fig. 5).

Because homopolymers of SA degrade over a period of days, while polymers based upon CPH degrade over periods of months to years, copolymerization of SA with CPH offers a means of creating poly(anhydride) materials with a wide range of controlled surface-erosion characteristics. Surface-degrading materials have better retention of mechanical strength than those that degrade via bulk mechanisms; scaffold structural integrity is compromised early in the bulk degradation process, and the majority of mass loss occurs toward the end of the degradation process,

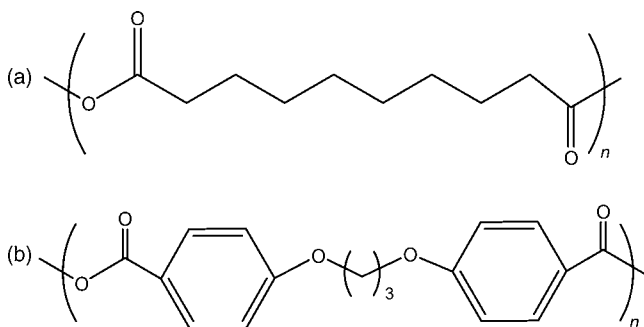


Figure 5. Polyanhydrides, such as SA (a) and CPH (b) are commonly used in tissue engineering and drug delivery applications.

often resulting in inflammation following the sudden burst of acidic degradation products. Such drawbacks make surface-eroding polyanhydrides advantageous for several tissue engineering related applications. Polyanhydrides have proven to be excellent vehicles for drug delivery, and are therefore particularly appropriate for the scaffold-only approach to tissue engineering, where these materials provide a structural platform that elicits *in situ* tissue ingrowth and regeneration through the release of bioactive and tissue-inductive agents. The tissue engineering potential of this class of materials was recently enhanced through the modification of SA and CPH to contain photopolymerizable moieties, meaning that the starting monomers are cross-linked to form scaffolds via exposure to a certain wavelength of light. As discussed in following sections, formation of scaffolds via photopolymerization affords several advantages over many other polymer processing techniques.

Poly(Ethylene Glycol) and Poly(Vinyl Alcohol). Hydrogels synthesized from poly(ethylene glycol) (PEG) or poly(vinyl alcohol) (PVA) represent yet another major class of biomaterials used in tissue engineering applications. An interesting feature of both PEG and PVA is that they are relatively biologically inert, meaning that they are resistant to protein and cell adhesion. While it may appear counterintuitive that nonadhesivity makes these materials attractive for tissue engineering purposes, the reason that tissue engineers value this property of PEG and PVA comes back to the need to control the biomaterial environment. Both PEG and PVA ultimately provide tissue engineers with a “blank” template that can be systematically modified in a controlled manner to possess a wide range of defined chemical, mechanical, and biological properties.

Unlike the polyesters and polyanhydrides described earlier in this section, PEG and PVA are not inherently degradable. However, PEG and PVA can be easily modified to contain either enzymatic or hydrolytic degradation sequences, or both. Thus, not only can the tissue engineer tailor the hydrolytic degradation time for a PEG scaffold, but they may also incorporate moieties that are sensitive to cell-secreted enzymes in order to facilitate creation of a system where scaffold degradation coincides with tissue growth.

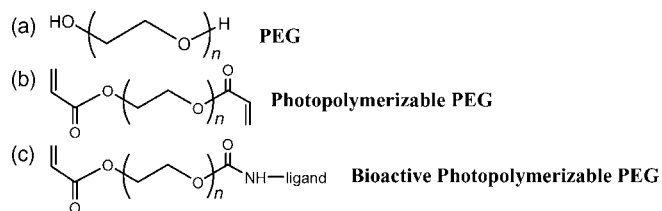


Figure 6. (a) Poly(ethylene glycol) may be chemically modified with vinyl groups (b) to form a material that is polymerized into hydrogel structures via exposure to light (photopolymerized PEG). (c) Bioactive molecules may also be covalently incorporated into these structures by modifying one end group of photoactive PEG with a biological ligand.

Fabrication of PEG and PVA scaffolds is commonly performed via photopolymerization. The end groups of these polymers are modified to contain vinyl groups (Fig. 6) that enable light-induced polymer crosslinking and hydrogel formation under specified conditions. Unlike other processing methods, the conditions of photopolymerization are very mild and can be executed in the presence of cells and sensitive bioactive agents. Photopolymerization may be performed *in situ*, thus offering a minimally invasive method of implanting biomaterials into the site of desired tissue repair. Moreover, the photopolymerized material can assume complex shapes, as it can fill irregularly sized defects *in situ* without complicated molding or shaping techniques. The hydrogel structures that result from photopolymerization are transparent in appearance and possess varying mechanical properties that are dependent on polymer molecular weight and concentration. These materials may also be covalently modified to present various bioactive agents that stimulate cell adhesion to the scaffolds, as discussed in later sections. The PEG and PVA hydrogels possess excellent biocompatibility and are very successful in supporting cell viability, growth, and function. Disadvantages of these materials include their poor tensile strength and relatively small pore size.

Other Synthetic Materials. Numerous other synthetic materials have demonstrated promise for tissue engineering applications, including poly(acrylamides), poly(urethanes), poly(orthoesters), poly(lactones), poly(propylene fumarates), and poly(phosphazenes). A more detailed discussion of the role of synthetic materials in tissue engineering is available in several reviews (9,11,12).

Limitations. Synthetic materials alone cannot provide biological cues to influence cell behavior and tissue formation; these materials serve as only structural supports for cell growth. Thus, it is difficult to recreate a cell's natural biological environment using only synthetic materials. Moreover, issues with material biocompatibility persist. Biocompatibility of scaffolds and their degradation products remains a significant obstacle in the development of new synthetic materials. Furthermore, biocompatibility does not necessarily imply hemocompatibility (blood compatibility), creating further challenges in the creation of materials for blood-contacting applications. Finally, because most synthetic materials degrade via hydrolysis,

tissue engineers continue to struggle with the problem of tailoring material degradation to tissue regeneration such that the material remains for a sufficient period of time to support tissue growth, but not so long as to inhibit tissue formation.

Bioreactors

The previous sections highlighted the importance of isolating cells that are capable of generating new tissues and creating materials that can support new tissue growth. Once the appropriate cell types are procured and material scaffolds are constructed, what is the optimal manner in which to combine these components to generate new tissues? Furthermore, what is the appropriate environment for cultivation of a new tissue as it develops? These are critical questions in tissue engineering approaches, and a variety of bioreactor (13,14) systems have been built to control and optimize cell engraftment cell survival, and proper cell function within biomaterial scaffolds.

Cell Seeding. An initial challenge posed in cell-based tissue engineering schemes involves simply placing cells within a biomaterial construct, a process termed cell seeding. In most strategies, it is important to encourage as many cells as possible to engraft within a biomaterial scaffold to promote generation of new tissues and to avoid squandering valuable cellular components. More efficient cell seeding limits the size of a biopsy needed for cell sourcing and may also reduce the extent of cell expansion necessary to produce adequate cellular components. In addition, studies aiming to generate cartilage, bone, and cardiac tissues have demonstrated that cell seeding density and homogeneity directly influence the growth of new tissues. Higher cell densities lead to enhanced tissue formation, while more homogeneous cell seeding distributions result in more uniform tissue formation. In view of these previous results, investigators have developed a variety of strategies to encourage efficient and homogeneous cell seeding.

The simplest method of cell seeding, termed static seeding, involves simply adding a cell suspension onto a scaffold construct and passively allowing the suspension to permeate the material. This method often leads to inefficient cell seeding and heterogeneous cell distribution. In fact, static seeding approaches often result in growth of only thin $\sim 100 \mu\text{m}$ layers of tissue due to limitations in cell seeding density and homogeneity as well as limited nutrient diffusion. The limitations of this static approach have led to development of more active cell seeding approaches, which include stirred bioreactor systems, direct perfusion bioreactors, and rotating wall bioreactors. In a stirred bioreactor, a dilute suspension of cells is continuously stirred around a stationary, porous scaffold construct, allowing for convective flow. Similarly, direct perfusion bioreactors encourage infiltration of cells by directly flowing a cell suspension through a stationary porous scaffold. Rotating wall bioreactors encourage mixing of cells with scaffolds by rotating the entire bioreactor casing around a central axis. Each of these strategies has enhanced the efficiency and

homogeneity of cell seeding, resulting in more copious and uniform new tissue development.

Mass Transport. Encouragement of mass transport to and from a developing tissue is among the most formidable challenges in tissue engineering. This challenge is particularly daunting *in vitro*, where a developing tissue cannot be exposed to a functional vascular network. Mass transport is vital in tissue engineering approaches, as it provides a means for delivery of oxygen and nutrients to cells within a developing tissue. Tissues that are grown in the absence of facilitated mass transport typically contain a shell of viable tissue surrounding an inner core of necrotic tissue. The thickness of healthy tissue is commonly $< 1 \text{ mm}$, which is not an appropriate scale for a majority of intended tissue engineering applications. To address this limitation several investigators have developed bioreactor systems that encourage mass transport throughout a developing tissue. These systems use mechanisms that are similar to the aforementioned cell seeding bioreactors, and in this case convective flow is used to transport oxygen, nutrients, and wastes during new tissue development. The result is a significant increase in the total amount of tissue grown, both *in vitro* and *in vivo*. Bioreactor systems that enhance mass transport have become important in bone, liver, and skeletal muscle tissue engineering approaches, and the clear benefit of mass transport during tissue development *in vitro* has motivated strategies aimed at increasing mass transport *in vivo* (e.g., by promoting vascular tissue in growth).

Inductive Signaling. Efficient and homogeneous cell seeding and optimal cell survival have a substantial impact on the success of a tissue engineering approach. However, it is also important to expose growing tissues to signals that direct their development into functional tissues. This is particularly crucial in emerging stem cell-based tissue engineering schemes, in which cells must be exposed to signals that direct differentiation and induce formation of the tissue of interest by cells capable of generating multiple mature tissue types. Therefore, there is a need to develop bioreactor systems that provide a range of signals to cells to encourage functional tissue growth *ex vivo*. To that end, investigators have developed bioreactor systems that are capable of delivering both biochemical signals (e.g., protein growth factors) and mechanical forces (e.g., fluid shear forces, compressive forces). As a result, new bioreactor systems allow for growth of tissues that more effectively mimic natural tissue structure and function. Illustrative examples include cartilage, bone, skeletal muscle, ligament, and cardiovascular tissues.

Of course, in many cases the most effective bioreactor system for new tissue development is actually the *in vivo* implantation site. Therefore, an important consideration in tissue engineering is how long, if at all, a cell-scaffold construct should be cultured *ex vivo* prior to implantation. For example, several recent approaches have focused on encouraging vascular tissue infiltration *in vivo* to address the challenge of mass transport to and from a developing tissue. For tissues that are particularly dependent on high oxygen tension (e.g., liver tissue) inducing vascular tissue

ingrowth *in vivo* may be a more effective strategy when compared with convective fluid transport in a bioreactor *ex vivo*. However, for poorly vascularized tissues (e.g., cartilage) that rely heavily on the local mechanical environment during their development, the mechanically and biochemically controlled environments of a bioreactor system may allow for optimization of tissue development *ex vivo*. Other tissue types may benefit from an initial culture period *in vitro* for cell expansion and early tissue development, followed by implantation into a supportive *in vivo* site. The amount of *in vitro* culture time that is appropriate prior to implantation of a tissue engineered construct is an important open question in several approaches.

Mechanical Signaling. The mechanical environment is a key factor during development of several tissue types, and mechanical forces become important as early as the eight-cell stage of embryonic development. Similarly, mechanical forces are important parameters during development of several engineered tissue types. Numerous physical factors have been shown to influence growth of engineered cartilage tissue, including hydrodynamic forces and cyclic mechanical compression and tension. In addition, systems for application of tensile stresses to growing skeletal muscle influence the ultimate contractile properties of the engineered muscle tissue. Mechanical forces also have a clear effect on bone development, repair, and regeneration, and investigators have taken advantage of these effects to build bioreactors that provide mechanical stimulation (see the section Case Studies in Tissue Engineering). To address the importance of mechanics, several investigators have generated instrumentation for application of biomechanical forces to developing tissues. Bioreactors can be designed to include screws to apply static strain or motorized load cells for dynamic actuation, and the biomechanical components are engineered to interface with other bioreactor components, including stirring tools for mass transport and media containing inductive agents. Biomechanical forces have played a particularly prominent role in emerging functional tissue engineering approaches, which often aim to provide factors that directly mimic the characteristics of the corresponding *in vivo* environment and improve the properties of the engineered tissue.

Emerging Designs. Although they are beyond the scope of this article, it is important to note that new concepts are emerging in bioreactor design at the cellular and subcellular scale. Several investigators are actively developing microfluidic systems that are capable of culturing multiple cell types simultaneously in highly controlled environments. These systems are being customized for coculture of multiple cell and tissue types (often termed organ-on-a-chip approaches), as well as highly controlled drug delivery. Microfluidic systems can also be interfaced with biomaterials to provide a powerful platform for highly controlled tissue development *ex vivo*. These approaches and others are enhancing the level of complexity that is possible in bioreactor design, and may allow for more direct mimicry of natural microenvironments.

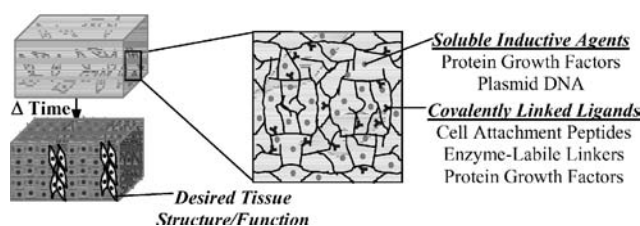


Figure 7. Schematic representation of inductive tissue engineering approaches. Cells are exposed to 3D environments that are engineered to contain covalently linked or soluble agents. The agents are chosen to influence cell activity, ultimately leading to induced growth of a desired tissue.

INDUCTIVE APPROACHES

A variety of investigators have designed 3D matrices to interface directly with mammalian cells, and these materials were described in the biomaterials section above. Both natural (e.g., collagen gels) and synthetic [e.g., poly(L-lactic acid)] materials have been used for decades as scaffolding to support tissue formation from cells *in vitro* or *in vivo*. These materials have historically been passive, in that they have not exerted a high level of control over the signals presented to resident cells. In the past decade, matrices have been designed to interact with cells via specific cell attachment ligands, or designed to deliver soluble signals, so that characteristics of the cell's microenvironment can be tightly controlled (Fig. 7). Bioactive matrices have been used to delineate the effects of specific cell attachment ligands, material properties (e.g., mechanics), and soluble signals on attachment, proliferation, migration, and differentiation of a wide variety of mature cell types. These strategies are often termed inductive approaches, as they are designed to induce a particular cell activity (e.g., tissue formation). The following paragraphs describe inductive schemes that have been used to add a bioactive component to biomaterial matrices.

Covalent Modification of Biomaterials

One widespread approach used to confer bioactivity involves covalently linking biologically active molecules to a biomaterial (15). In this general approach, a biological molecule is engineered to include a chemically reactive group, which is then reacted with a functional group within a biomaterial. The result is a material that has physical properties defined by the original material, and with biological properties defined by the linked biological molecule. This approach has been primarily used to link peptide sequences that promote cell attachment, including RGDS, RGDY, YIGSR, and IKVAV. For example, synthetic PEG hydrogels and natural alginate hydrogels, which do not intrinsically interact with cell surface receptors, can now be readily modified with peptides that promote cell attachment and desired function. In addition to these methods that deliver signals to cells, investigators have also developed materials that respond to cell activity. One example of this general approach is materials that are held together, or cross-linked, by biochemically labile chemistries. These

materials are designed so that their labile linkages are degraded in response to enzymes that are produced by resident cells. Thus, by creatively designing these linkages investigators have been able to generate materials that respond to cell migration and differentiation, resulting in materials that can effectively "listen" to cells. Taken together, the inductive biomaterials created by covalent modification can be considered truly cell interactive, as they are able to deliver signals (e.g., cell adhesion ligands) to cells, and respond to cell activity (e.g., by degrading).

Delivery of Soluble Factors

Another important focus of recent tissue engineering approaches has been delivery of proteins, called growth factors, which can directly modulate cell activity (16). Traditional growth factor delivery approaches have focused on embedding proteins in plastic micron-scale spheres or suspending proteins in highly hydrated gels. In each case, the protein is released via diffusion as the carrier materials degrade, resulting in a high local concentration of the protein. The advent of these technologies has had a revolutionary effect on medicine, and the worldwide market for drug delivery technologies is expected to exceed \$100 billion by 2005. Although these technologies have been useful in a wide variety of biomedical applications, their application to tissue engineering is pragmatically limited. Plastic microspheres do not provide a structural matrix for tissue ingrowth and are difficult to process into structural matrices while maintaining the biological activity of the embedded protein.

Hydrated gels are also nonideal carriers, as growth factors typically diffuse out of the gel matrix rapidly, resulting in limited signaling. Investigators have recently developed innovative approaches to address the limitations of these systems and allow for longer term protein release within structural matrices that support tissue development. For example, plastic microspheres have been engineered to form aggregates with cells or to form a matrix with interconnected pores, and these new strategies allow for inclusion of biologically active proteins. The proteins are then released into the cell population in a sustained manner as the material degrades. These methods are flexible in the type of growth factors that can be included, and have recently been extended to delivery of multiple active growth factors simultaneously. Another recent approach involves covalent immobilization of active growth factors within a hydrogel matrix to locally contain growth factors and limit diffusion out of the hydrogel. This approach has been applied to delivery of vascular and neural growth factors, and could potentially be applied to other proteins, provided they maintain biological activity when covalently immobilized. Taken together, these approaches have been quite successful in actively influencing cell activity within structural matrices during new tissue development. The next generation of protein delivery approaches is likely to exert a higher level of control over where and when cells are exposed to inductive signals. Spatial patterning and tightly regulated timing of protein delivery are routine in natural tissue development, and may be similarly important in complex tissue engineering applications.

Gene Delivery

Another set of inductive approaches does not rely on delivery of peptides or proteins to cells from a synthetic system, but instead focuses on genetic manipulation of cell activity (17). Investigators have developed a variety of strategies to deliver genes to cells, and the genes are typically designed to induce cells to secrete inductive proteins. The gene delivery approaches often utilize the same core technologies that have been used for protein delivery, including sustained release from polymer scaffolds and release from hydrogel matrices, but they substitute plasmid DNA for proteins. Plasmids are circular strands of DNA, which typically contain a single gene under the control of a promoter, which is chosen to enable production of an encoded protein in the target cell type. The encoded protein, often a protein growth factor, is chosen to influence cell activity. Therefore, this approach generates cells that act as bioreactors to produce an inductive protein, and the protein, in turn, encourages new tissue development. This gene delivery approach has been successfully applied to engineering of bone, skeletal muscle and vascular tissues, and the general concept could find broad applicability in engineering of essentially any tissue type.

CASE STUDIES IN TISSUE ENGINEERING

Bone Tissue Engineering Strategies

Vertebrate bone tissue is a dynamic organ with a range of vital functions. Bone serves as a storage depot for mineral ions, a protective barrier for internal organs, and a solid support for muscle actuation. Perhaps the most unique and intriguing property of bone is its ability to continuously regenerate its structure, thereby maintaining consistent structural and mechanical properties. A coordinated set of cell and molecular activities mediates this dynamic remodeling process, and these activities are orchestrated by biochemical, mechanical, cellular influences. The dynamic nature of natural bone tissue makes it perhaps the ultimate "smart material", and the characteristics of natural bone have served as an inspiration to chemists and materials scientists who aim to develop stimulus-responsive and self-healing materials. Furthermore, the constant generation of new bone tissue in natural systems serves as an excellent model for tissue engineering, a field that similarly strives to regenerate natural tissue structure and function. Therefore, bone can be considered an exemplary system for a variety of tissue engineering approaches, including conductive, inductive, and cell-based strategies.

It is perhaps surprising that there is any need for engineered bone tissue at all in view of the efficiency of bone regeneration in vertebrates. However, there are several pathological conditions that result in permanent bone loss or damage. Costs of musculoskeletal conditions represent an average of 3% of the gross domestic product of developed countries (an estimated \$254 billion annually in the United States) and >600,000 inpatient fracture reduction procedures are performed in the United States annually. Furthermore, bone and joint diseases account for one-half of all chronic conditions in people over the age

of 50. The predicted doubling of this age group's population by 2020 suggests that the tremendous need for novel bone repair and replacement therapies will continue to grow rapidly. Regeneration of natural tissue represents a promising approach to replace bone, and could both supplant many of the current metallic hardware-based bone replacement methods and expand the range of bone loss conditions that can be effectively treated. Filling of bone voids in non-union fractures or maxillofacial deformities, bridging of gaps in spine fusion surgeries, and stabilization of vertebral compression fractures provide illustrative potential targets for new bone tissue engineering approaches. Current clinical strategies aimed at repairing or replacing natural bone tissue are typically passive, relying on mechanically sound hardware (e.g., titanium, polyethylene) to replace the structural properties of natural bone. These approaches do not exert a high level of control over the process of new bone formation in a defect site. Limitations in these approaches invoke new therapies for bone replacement, including bone tissue engineering (18–20). The following paragraphs give a brief description of representative tissue engineering approaches that have been developed to repair or replace bone tissue. A more focused and detailed treatment of this topic is given in several outstanding review papers (18–20).

Cell Sources. Osteoblasts are the chief bone-forming cells in natural bone tissue. Based on the regenerative capacity of natural bone tissue it seems obvious that mature osteoblasts would be an excellent candidate as a cell source for bone tissue engineering. These cells can be isolated from the patient (autologous transplantation) or from a donor (allogeneic transplantation), and the typical donor sites are the calvaria or the iliac crest. Investigators have used osteoblasts seeded within various biomaterial scaffolds to engineer bone tissue. However, there are significant limitations to the use of these differentiated cell types. Isolation procedures result in a limited number of autologous or allogeneic osteoblasts, and these cells are difficult to expand in culture. In addition, allogeneic cells harbor the potential to bring about unwanted immune responses. These limitations have led rapidly to the identification and use of alternative cell sources in bone tissue engineering.

The best characterized cell sources in common use in current bone tissue engineering approaches are bone marrow derived fibroblasts. These cells include the aforementioned mesenchymal stem cells, as well as other more committed osteoblast precursors, called preosteoblasts, isolated from bone marrow. These cell types, which are typically distinguished from other cells that reside in the marrow based on their ability to attach and form colonies on standard cell culture substrates, have a long history of use in bone biology and orthopedic regeneration. Between 1968 and 1974, Friedenstein and co-workers (see Ref. 5 for details) undertook a series of studies on the bone-forming capacity of bone marrow derived cells. The studies resulted in isolation of a specific cell type, termed the colony-forming unit fibroblast, which was characterized by its ability to attach to standard cell culture substrates and form isolated colonies in culture. It was not until two decades later that

these cells were further characterized by Caplan and co-workers (see Ref. 5 for details) and called mesenchymal stem cells (MSCs). The pioneering work of Caplan's laboratory in skeletal tissue engineering using MSC sources, coupled with a high profile publication by Pittenger et al. (see Ref. 5 for details) describing the tremendous multipotential of these stem cells, has facilitated use of these cells in bone tissue engineering applications. Recent studies have also identified other adult precursor cells that are capable of forming bone tissue, including skin-, adipose-, and dental pulp-derived mesenchymal stem cells. Investigators are now using each of these cell types as integral components in bone tissue engineering schemes.

Biomaterials. Natural long bones (e.g., the femur) develop upon a cartilage template matrix during a process termed endochondral ossification. In many ways, this natural cartilage template serves as an ideal support for development of new bone tissue. The cartilage matrix: (1) is porous and therefore allows for infiltration of bone-forming cells and vascular tissue; (2) is capable of withstanding the mechanical environment in an orthotopic location; (3) provides a substrate that is conducive to new bone formation; (4) is biodegradable and can be remodeled by infiltrating bone cells; and (5) contains signals that induce new bone formation. These natural characteristics mirror the parameters that are important for design of natural and synthetic scaffolds for bone tissue engineering, including pore structure, mechanical properties, degradability, osteoconductivity, and osteoinductivity. Osteoconductivity is generally defined as the ability of a material to support formation of bone by bone-forming cells (e.g., in a bone defect), while osteoinductivity is defined as the ability of a material to induce formation of bone tissue in conditions that are not otherwise conducive to bone formation (e.g., in a nonorthotopic location). Investigators have developed several classes of biomaterials to address these parameters and to successfully engineer functional bone tissue.

A broad range of natural and synthetic materials have been explored in bone tissue engineering, including poly(anhydrides), poly(fumarates), self-assembling peptide amphiphiles, and alginate hydrogels. Although these materials are well characterized in the context of bone engineering, the most commonly used bone tissue engineering scaffolds have been poly(α -hydroxy esters), type I collagen-based materials, and calcium phosphate based minerals. Each of these base materials can be processed into porous scaffolds that degrade into nontoxic byproducts, allowing for formation of new bone tissue in concert with material degradation. In addition, the mechanical properties of these matrices are dictated by their structure and composition, allowing for significant mechanical tailoring. A particular advantage of calcium phosphate based materials is a property known as bioactivity, which is the ability of these materials to serve as an excellent template for synthesis of mineralized tissue by bone cells and bone precursor cells *in vitro* and *in vivo*. For example, mesenchymal stem cells differentiate and form bone tissue more readily on calcium phosphate materials when compared with other types of scaffold, including polyesters and

protein-based hydrogels, and this phenomenon is attributed to calcium phosphate bioactivity. Many bone tissue engineering approaches exploit this bioactivity by processing or synthesizing new types of calcium phosphate based materials, including both natural (e.g., coralline hydroxyapatite) and synthetic (e.g., sintered hydroxyapatite) scaffolds. Recent approaches have also combined natural and synthetic polymers with calcium phosphate minerals to create hybrid materials for bone tissue engineering. These materials directly mimic the organic–inorganic composite structure of the natural bone extracellular matrix, and they have shown substantial promise as supportive scaffolds for new bone formation.

Bioreactors. *In vitro* engineering of bone tissue has been an active area of study for over a decade. However, early attempts at bone tissue engineering demonstrated limitations that highlight the importance of bioreactor design. Previous studies indicate that growth of bone tissue within 3D scaffolds in static culture conditions is limited to the outer 200–800 μm of a scaffold construct. The poor tissue growth and cell death observed at locations deeper within these scaffolds was likely caused by poor nutrient transport into the developing tissue. These studies and others have provided an impetus to create new bioreactor systems that enhance mass transport and allow for development of more homogeneous bone tissue *in vitro*. Spinner flasks and rotating wall bioreactors have been shown to substantially enhance bone synthesis by mesenchymal stem cells cultured within 3D polyester scaffolds. In addition, perfusion culture systems successfully enhanced generation of bone tissue by osteoblasts and bone marrow-derived precursor cells *in vitro*. In each case, the success of these approaches is attributed to improved nutrient transport within the developing tissue, which is generated by fluid flow.

Other *in vitro* bioreactors have been designed to understand and manipulate the effect of mechanics on engineered bone tissue growth. Mechanical forces have a well-known influence on remodeling of adult bone tissue and, in turn, regulation of bone strength over time. Mechanical stimulation is also vital to proper early development of cartilage and bone, and the mechanical properties of developed bone tissue can be correlated to the load applied during development. In view of these mechanical effects on bone maintenance and development, it is not surprising that mechanical forces have a pronounced effect on bone formation by osteoblasts and osteoblast precursor cells in cell and tissue culture. Culture systems that apply compressive or shear forces to bone-forming cells have shown that the cells respond to mechanical stress by enhancing synthesis of bone matrix.

Inductive Approaches. It is clear that protein growth factors are a vital component of natural bone development and repair processes, due to their effects on bone forming cells and blood vessel ingrowth. Based on these observations, investigators are developing several novel delivery systems that allow for controlled delivery of protein growth factors to bone defect sites to improve or accelerate bone healing. The most potent known growth factors related to

bone regeneration are the class of molecules known as bone morphogenetic proteins (BMPs). Discovery of the unique ability of demineralized bone matrix to induce bone tissue formation in extraskeletal sites led to the isolation and discovery of BMPs as inductive factors. Since their discovery, BMPs have been delivered from several materials to induce formation of new bone tissue in a variety of skeletal and extraskeletal sites. The BMP delivery vehicles examined thus far include demineralized bone matrix, polyester scaffolds, β -tricalcium phosphate, and hydroxyapatite. In each case, BMPs in conjunction with a natural or synthetic carrier material *in vivo* induced dramatic increases in the quantity and functionality of regenerated bone tissue when compared with the carrier materials alone.

Recent inductive approaches have also addressed the importance of angiogenesis in natural bone growth, development, and repair. Bone does not grow, develop, or heal properly when angiogenesis is artificially blocked. However, until recently there had been little interest in specifically inducing ingrowth of a functional vascular supply to support developing bone tissue. Studies have recently demonstrated that delivery of growth factors that induce blood vessel ingrowth (e.g., vascular endothelial growth factor) can substantially increase bone repair and regeneration. These studies provide a new mechanism for actively inducing bone regeneration.

Inductive gene delivery strategies have also been explored in bone tissue engineering. Investigators have delivered plasmid deoxyribonucleic acid (DNA) encoding bone morphogenetic protein-4 (BMP-4) and parathyroid hormone from collagen-based scaffolds. Localized delivery of these plasmids led to the formation of local bone foci similar to natural bone. In addition, multiple studies have achieved transfection of fibroblasts with a gene encoding for bone morphogenetic proteins, resulting in more extensive bone formation. These genetic approaches will likely gain more significance and acceptance with the emergence of more efficient methods for gene transfer, and they could be used in conjunction with novel stem cell based approaches.

Combination Approaches. In this section, bone tissue engineering components are separated into distinct categories to facilitate the readers understanding of the field. However, it is important to note that the vast majority of strategies for bone tissue engineering involve a combination of multiple components, including materials, cells, and biological molecules. In fact, emerging tissue engineering schemes almost exclusively utilize combinations of two or more of the categorized components in this section, and it has become more common for investigators to unite materials, molecules, cells, and highly controlled bioreactor environments to direct bone tissue development. Development of combined approaches requires a more complete understanding of the interdependence of distinct components. For example, the combination of BMPs with a carrier material is critical, and the identity of the carrier material can significantly influence BMP activity. In addition, surface characteristics and geometry of the scaffold material significantly influence induction by BMPs. The complex interplay between substrates and growth factors in synthetic tissue engineering scaffolds emulates the

cross-talk between extracellular matrix signals and soluble signals in natural tissue development and regeneration. Indeed, as the complexity of bone tissue engineering systems increases, researchers may approach the intricacy and control demonstrated during natural bone development and regeneration.

Cardiovascular Tissue Engineering Strategies (21–23)

Cardiovascular disease is a significant cause of morbidity and mortality in the United States and developed countries. Successful treatment has often been limited by the poor performance of synthetic materials utilized for tissue replacement, as hemocompatibility remains a significant challenge in biomaterial design. A lofty goal for cardiovascular tissue engineering is the development of a completely tissue engineered heart. Progress toward this goal will likely be made through the parallel development of effective tissue engineered components of the cardiovascular system. These individual components include blood vessels, heart valves, and cardiac muscle. Each of these structures will be briefly discussed with respect to the cell source, biomaterial, and bioreactor considerations.

Heart Valves. The human heart contains four valves, each with slightly different characteristics and each experiencing a different hemodynamic environment. While heart valves are relatively small, thin structures (on the order of a few hundred microns thick), their composition is surprisingly complex. Valve dysfunction may occur via a variety of mechanisms, and the current treatment for diseased valves is predominantly valve replacement. Over 290,000 people received heart valve replacements in 2003; this number has been steadily rising over the last decade and is expected to reach 850,000 by 2050. Currently available valve substitutes have enabled these patients to experience an enhanced quality of life and have extended patient survival. Yet, in 50–60% of patients with substitute valves, complications associated with these valve replacements necessitate reoperation or cause death within 10 years postoperatively. Present heart valve substitutes consist of either mechanical prosthetic valves or tissue valves that are derived from either human or animal tissue. Perhaps the greatest shortcoming of current valve substitutes is their inability to grow or remodel in response to the physiological environment. Valve replacement is particularly challenging for pediatric patients, who not only outgrow mechanical valves quickly, but also experience rapid calcification (pathological hardening) of transplanted tissue valves. Valve replacement in the elderly has also become more complex; as the average life span increases, but the age at which heart valve disease occurs does not, patients > 65 now need valves that will last 20 years, not just 10. Fabrication of a tissue engineered valve using a biodegradable scaffold may enable the creation of functional valve tissue capable of growth and remodeling in response to changes in its physiological environment. Significant advances have recently been made toward the creation of a functional tissue engineered heart valve via the immobilization of cells within a variety of natural and synthetic matrices, as will be briefly discussed here. A

durable, nonobstructive, nonthrombotic, self-repairing tissue valve that would grow with the patient and remodel in response to *in vivo* stimuli is the current goal for tissue engineered heart valves.

Cell Sources. The source of cells to use in valve regeneration remains a significant concern. Valvular interstitial cells (VICs) comprise the majority of the cell population in heart valves, with a thin layer of endothelial cells (ECs) providing the valve with a nonthrombogenic surface. The adaptive, complex, and dynamic structure of heart valves can be primarily attributed to the VICs, which are responsible for valve extracellular matrix production as they constantly remodel and repair the valve. The organization and relative proportions of the valve matrix are paramount to valve function, thus emphasizing the importance of these interstitial cells. Although these cells readily proliferate and function in *in vitro* culture, they are very difficult to obtain from a patient or living donor. Other mature cell types, such as smooth muscle cells, dermal fibroblasts, and myofibroblasts, can perform some of the functions of VICs, but ultimately do not exactly mimic all of the properties of VICs. Recent studies suggest that marrow stromal cells may be a viable immature cell source for valve tissue engineering. Embryonic stem cells also hold promise as a cell source for VICs, although the method of inducing their differentiation to VICs has not yet been identified.

Construction of fully functional heart valves will also require endothelialization of the valve surface. The ECs play a critical role in both the maintenance of valve homeostasis and the pathogenesis of valvular disease. Mature cell sources for ECs include dermal microvascular endothelial cells (HDMECs), which are isolated from human foreskin, and endothelial cells isolated from human umbilical veins (HUVECs). However, neither of these sources provide autologous cells for a patient in need of valve replacement. Mesenchymal stem cells have recently been shown to differentiate into ECs, indicating that MSCs are a promising source for autologous ECs. Researchers have also identified techniques to induce the differentiation of ECs from embryonic stem cells.

Biomaterials. Construction of a tissue engineered heart valve requires a material that is readily processed into a complex shape, can withstand harsh hemodynamic stresses, and yet is flexible enough to allow for valve opening and closure. To this end, synthetic scaffolds have been constructed using PGA, PLGA, polyhydroxyalkanoate (PHA), or poly-4-hydroxybutyrate (P4HB). Studies using PGA/P4HB in particular have been very successful in large animal models. Much work has also focused upon using decellularized native valves as a scaffold material; in this approach, immunogenicity of a donor valve is lessened by removing the cells, leaving only the extracellular matrix structure. The primary advantage of this strategy is that the appropriate matrix composition of the valve is already present and does not have to be reconstructed. These valve scaffolds can then be recellularized with autologous (non-immogenic) cells, with the goal of restoring the valve's regenerative capacity. Finally, hydrogels made from

several naturally derived materials, such as collagen or hyaluronic acid (HA), have also been used for heart valve tissue engineering. Although HA-based valve scaffolds have not yet progressed to *in vivo* studies, the use of HA is particularly exciting, as this polysaccharide is required for cardiac morphogenesis and native heart valve formation.

Bioreactors. Because it is difficult to survive with only a partially formed or semifunctional heart valve, tissue engineered matrices must first be cultured in a bioreactor environment prior to implantation. Culture of tissue engineered valves in bioreactors aims to create mechanically stable and reliable valve tissue with significant (or complete) degradation of the material scaffold prior to implantation. In addition to providing appropriate nutrients to the cell-seeded scaffolds, such bioreactors are intended to provide physiologically relevant flow and shear stresses in order to condition the tissue for *in vivo* implantation and function. Furthermore, tissue engineered heart valves exposed to pulsatile fluid shear stresses *in vitro* display improved function over constructs cultured in bioreactors without a mechanical signal. Investigators working on tissue engineered heart valves have made significant advances in bioreactor design, and these bioreactors may have utility for the culture of other tissues that require pulsatile stresses.

Cardiac Muscle. Because adult cardiac muscle cells (cardiomyocytes) cannot regenerate, myocardial damage induced by a myocardial infarction (heart attack) or other injury results in loss of cardiac function and progressive deterioration leading to congestive heart failure. Necrotic cardiomyocytes in infarcted ventricular tissue are replaced by fibroblasts, leading to the formation of scar tissue and creating regional contractile dysfunction. Current treatments consist of mechanical support using left ventricular assist devices and, ultimately, heart transplantation. The incidence of heart failure at 1 year postheart attack is >20%, and the 1 year postheart attack mortality rate (~30%) did not change from 1975–1995, highlighting the lack of effective treatments to combat this problem. Multiple tissue engineering strategies, including cell-only transplantation, inductive approaches, and implantation of cell-seeded scaffolds have been investigated in order to generate a viable method of repairing damaged heart tissue.

Cell Sources. Cardiomyocytes (CMs) are highly differentiated cells that comprise 70–80% of the heart's mass, yet only 20–30% of the total cardiac cell number. The remaining cell population consists of cardiac fibroblasts and endothelial cells, although CMs are primarily responsible for the contractile activity of the heart. Electromechanical coupling between CMs enables the transduction of electrical signals into muscular contractions. Use of mature autologous cells is feasible for some tissue repair applications, but is likely not possible for myocardial repair, primarily because adult CMs do not proliferate, and removal of any cardiac tissue to obtain CM progenitor cells could impair the function of the already-injured heart.

Cellular cardiomyoplasty (CCM) is a promising therapy that has recently emerged for the repair of damaged cardiac muscle and involves the injection of a suspension of cardiac-related cells into the injured heart (cell transplantation). The therapeutic goal of CCM is the replacement of dead heart muscle with functionally competent and contractile myocardium. Sources for these injected cells have included skeletal myoblasts, fetal cardiomyocytes, hematopoietic stem cells, mesenchymal stem cells, embryonic stem cells, and endothelial progenitor cells, yet the functional fate of these cells following transplantation remains unresolved, and normal electromechanical coupling between implanted cells and host myocardium has been absent. While CCM has been shown to augment myocardial function, there exists debate about whether the transplanted cells are actually functioning as cardiomyocytes. Strategies that do not involve regeneration of cardiomyocytes are limited to rescuing injured tissue and are unable to contribute directly to the restoration of contractile function or increase systolic function. Hence, if the transplanted cells are not differentiating into cardiomyocytes, then contractile, electrically coupled, fully integrated cardiac tissue is not being regenerated, leading to suboptimal improvement of cardiac function. Furthermore, because only 20% of the heart's cells are CMs, stem cells injected into the heart simultaneously receive developmental cues from many different cell types, making it difficult to predict the differentiation pathway that the transplanted cells will follow. Because researchers currently cannot regulate the differentiation factors to which cells are exposed *in vivo*, transplantation of stem cells that are committed to the cardiomyocyte lineage may be a more effective method of regenerating CMs *in vivo*. These cell sourcing issues observed with cellular cardiomyoplasty also extend to the tissue engineering strategy of implanting cell-seeded biomaterial constructs.

Biomaterials. While cellular cardiomyoplasty has been successful in augmenting cardiac function, it has still not been proven to reliably regenerate cardiac muscle. Thus, many researchers are combining cardiac-related cells with biomaterials in order to develop an approach that enables greater control over tissue formation. Numerous synthetic, natural, and biomimetic materials have demonstrated promise for cardiac tissue engineering applications. Seeding of cardiomyocytes on PGA- or PLGA-based materials has produced cardiac grafts whose *in vitro* structural and electrophysiological properties approach those of native heart muscle. However, the acidic degradation products and brittleness of these materials have hindered their *in vivo* success. Several collagen-based matrices have been used to create small myocardial structures that morphologically resemble and possess many of the properties of native myocardium, while alginate matrices also show good promise as scaffold materials for cardiac repair. Bioactive or inductive materials may also be used in a scaffold-only approach for cardiac tissue regeneration; recent studies have demonstrated the existence of a small number of cardiac progenitor cells within heart muscle, and implantation of a bioactive material containing appropriate agents may recruit these myocardial progenitors to the site of injury. While the number of polymers that have

been used to develop cardiac muscle is small at this point compared to the number used for other tissues, (e.g., bone, skin, and blood vessel) this number is likely to increase over the next decade as the push to develop a completely tissue engineered heart increases.

Bioreactors. *In vitro* development of cardiac muscle requires the use of appropriate nutrient delivery systems. Cell-seeded constructs cultured in perfusion bioreactors contain more uniformly distributed cells with enhanced differentiation. While direct perfusion of cell-scaffold constructs has been shown to be beneficial, culturing the constructs under laminar flow rather than turbulent flow has also been found to enhance engineered cardiac muscle. Laminar flow conditions can be provided by using a rotating bioreactor rather than a spinner flask. Another approach for *in vitro* culture of cardiac muscle is to subject cell-scaffold constructs to mechanical stimuli, such as pulsatile flow and cyclic stretch, as would be experienced *in vivo*. Bioreactors that combine perfusion mechanisms with mechanical signals, such as compression or stretching, significantly improve cardiac tissue formation and function over perfusion-only systems. The nutrient requirements for cardiac tissue are quite high, and this tissue is particularly sensitive to ischemic conditions. The thickness of a tissue engineered myocardial construct that can be developed using current techniques is limited to 100 μm , contrasting the 1 cm thickness of native heart muscle. Recent developments in the design of bioreactors for heart valves, which employ adjustable pulsatile flow and varying levels of pressure, may be applicable to the culture of engineered myocardium.

Large Diameter Blood Vessels (>6 mm). Vascular disease is a prominent problem in the United States, with approximately 500,000 coronary artery bypass surgeries performed each year. Natural tissues, primarily saphenous veins or internal mammary arteries, are generally used for coronary artery replacement. In general, the results have been quite favorable for these procedures. Unfortunately, as discussed earlier, the availability of donor vessels and issues with donor site morbidity often preclude the use of natural tissue grafts. Moreover, grafted vessels are often unable to adjust to the increased pressure and wall shear stress in the grafted position, resulting in inadequate vessel performance. Blood vessel replacements composed of entirely synthetic materials were first developed in the 1950s using polymers, such as polyesters, polyethylene terephthalate (PET, Dacron), and expanded polytetrafluoroethylene (ePTFE, Gore-Tex). However, implantation of these materials is permanent, with no regeneration of the biological blood vessel. Furthermore, several problems such as platelet adhesion and activation, and decreased mechanical compliance compared with adjacent arterial tissue, are also frequently associated with synthetic grafts. These problems have motivated investigations of tissue engineered alternatives.

Cell Sources. Blood vessels consist of three cell types: fibroblasts, smooth muscle cells, and endothelial cells. Most tissue engineering endeavors focus upon the smooth muscle and endothelial cells, which are generally obtained

from harvested autologous blood vessels. However, issues, such as donor site morbidity, vessel availability, and performance of an invasive surgery complicate the use of these mature, autologous cells. A solution to these problems may lie in the use of mesenchymal stem cells, which are capable of differentiating into vascular smooth muscle cells and endothelial cells, and may therefore represent an excellent cell source for vascular tissue engineering.

Biomaterials. Attempts to tissue engineer blood vessels date back to the 1970s, and the number of different materials that have since been used in these endeavors is numerous. Construction of a tissue engineered blood vessel requires a material that is resistant to platelet adhesion (nonthrombogenic), can withstand transmural pressure acting normal to the vessel wall and tangential shear stresses, and has similar mechanical properties to that of the adjoining native vessel. Much research has focused upon the use of collagen gels for vascular tissue engineering. Although collagen gels have poor mechanical properties, proper alignment and function of vascular cells in these gels has been observed. The PGA Polymer has been used alone and in combination with other synthetic materials (polyhydroxyalkanoate) to form tissue engineered vessels that remained functional for several months following implantation in an animal model. Recent research has focused upon the use of PEG hydrogels as vascular tissue engineering scaffolds, as the physical and biological properties of these materials are well suited for vascular applications. Finally, a unique vascular tissue engineering approach used no material, but instead rolled sheets of vascular cells into tubular shapes and observed excellent mechanical and functional properties of the tissue.

Bioreactors. With respect to shear stresses, flow, and application of appropriate mechanical signals, the bioreactor culture of tissue engineered vascular grafts possesses many similarities to the other cardiovascular applications discussed above. Application of fluid shear stresses is necessary in order to generate the cellular alignment found in native vessels and to stimulate proper function of vascular cell types. Because *ex vivo* studies of native intact vessels often have similar culture requirements as tissue engineered vascular grafts, a diverse range of researchers has been involved in the development of such bioreactors, yielding numerous configurations and approaches. Variables in the design of vascular bioreactors include constant or pulsatile flow through the vessel lumen, application of longitudinal strain, and perfusion of exterior surface of the vessel.

BIBLIOGRAPHY

1. Langer R, Vacanti JP. Tissue engineering. *Science* 1993;260:920–926.
2. National Science Foundation. Report: The emergence of tissue engineering as a research field; 2003.
3. Saltzman WM, Tissue engineering: Engineering principles for the design of replacement organs and tissues. New York: Oxford University Press; 2004.
4. Lavik E, Langer R. Tissue engineering: Current state and perspectives. *Appl Microbiol Biotechnol* 2004;65:1–8.

5. Lanza R, et al., editors. Handbook of Stem Cells. Volumes 1 and 2. New York: Elsevier Academic Press; 2004.
6. Zandstra PW, Nagy A. Stem cell bioengineering. *Ann Rev Biomed Eng* 2001;3:275–305.
7. Li L, Xie T. Stem cell niche: structure and function. *Ann Rev Cell Dev Biol* 2005;21:605–631.
8. Yannas IV. Natural materials. In: Ratner BD, Hoffman AS, Schoen FJ, Lemons JE, editors. *Biomaterials Science*. New York: Elsevier Academic Press; 1996.
9. Lee KY, Mooney DJ. Hydrogels for tissue engineering. *Chem Rev* 2001;101:1869–1879.
10. Baier Leach J, et al., editors. *Encyclopedia of Biomaterials and Biomedical Engineering*. New York: Marcel Dekker; 2004.
11. Cannizzaro S, Langer R. Biomaterials: Synthetic and engineering strategies. In: Dillow AK, Lowman AM, editors. *Biomimetic Materials and Design: Biointerfacial Strategies, Tissue Engineering and Targeted Drug Delivery*. New York: Marcel Dekker; 2002.
12. Thomson RC, Ishaug SL, Mikos AG, Langer R. Polymers for biological systems In: Meyers RA, editor. *Encyclopedia of Molecular Biology: Fundamentals and Applications*. New York: VCH Publishers; 1996.
13. Ratcliffe A, Niklason LE. Bioreactors and bioprocessing for tissue engineering. *Ann NY Acad Sci* 2002;961:210–215.
14. Martin I, Wendt D, Heberer M. The role of bioreactors in tissue engineering. *Trends Biotechnol* 2004;22:80–86.
15. Lutolf MP, Hubbell JA. Synthetic biomaterials as instructive extracellular microenvironments for morphogenesis in tissue engineering. *Nat Biotechnol* 2005;23:47–55.
16. Saltzman WM, Olbricht WL. Building drug delivery into tissue engineering. *Nat Rev Drug Discov* 2002;1:177–186.
17. Bonadio J, Goldstein SA, Levy RJ. Gene therapy for tissue repair and regeneration. *Adv Drug Del Rev* 1998;33:53–69.
18. Salgado AJ, Coutinho OP, Reis RL. Bone tissue engineering: State of the art and future trends. *Macromol Biosci* 2004;4:743–765.
19. Jadlowiec JA, Celil AB, Hollinger JO. Bone tissue engineering: Recent advances and promising therapeutic agents. *Exp Opin Biol Ther* 2003;3:409–423.
20. Green D, Walsh D, Mann S, Oreffo ROC. The potential of biomimesis in bone tissue engineering: Lessons from the design and synthesis of invertebrate skeletons. *Bone* 2002;30:810–815.
21. Leor J, Amsalem Y, Cohen S. Cells, scaffolds, and molecules for myocardial tissue engineering. *Pharmacol Ther* 2005;105:151–163.
22. Masters KS, Mann BK. Tissue engineered heart. In: Nedovic V, Willaert R, editors. *Applications of Cell Immobilisation Biotechnology*. New York: Kluwer Academic; 2005.
23. Mann BK, West JL. Tissue engineering in the cardiovascular system: Progress toward a tissue engineered heart. *Anat Rec* 2001;263:367–371.

See also BIOMATERIALS: TISSUE ENGINEERING AND SCAFFOLDS; BIOSURFACE ENGINEERING; ENGINEERED TISSUE; SKIN TISSUE ENGINEERING FOR REGENERATION; SKIN SUBSTITUTE FOR BURNS, BIOACTIVE.

TISSUE ENGINEERING. See BIOMATERIALS: TISSUE ENGINEERING AND SCAFFOLDS.

TISSUE EQUIVALENTS. See PHANTOM MATERIALS IN RADIOLOGY.

TMJ. See TOOTH AND JAW, BIOMECHANICS OF.

TOMOTHERAPY

TIMOTHY HOLMES
St. Agnes Cancer Center
Baltimore, Maryland

THOMAS R. MACKIE
University of Wisconsin
Madison, Wisconsin

INTRODUCTION

Tomotherapy is intensity-modulated rotational radiotherapy utilizing a photon fan beam (1). The term tomotherapy derives from *tomographic radiotherapy*, literally meaning “slice” radiotherapy. Tomotherapy treatment delivery is conceptually similar to computerized tomographic (CT) imaging where a three-dimensional (3D) image volume is acquired as a stack of two-dimensional (2D) cross-sectional images. By analogy, a tumor volume can be subdivided into a stack of slices that are independently irradiated to achieve a 3D conformal dose distribution. Dose conformation is achieved by modulating the intensity pattern of the incident X-ray beam during rotation of the X-ray source using a fan-beam multileaf (MLC) collimator whose leaves are pneumatically driven to achieve near instantaneous leaf transitions between open and closed states. As shown in Fig. 1, the table motion can be incrementally stepped the width of the fan beam after each completed arc (serial tomotherapy) (2,3) or it can be

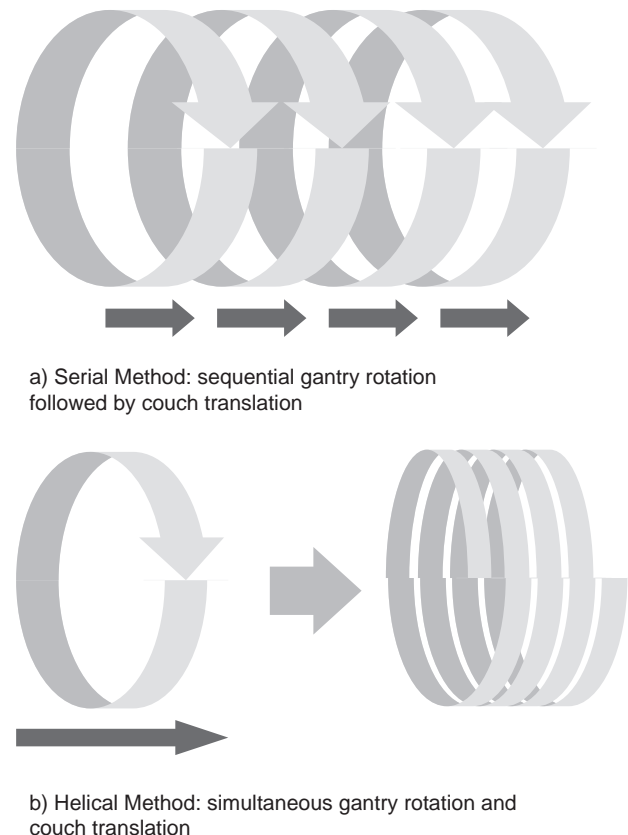
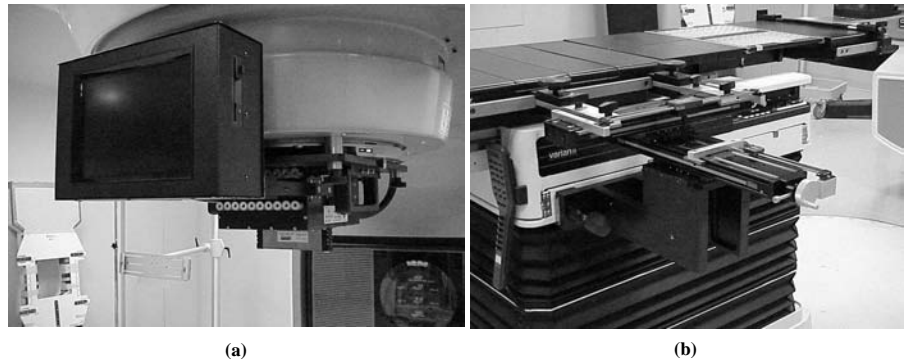


Figure 1. Tomotherapy methods: (a) serial tomotherapy, (b) helical tomotherapy.

Figure 2. Components of the NOMOS Peacock System: (a) gantry mounted MIMiC binary MLC and controller, and (b) table mounted precision positioner for accurately stepping the table between gantry rotations. (Photos courtesy Jim Hevzi, PhD, Cancer Treatment and Research Center, San Antonio, TX.)



translated simultaneously with continuous source rotation (helical tomotherapy) (4).

Two types of tomotherapy, serial and helical, have been developed and implemented clinically. Serial tomotherapy, embodied in the NOMOS Peacock System (North American Scientific - NOMOS Radiation Oncology Division, Cranberry Township, PA) is an add-on component for conventional C-arm medical linac gantries (Fig. 2a). It is feasible to deliver non-coplanar serial tomotherapy treatments using this implementation. Alternatively, the Hi-ART II helical tomotherapy system (TomoTherapy, Inc., Madison, WI) is a dedicated radiotherapy treatment unit built upon a helical CT ring gantry. The constraint of a ring gantry is minimal since few patients are treated with noncoplanar radiation fields and IMRT diminishes the need for these types of field arrangements. Most importantly, a ring gantry is a very stable platform for CT scanning and is used in all diagnostic CT scanners.

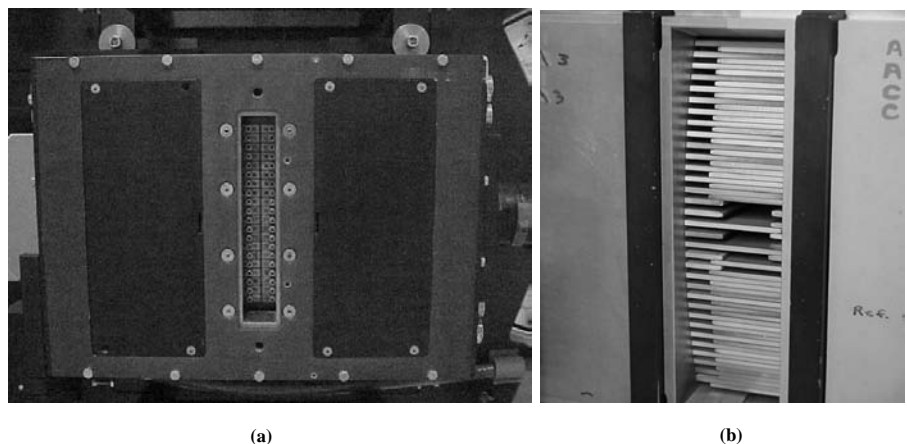
Both tomotherapy implementations have a high level of system integration with common components being (1) an inverse treatment planning system, and (2) computer controlled beam modulation using a pneumatically driven multileaf collimator (Fig. 3). One major difference between the two implementations is the integration of MVCT image guidance into the treatment delivery process of the Hi-ART II (5). The MVCT imaging is primarily used for daily target location to improve precision and accuracy of IMRT treatments. In addition, the detector can be used for machine quality assurance, and for reconstructing daily estimates of delivered dose using measured profiles of X-ray fluence

transmitted through the patient during treatment delivery (6). This latter capability, called *dose reconstruction* is useful to the process of *adaptive radiotherapy* wherein a daily discrepancy between planned and delivered dose due to errors in patient setup or changes in gross tumor volume can be corrected in future treatments.

HISTORICAL BACKGROUND ON THE DEVELOPMENT OF TOMOTHERAPY

Rotational treatment techniques were very popular in the era of orthovoltage (250 kVp) X rays prior to the development of megavoltage photon sources. Rotational delivery allowed improved dose at the tumor relative to the surface compared to treatments using a few fixed-beams portals. This was a consequence of the peak dose for orthovoltage X rays occurring at the skin surface, which for nonrotational deliveries was the dose-limiting feature of the treatment. The development of high energy photon beams from Co-60 and medical linear accelerators reduced the need for rotational treatments since the peak dose was shifted below the skin surface resulting in improved skin sparing when few field techniques were used. As a consequence, treatment techniques shifted almost exclusively to simpler few field beam arrangements, typically of opposed beams, that were easy to verify using planar radiography. This in turn eventually led to the development of dual energy medical linear accelerators, or linacs, with two X-ray energies: a low X-ray energy (4, 6 MV) for treatments in the

Figure 3. Tomotherapy binary MLCs: (a) the NOMOS MIMiC design has 20 pairs of opposed leaves that allow two slices to be treated per rotation. (Photo courtesy Jim Hevzi, PhD, Cancer Treatment and Research Center, San Antonio, TX). (b) The TomoTherapy binary MLC design consists of 64 interdigitating leaves that are used to treat a single slice per rotation. (Photo courtesy TomoTherapy, Inc, Madison, WI.)



head–neck region, and a high X-ray energy (10 MV or greater) for treating deep seated lesions in the trunk and pelvis. In situations where a single low energy machine was the only treatment unit available, rotational techniques were often used for deep-seated lesions, owing to the fact that the relative shape of the rotational dose distribution is insensitive to the X-ray energy used.

In 1988, Anders Brahme of the Karolinska Institute published an article that described the advantages of using nonuniform beam profiles to achieve conformal dose distributions of arbitrary shape (7). A key feature of this work was the application of a mathematical optimization method to determine the nonuniform profiles for few-field delivery. Subsequent efforts by Bortfeld et al. (8), Webb (9), and Holmes et al. (10) applied iterative tomographic image reconstruction methods to inverse planning of intensity modulated beams with the latter two investigations providing methods applicable to rotational IMRT. Webb's application of the simulated annealing algorithm for planning rotational IMRT (9) was implemented in the first commercial inverse treatment planning system, PeacockPlan (NOMOS Corp., Sewickley, PA), to support serial tomotherapy. The work of Holmes et al. (10,11) and Swerdloff et al. (12) under the direction of Dr. T. R. Mackie at the University of Wisconsin Department of Medical Physics began the formulation of a dedicated helical tomotherapy unit (1).

SERIAL TOMOTHERAPY: NOMOS PEACOCK SYSTEM

The Peacock System was developed during the early 1990s by the NOMOS Corporation under the leadership of Mark Carol, a neurosurgeon who became interested in the radiation therapy problem during his medical training (2,3). This system was designed as an add-on device to existing medical linear accelerators, many of which in the early 1990s lacked the features of a computer controlled MLC and a record and verify system needed to perform IMRT delivery. The first serial tomotherapy treatments were performed at the Methodist Hospital, Baylor College of Medicine, in March of 1994 under an Investigational Device Exemption with clearance from the Food and Drug Administration (FDA) coming in 1996. Since then, >100 systems have been installed clinically with over 10 thousand IMRT treatments completed successfully.

The Peacock System consists of four components: (1) CORVUS: an inverse treatment planning system; (2) MIMiC: a computer controlled multileaf collimator; (3) the controller: a dual computer system for control and continuous checking of the MIMiC; and (4) the Auto-Crane: a computer controlled table positioner. The latter three components are shown in Fig. 2.

CORVUS was the first commercial implementation of inverse treatment planning for IMRT. The planning system consists of (1) tools for import and fusion of CT and MRI images for subsequent contouring of treatment and avoidance regions including organs at risk; (2) a finite-size pencil beam 3D dose computation model; (3) an optimization engine based on simulated annealing and gradient-based optimization algorithms; (4) algorithms to convert optimized X-ray intensity maps to multileaf collimator field

shapes and their associated treatment times; and (5) tools to map patient treatments onto dosimetry phantoms for treatment plan delivery verification.

The MIMiC is a binary multileaf collimator. It consists of 20 pairs of opposed tungsten leaves that are 8 cm tall and project to nominally 1 cm wide at 100 cm from the radiation source. The leaves move quickly in and out of a fan beam to provide the intensity modulation. The motive power for the leaves was compressed air pushing and pulling on pistons. The binary collimator was licensed from the University of Wisconsin patent invented by Swerdloff et al. (12). The opposed leaf design allows two adjacent treatment slices to be treated simultaneously during a single arc producing a more efficient delivery. The treatment slice width can be set to 1 or 2 cm using mechanical stops, or to 0.4 cm using a tertiary collimator called the BEAK. The maximum size of the "modulated" volume is a 20 cm diameter cylinder whose length is determined by the longitudinal motion of the treatment table.

Tomotherapy treatments typically require an order of magnitude greater number of monitor units than conventional 3D conformal radiotherapy, consequently the leakage characteristics of the collimation system and the treatment unit are of concern as they influence the whole body dose equivalent received by the patient and the risk of a second malignancy that might occur as a consequence of treatment (13,14). Leakage transmission through a MIMiC leaf and through the gap between leaves is <1%; the leakage transmission outside of the defined slice is determined by the leakage characteristics of the treatment unit jaws and head shielding, which is typically 0.1%. These leakage characteristics were improved upon in the development of the dedicated helical tomotherapy unit described in the next section.

Attached to the MIMiC housing is a pc-controller. This device is a dedicated record and verification system used to monitor the (1) the angular position of the treatment gantry using onboard inclinometers, and (2) the open and closed state and transition times (<140 ms maximum) of the MIMiC leaves during treatment delivery. During treatment delivery, the gantry angle defines the open–close state of each leaf as defined by the treatment plan, which is created assuming constant dose rate and gantry rotation speed. Unacceptable changes in gantry speed will be detected by the controller causing treatment interruption by tripping the interlock circuit of the treatment room door thereby terminating the radiation beam. The controller also verifies the integrity of the plan data, which is stored on a 3.5 in. password protected floppy disk that is originally created by CORVUS as a means of transferring the plan to the controller. Data integrity is ensured using CRC check sums of the data files on the floppy disk. Modification of a plan requires the user to decommission the floppy disk using a password while inserted in the CORVUS planning station.

A continuous gantry rotation is modeled as a set of fixed beams spaced at 10° intervals from ~±160° from the vertical gantry position. A 10–15° interval is required at startup to allow the gantry speed and beam output to stabilize before leaves are opened to modulate the beam. Intensity modulation is achieved by opening a leaf for a

fraction of the interval with opening and closing occurring at angles symmetrical to the center of the interval. For example, a 50% modulation is achieved by opening the leaf at -2.5° and closing it at $+2.5^\circ$ about the center of a 10° interval. Typically, 10 intensity levels are used to approximate an intensity profile so leaf transitions occur at 1° boundaries within the 10° interval.

Serial tomotherapy requires precise positioning of adjacent treatment slices at the submillimeter level to avoid unacceptable dosimetry errors in the abutment region between slices (15). In recognition of this issue, the initial implementation of the Peacock System used a manual table positioner called the Crane that was a professional grade photography studio stand (Cambo BV, Kampen, The Netherlands) modified to attach to the treatment table side rails with table position determined by high precision digital linear scales. This device was subsequently replaced by a computer-controlled positioning device called the Auto-Crane, that was originally developed as the Xlator by the University of Texas Health Sciences Center, San Antonio, TX (16).

HELICAL TOMOTHERAPY: TOMOTHERAPY, INC. HI-ART II

Figure 4 shows a TomoTherapy Hi-ART II tomotherapy unit (TomoTherapy, Inc., Madison, WI) with its covers off and its various subsystems labeled. The linac is a 30 cm long, 6 MV, S-band (nominal 3 GHz) magnetron-powered device with a gridded-gun and a solid-state modulator. The linac is used for both CT imaging and treatment delivery, with imaging carried out using a lower beam current to reduce patient dose, and lower X-ray energy to improve image contrast. Given its compact size, the linac is aligned parallel to the beam axis with the flattening filter removed to increase output to 8–10 Gy/min at the gantry rotation axis located 85 cm from the fixed target. Elimination of the field-flattening filter improves the energy spectrum of the

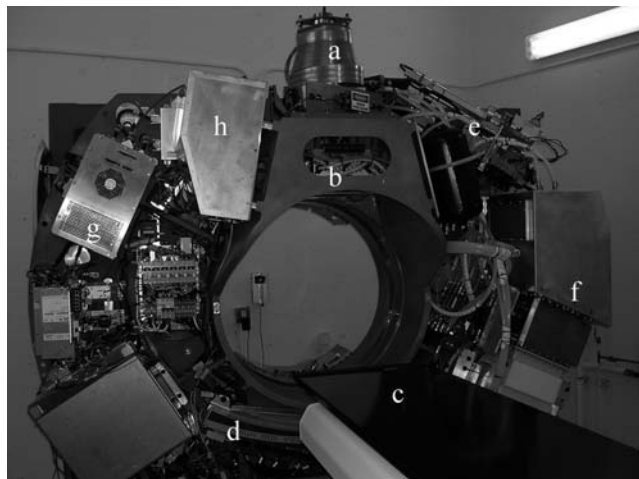


Figure 4. Helical tomotherapy unit in a factory test cell. The major components are: (a) linac, (b) binary MLC, (c) couch, (d) detector, (e) RF system, (f) modulator, (g) gantry control, and (h) RF power.

beam and reduces scatter outside the field boundary, thereby simplifying treatment beam modeling and CT image reconstruction.

In recognition of the increased number of monitor units required for tomotherapy delivery, the treatment head shielding was designed to limit the primary leakage to 0.01% of the primary beam, or one-tenth the limit (0.1%) used for conventional radiotherapy treatments. The shielding includes integrated primary collimators that are used to define field widths from 5 to 50 mm (nominal). Below the primary collimators resides the binary multileaf collimator module, consisting of 64 leaves made from 10 cm high tungsten with leakage $<0.5\%$. A 13 cm thick lead counter weight is attached to the ring gantry opposite the treatment head that acts as a rotating primary barrier or beam stop.

Each binary MLC leaf completely blocks a portion of the fanbeam with a projected shadow of 6.25 mm (nominal) at the gantry rotation axis. The 64 leaves define a 40 cm diameter treatment field of view, which combined with up to 160 cm (nominal) couch travel enables very large treatment volumes to receive IMRT. Intensity modulation is achieved using pneumatic control of the binary MLC leaves by rapidly (~ 20 ms) switching the open-closed state of leaves during gantry rotation. As in the MIMiC collimator, the intensity level is proportional to the time a leaf is open and effectively there are ~ 50 intensity levels that that can be delivered.

The linac and gantry systems of the tomotherapy system are highly favorable for CT imaging where mechanical stability of the source-detector positions during rotation and a small source size are desirable. The gantry sag of the tomotherapy system is ~ 0.1 mm so no sag corrections are required in the CT reconstruction algorithm. The size of the electron beam on the target is ~ 1 mm so that the resolution is ~ 1.2 – 1.6 mm, which is comparable to a conventional CT scanner for high contrast objects. Operating at an average dose to the patient of 1–3 cGy, the images produced have soft tissue contrast of 2–3%, which is poorer than a modern CT scanner, yet are of sufficient quality for adaptive radiotherapy processes. The tomotherapy unit's xenon gas detector elements have tungsten septa separating ionization cavities. In addition to the ionization collectors, the tungsten plates are embedded photon converters intercepting the megavoltage photons and yet are thin enough to let an appreciable fraction of the electrons set in motion to deposit energy in the xenon gas. The interception of the beam by the tungsten means that the quantum efficiency of the system is $\sim 25\%$, which is much more than the few percent collection efficiency of modern portal imaging systems.

Modeling the treatment delivery process requires discretization of the continuous motions of the gantry and table as well as the continuous intensities of the modulated beams. Proper sampling reduces the chance for computational aliasing that can produce “streak” or “thread” artifacts in the dose distribution (17). Consequently, each 360° gantry rotation is modeled as 51 beams spaced at 7.06° apart: a number chosen to allow a 40 cm diameter target volume to be homogeneously treated with a 2.5 cm completely blocked central avoidance structure (18). Following

optimization, the intensity levels are discretized for treatment delivery, with 50 levels chosen to reduce the uncertainty in the target dose due to intensity discretization to $<0.1\%$. Discretization of couch travel is determined by the pitch ratio: the ratio of the couch travel distance per rotation to the field width defined at the axis. In helical tomotherapy delivery, the pitch is usually set to be less than one-half to avoid thread-like dose artifacts developing near the edge of the field becoming clinically significant (17). Given a typical pitch of 0.3 for a 25 mm field width, the table motion is modeled by offsetting adjacent beams by 0.147 mm (e.g., $0.3 \times 25 \text{ mm}/51$) increments parallel to the direction of table motion.

A Hi-ART II treatment delivery typically takes <5 min for small target volumes like a prostate and <10 min for larger volumes such as head and neck. Mackie et al. (19) provided an expression for estimating the irradiation time given the target length L , the beam width W , the prescribed dose D , the average dose rate at the target R , and the modulation factor M :

$$T = MD(L + W)/WR \quad (1)$$

The modulation factor is a user selectable planning parameter along with the beam width and pitch. It is defined as the ratio of the maximum leaf open time of any leaf to the average leaf opening time of all the nonzero values, and is typically selected in the range of 1.5–3.0. As an example, a 7 cm prostate volume irradiated at an average dose rate at the target volume of 4 Gy/min by a 2.5 cm wide fan beam with a modulation factor of 2.5 would yield a beam-on time of ~ 4.75 min.

THE HI-ART II CT IMAGING SYSTEM

The detector resolution of the Hi-ART II unit projected to the axis of rotation is ~ 0.6 mm in the transverse direction and equal to the slice width in the longitudinal direction. The rotation period of the Hi-ART II helical tomotherapy unit is 10 s (6 rpm) and the typical slice thickness used is 4 mm, however, a small slice width (e.g., 2 mm) could be used for the fine resolution needed for small target volumes. The unit takes ~ 800 projections (or views) per rotation. For each rotation, two CT slices can be obtained. Pitches of 1, 1.5, and 2 are available, which means that up to 0.8 cm length can be scanned in 10 s. A typical tumor of 8–10 cm length would take as little as 2 min to acquire 25 CT slices. Longer lengths and smaller pitches take more time proportionately. Acquisition occurs on the fly so there is little delay following acquisition for the images to be analyzed. Pixel resolution at the center of the image is dominated by the detector resolution and at the edge of the circle of reconstruction by the number of projections. The reconstructed images shown in Fig. 5 indicate that the Hi-ART II is capable of resolving 1.2–1.6 mm objects near the edge of a 30 cm diameter phantom.

The verification CT uses an ~ 3.5 MV beam, which means that the photons interact almost exclusively by Compton interactions so that the linear attenuation coefficient is linear with the electron density of the medium (20). Metal artifacts arise in conventional CT scanners because the attenuation of the metal is greatly enhanced due to the photoelectric effect. In helical CT, the beam is penetrating enough to eliminate artifacts arising from metal objects like hip prostheses and dental filings. This means that the

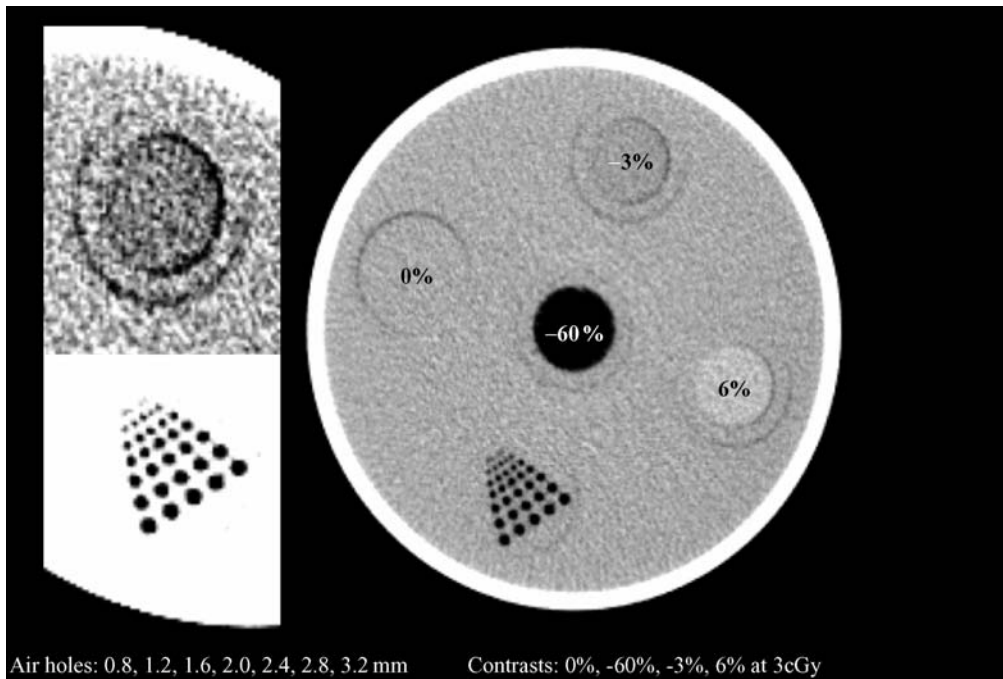


Figure 5. Verification CT at megavoltage (MV) energies of a RMI Solid Water CT phantom. The 3% contrast plug is clearly seen as are the 1.2 mm and 1.6 mm air holes in the solid water phantom. The dose was estimated at 3 cGy at the center of the phantom.

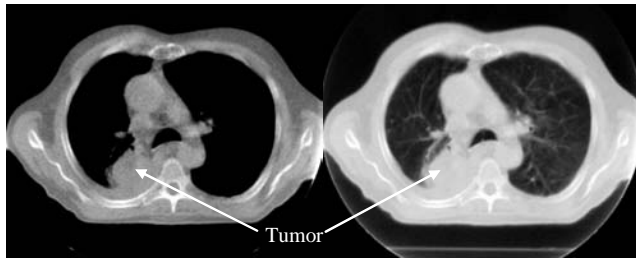


Figure 6. Verification CT of a lung patient. The panel on the left is a soft-tissue window and the panel on the right is a lung window. The difference between muscle and fat and bone and soft tissue is clearly distinguishable. Some of the vascular structures are visible in the lung. The tumor boundary in the lung is discernible, but its extension into the mediastinum is not visible. The dose to the patient was ~3 cGy.

representation supplied by a verification CT is a more reliable CT system for patients with metal implanted appliances. Figure 6 illustrates that bone has less contrast than a conventional CT scan, but it is still clearly discernible on the Hi-ART II unit. The boundary of lung with internal major airways and vascular structures are evident. The boundary between fat and muscle is clearly distinguished, which means that the breast and prostate are discernable. Other organs such as the kidney and

bladder are visible. Unlike the highest quality conventional CT scanners, the contrast between white and gray matter in brain is not visible.

The CT imaging is performed prior to each treatment to reduce the possibility of a geometrical miss of the target and sensitive structures. An automated comparison of verification and planning image sets is carried out immediately following image acquisition to guide the adjustment of the patient setup. The patient is assumed to be a rigid object requiring translations and rotations to bring the target anatomy and important sensitive structures into alignment with the treatment plan. The patient is positioned by aligning the patient's skin marks with lasers located outside of the bore of the unit. A sagittal representation of the patient's planning CT is shown on the operator console to aid in selecting the slices to be scanned. A verification scan is taken and reconstructed during the acquisition. The patient is then transported to the same position outside of the gantry bore while the verification image set is fused onto the planning image set and the translation and rotation offsets are reported. Typically the image fusion is first done automatically using a mutual information algorithm (21). Following automated registration, the patient registration can be fine-tuned manually. This allows the operator to take into account, as best as possible, the nonrigid nature of the transformation. Once

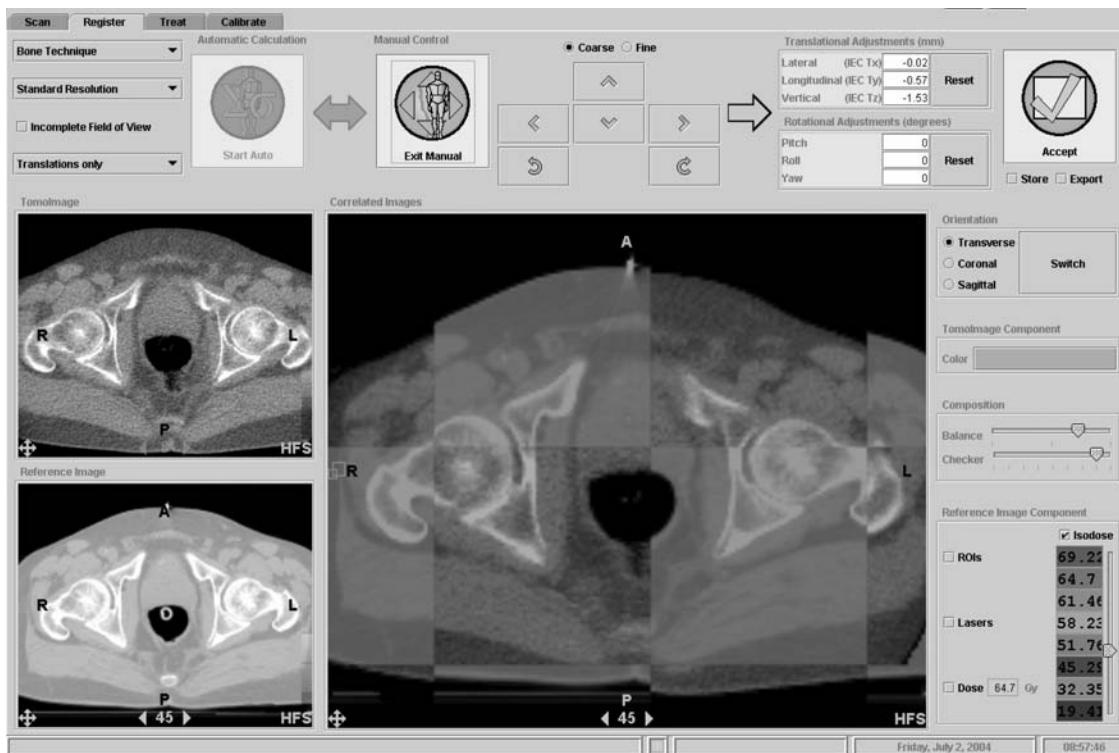


Figure 7. The registration window for a prostate patient. The gray squares are from the planning CT. The verification CT is shown in the upper left and the planning CT is in the lower left. The yellow squares in the large panel are from the tomotherapy verification CT. The rectal boundary and the fat pad surrounding the prostate are clearly aligned on these transverse images. Note that the skin boundary and the leg bones are not as well aligned as the prostate. Regions of interest and the dose distribution obtained from the planning system can be superimposed on the images, but these capabilities have been turned off in this presentation. The translational alignments suggested were 0.0 mm lateral, -0.6 mm longitudinal, and -1.5 mm vertical.

the image registration is completed, the offsets also describe how the patient must be adjusted. Figure 7 shows an example of the graphical user interface for the registration utility being used to register a prostate patient.

If the patient requires adjustment the patient can be translated accordingly. The Hi-ART II CT couch has automated vertical (elevation) and longitudinal translations, and automated gantry start angle adjustments to account for patient roll. The couch top can be manually adjusted in the lateral direction (x direction). Yaw and pitch rotations can be accommodated using angularly calibrated immobilization and positioning aids, which are especially useful for the head and neck. The Hi-ART II includes a set of moveable CT-simulator lasers so that the modified position of the patient can be confirmed.

IMAGE GUIDANCE AND ADAPTIVE RADIOTHERAPY

Adaptation of future treatments from information gleaned from past treatments is a further refinement based on daily CT verification. The daily CT image set forms a model of the patient that can be used to compute an estimate of the dose delivered for that treatment. *Dose reconstruction* is a determination of the dose delivered superimposed on the CT at the time of treatment (6). The CT detector runs at the time of treatment recording the treatment beam exiting through the patient and couch. Using the CT image set acquired just before treatment, the energy fluence incident on the patient can be computed and used to estimate the dose distribution in the patient using the same convolution-superposition method used for helical tomotherapy treatment planning. The total dose delivered up to the current treatment is analyzed for regions of under- and overdosage and these regions reoptimized to bring them inline with the original plan with the corrections applied in one or more future treatments: a process called *adaptive radiotherapy*.

The CT imaging capability of the Hi-ART II has proven extremely useful for reducing daily geometrical misses, which are the most important errors that can compromise treatment efficacy. The increased confidence provided by daily CT setup verification has allowed clinicians to reduce geometrical margins that account for setup uncertainty, and to explore accelerated treatment protocols using a larger dose per fraction than conventional treatments (22). It is expected that CT image guidance technology will facilitate further changes in radiotherapy practice in the coming decade as more facilities adopt this capability.

BIBLIOGRAPHY

- Mackie TR, Holmes T, Swerdloff S, Reckwerdt P, Deasy JO, Yang J, Paliwal B, Kinsella T. Tomotherapy: A new concept for the delivery of dynamic conformal radiotherapy. *Med Phys* 1993;20:1709–1719.
- Carol MP. Peacock: A system for planning and rotational delivery of intensity-modulated fields. *Int J Imag Syst Technol* 1995;6:56–61.
- Curran B. Where goest the Peacock? *Med Dos* 2001;26(1):3–9.
- Mackie TR, Holmes TW, Reckwerdt PJ, Yang J. Tomotherapy: Optimized planning and delivery or radiation therapy. *Int J Imaging Sys Tech* 1995;6:43–55.

- Mackie TR, Kapatoes J, Ruchala K, Lu W, Wu C, Olivera G, Forrest L, Tome W, Welsh J, Jeraj R, Harari P, Reckwerdt P, Paliwal B, Ritter M, Keller H, Fowler J, Mehta M. Image guidance for precise conformal radiotherapy. *Int J Radiat Oncol Biol Phys* 2003;56(1):89–105.
- Kapatoes JM, Olivera GH, Balog JP, Keller H, Reckwerdt PJ, Mackie TR. On the accuracy and effectiveness of dose reconstruction for tomotherapy. *Phys Med Biol* 2001;46:943–966.
- Brahme Optimization of stationary and moving beam radiation therapy techniques. *Rad Oncol* 1988;12:129–140.
- Bortfeld T, Burkelbach J, Boesecke R, Schlegel W. Methods of image reconstruction from projections applied to conformal radiotherapy. *Phys Med Biol* 1990;35(10):1423–1434.
- Webb S. Optimization by simulated annealing of three-dimensional conformal treatment planning for radiation fields defined by a multileaf collimator. *Phys Med Biol* 1991;36(9):1201–1226.
- Holmes T, Mackie TR, Simpkin D, Reckwerdt P. A unified approach to the optimization of brachytherapy and external beam dosimetry. *Int J Radiat Oncol Biol Phys* 1991;20(4):859–873.
- Holmes TW. A Model for the Physical Optimization of External Beam Radiotherapy, Ph.D. dissertation. Madison (WI): Department of Medical Physics, University of Wisconsin; 1993.
- Swerdloff S, Mackie TR, Holmes TW. Method and apparatus for radiation therapy. US patent 5,317,616. 1994.
- Followill D, Geis P, Boyer A. Estimates of whole-body dose equivalent produced by beam intensity modulated conformal therapy. *Int J Radiat Oncol Biol Phys* 1997;1;38(3):667–672. Erratum in: *Int J Radiat Oncol Biol Phys* 1997;1;39(3):783.
- Mutic S, Low DA. Whole-body dose from tomotherapy delivery. *Int J Radiat Oncol Biol Phys* 1998;1;42(1):229–232.
- Low DA, Mutic S, Dempsey JF, Markman J, Goddu SM, Purdy JA. Abutment region dosimetry for serial tomotherapy. *Int J Radiat Oncol Biol Phys* 1999;1;45(1):193–203.
- Salter BJ, Hevezi JM, Sadeghi A, Fuss M, Herman TS. An oblique capable patient positioning system for sequential tomotherapy. *Med Phys* 2001;28(12):2475–2488.
- Kissick MW, Jeraj R, Kapatoes JM, Keller H, Mackie TR, Olivera GH. The thread effect of helical tomotherapy. The XIVth International Conference on the Use of Computers in Radiation Therapy; 2004. p 185–186.
- Mackie TR, Olivera GH, Kapatoes JM, Ruchala KJ, Balog JP, Tome WA, Hui S, Kissick M, Wu C, Jeraj R, Reckwerdt PJ, Harari P, Ritter M, Forrest L, Welsh J, Mehta MP. Helical tomotherapy. In: Palta J, Mackie TR, editors. *Intensity-Modulated Radiation Therapy: The State of the Art*. College Park (MD): American Association of Physicists in Medicine; 2003. p 247–284.
- Mackie TR, Hughes J, Olivera GH, Kapatoes J, Ruchala K, Ramsey C, Kissick M, Jeraj R. The delivery time for helical tomotherapy. The XIVth International Conference on the Use of Computers in Radiation Therapy; 2004. p 750–752.
- Jeraj R, Mackie TR, Balog J, Olivera G, Pearson D, Kapatoes J, Ruchala K, Reckwerdt P. Radiation characteristics of helical tomotherapy. *Med Phys* 2004;31(2):396–404.
- Ruchala KJ, Olivera GH, Kapatoes JM. Limited-data image registration for radiotherapy positioning and verification. *Int J Rad Onc Biol Phys* 2002;54:592–605.
- Fowler JF, Tome WA, Fenwick JD, Mehta MP. A challenge to traditional radiation oncology. *Int J Radiat Oncol Biol Phys* 2004;60(4):1241–1256.

See also COMPUTED TOMOGRAPHY; RADIATION THERAPY, INTENSITY MODULATED; RADIOTHERAPY, THREE-DIMENSIONAL CONFORMAL.

TONOMETRY, ARTERIAL

JOSEPH S. ECKERLE
 SRI International
 Menlo Park, California

INTRODUCTION

The arterial tonometer is an instrument for measuring arterial blood pressure. It differs from the familiar sphygmomanometer in that, rather than measuring the pressure only at greatest contraction and greatest heart dilation (systolic and diastolic), it provides continuous measurement throughout the heart's pumping cycle. Typically, the instrument sensor is placed over a superficial artery; the radial artery pulse point at the wrist is one convenient site for tonometer measurements. Figure 1 shows how a tonometer sensor would be placed for measurements at this site.

A catheter can be used for accurate, continuous measurement of blood pressure, but the instrument is invasive and numerous risks are associated with its use. In contrast, the noninvasive tonometer can provide an accurate, continuous blood pressure measurement with negligible risk.

Sphygmomanometric instruments of several types are available for noninvasive blood pressure measurements. However, these instruments are generally not capable of continuous blood pressure measurement, nor is their long-term use feasible. The familiar blood pressure cuff hinders venous return and results in peripheral edema. In contrast to sphygmomanometric instruments, the tonometer can be used for beat-by-beat blood pressure measurement over long periods of time with minimal edema. Important disadvantages of the tonometer include its sensitivity to sensor placement and movement artifacts, effects of anatomical variations, and the greater complexity and cost relative to a conventional sphygmomanometer.

In what follows, the physical principles that form the theoretical basis for tonometric blood pressure measurement are first presented; these include techniques for identifying the location of an artery beneath a tonometer sensor and for adjusting the force with which the sensor is

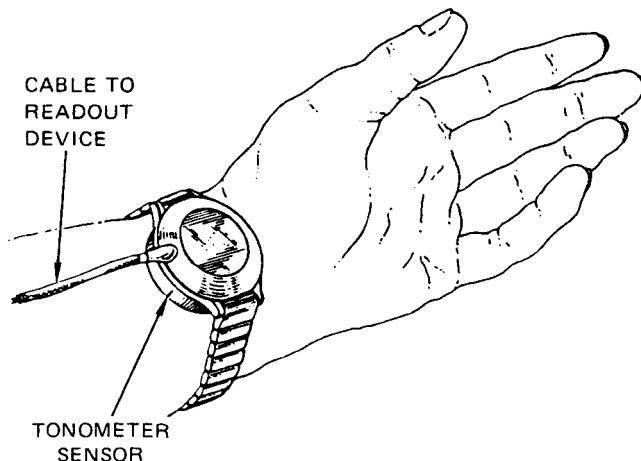


Figure 1. Arterial tonometer sensor on radial artery.

pressed against the skin. Next, the technical evolution of tonometer transducers is discussed, leading to a description of modern, multiple-element transducers. This is followed by discussion of several considerations that are unique to tonometric measurements, and may influence the usefulness of the technique for certain applications. Various applications of tonometry, particularly surgical monitoring and cardiovascular evaluation, are discussed next. Finally, the measurement accuracy of tonometry is addressed.

PRINCIPLES OF OPERATION

General Principles

The fundamental principles underlying arterial tonometry are similar to those for ocular tonometry (1,2). Figure 2 shows an idealized model that helps illustrate these principles. In Figure 2, P represents the blood pressure in a superficial artery and F is the force measured by a tonometer transducer. The membrane is the artery wall. Figure 2b is a "free body diagram" showing all the forces and moments acting on the frictionless piston of Fig. 2a. As shown in Fig. 2b, an ideal membrane transmits only a tensile force, T , and does not transmit any bending moment. The tension vector shown, T , is perpendicular to the pressure vector, so the force, F , is independent of T and depends only on the blood pressure and the area of the frictionless piston, A . Following common practice, the integrated effect of arterial pressure acting on the segment of arterial wall is represented by a vector of magnitude PA , oriented perpendicular to the wall. Thus, measurement of the force, F , permits one to directly infer the intraarterial pressure.

Figure 3 shows a superficial artery and a tonometer sensor in cross-section. The tonometer sensor is represented schematically, and is modeled as an assemblage of springs with spring constants, K , as shown. By careful design of the tonometer sensor and selection of an appropriate superficial artery, it is possible to satisfy several conditions:

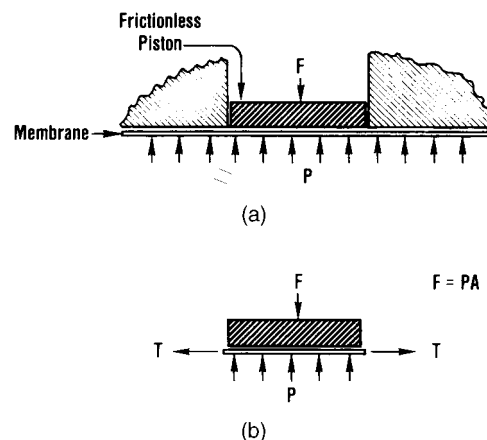


Figure 2. Idealized model for a tonometer.

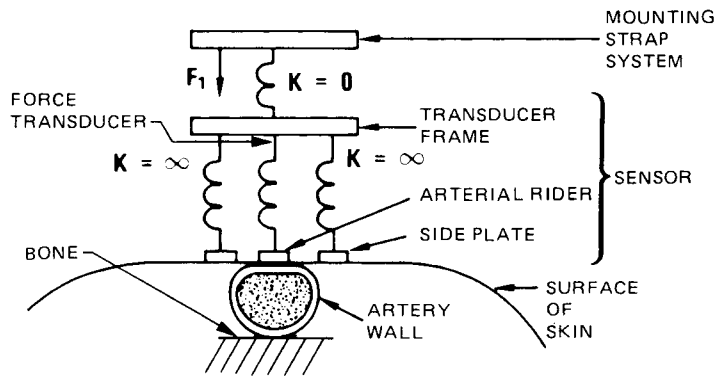


Figure 3. Schematic diagram of a tonometer sensor and a superficial artery.

1. The artery is supported from below by bone (e.g., the radius).
2. The hold-down force, F_1 , flattens a portion of the artery wall, but does not occlude the artery.
3. The thickness of the skin over the artery is insignificant, compared to the artery diameter.
4. The artery wall behaves essentially like an ideal membrane.
5. The arterial rider is smaller than the flattened area of the artery, and is centered over the flattened area.
6. The spring constant of the force transducer, K_T , is large compared to the effective spring constant of the artery.

When all these conditions are satisfied, Pressman and Newgard (3) showed theoretically that the conditions of Fig. 2 apply where the arterial rider of Fig. 3 corresponds to the frictionless piston of Fig. 2. Thus, the electrical output signal of the force transducer is directly proportional to the intraarterial blood pressure.

The arterial tonometric measurement depends on the membrandlike behavior of the artery wall. Drzewiecki et al. (4) have shown analytically and by experiments with an excised canine femoral artery (5) that the desired behavior can be obtained, provided that the artery is flattened sufficiently. This work provides an important theoretical foundation for the arterial tonometer and helps to explain the observations of several prior *in vivo* studies.

Multiple-Element Sensors

A major practical problem with the simple arterial tonometer of Fig. 3 is the requirement that the arterial rider be precisely placed over the superficial artery. Reliable measurements can be obtained only after painstaking adjustment of the sensor location by a trained operator (6). Apparently, there are differences of opinion concerning the severity of this positioning problem and these are discussed in greater detail below.

To ameliorate this problem, multiple-element tonometer sensors, shown schematically in Fig. 4, have been developed (7-9). The sensor worn by the patient actually consists of a multiplicity of individual sensors. Typically, the sensors are arranged to form a linear array of force transducers and arterial riders. The array need only be

positioned with enough precision so that some element of the array is centered over the artery. A computer then automatically selects the sensor element that is correctly positioned over the artery.

One algorithm for selection of the correct element from the multiple-element sensor array exploits two characteristics of the pressure distribution in the vicinity of the artery (8). The first is that the pulse amplitude (i.e., the pressure difference between the systolic and diastolic points on the transducer output waveforms) exhibits a broad maximum over the artery. The algorithm searches for the largest pulse amplitude; the corresponding sensor element will then be within about one artery diameter of the center of the artery. However, this element will, in general, not be precisely centered over the artery.

To identify the centered sensor element more precisely, the algorithm then exploits a second phenomenon, illustrated in Fig. 5. This figure shows a multiple-element tonometer sensor and the underlying, partly compressed artery. For purposes of illustration, assume that the diastolic pressure in the artery is 80 mmHg (10.7 kPa). At the instant of diastole, the pressure measured by each element of the sensor is shown plotted at the top of the figure. Elements 4-6, which all lie over the flattened part of the artery wall, measure the intraarterial pressure [80 mmHg (10.7 kPa)] with good accuracy. However, the pressures measured by elements 2, 3, 7, and 8 are all significantly greater than the intraarterial pressure. This higher pressure can be explained by noting that the artery wall is bent to a very small radius in the regions below the latter elements. As a result, large bending moments are transmitted by the artery wall and are manifested as increased

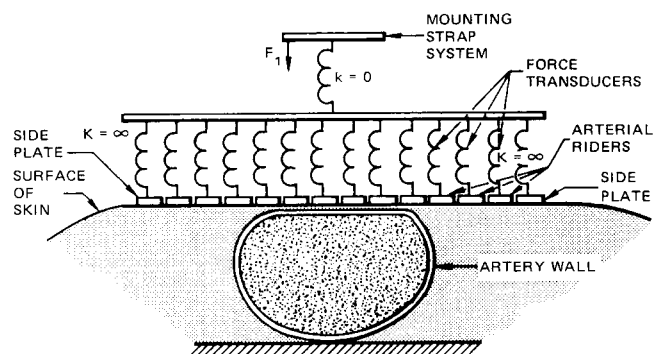


Figure 4. Multiple-element arterial tonometer.

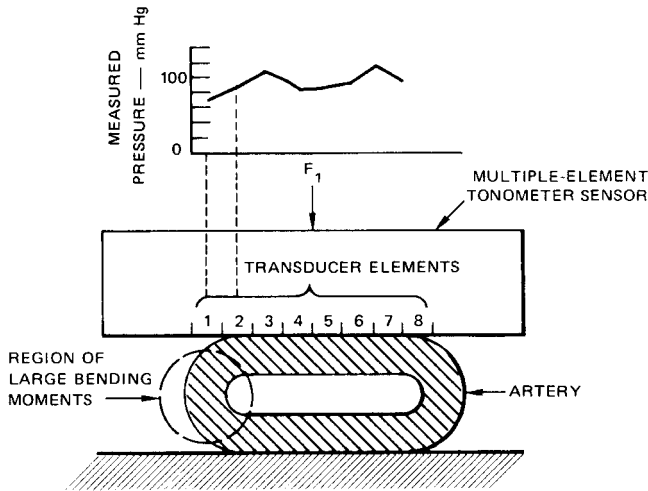


Figure 5. Illustration of pressure distribution near a superficial artery.

pressure on the adjacent sensor elements. A more detailed treatment of this phenomenon can be found in Drzewiecki et al. (4). The element-selection algorithm exploits this phenomenon by searching for a (spatial) local minimum in diastolic pressure (e.g., element 5 of Fig. 5) in a region near the maximum pulse amplitude. The term “local minimum” has a precise meaning in mathematics: When moving in either direction from a local minimum, the value of the function increases. The sensor element corresponding to the local minimum is then assumed to be centered over the artery and blood pressure is measured with this element (10).

Recently, there have been several efforts (11–13) to improve the basic algorithm described above. One algorithm or another might be most effective depending on numerous factors, such as the application, the patient population, and the precision of the sensor and associated amplifiers.

To further ameliorate the positioning problem, Shinoda and others (14,15) have developed motor-driven mechanisms to move a multiple-element sensor laterally within a larger housing strapped to the wrist.

Hold-Down Force

Adjusting the tonometer sensor’s transverse location with respect to the artery is not enough; the degree of arterial flattening is also important for accurate tonometric pressure measurement. Arterial flattening depends on the interaction of anatomical factors with the value of the hold-down force, F_1 , in Fig. 3. The appropriate value of the hold-down force must be determined for each subject before accurate tonometric measurements can be made. The procedure commonly employed involves increasing (or decreasing) the hold-down force gradually while recording the signal from the tonometer sensor. Figure 6 is an example of such a recording (16). In Fig. 6, hold-down force decreases with time through regions A, B, and C. Region B, where the pulse amplitude is greatest, is considered (3,16,17) to be the region where the most accurate blood pressure measurements are made. This region corresponds to flattening of the artery (as shown in Fig. 5) that is insufficient to cause its occlusion.

Drzewiecki et al. (5) discuss the effects of hold-down force from a theoretical perspective. Eckerle (18) developed algorithms for automatic identification of the center of region B (Fig. 6) and for recognition of difficult subjects (see below) based on various parameters that define this region. Briefly, the algorithm fits a third-order polynomial to the data of Fig. 6. The locations of the regions of Fig. 6 can then be directly computed from the polynomial coefficients. Recently, alternative algorithms (19,20) to control hold-down force have been developed. Again, the optimum algorithm may depend on factors such as application, population, and equipment precision.

TONOMETER SENSOR DESIGN

The size and precision requirements for tonometric blood pressure sensors are severe. Eckerle et al. (21) conclude that, ideally, the arterial rider should be less than ~0.2 mm wide and the associated transducer should be accurate to better than ±2 mmHg (270 Pa). For multiple-element sensors, at least 25 elements with interelement spacings of ~0.2 mm are desirable. As of 1984, these design goals had been approached but not met (21). Then, in 1990, an integrated circuit (IC) based tonometer sensor array was

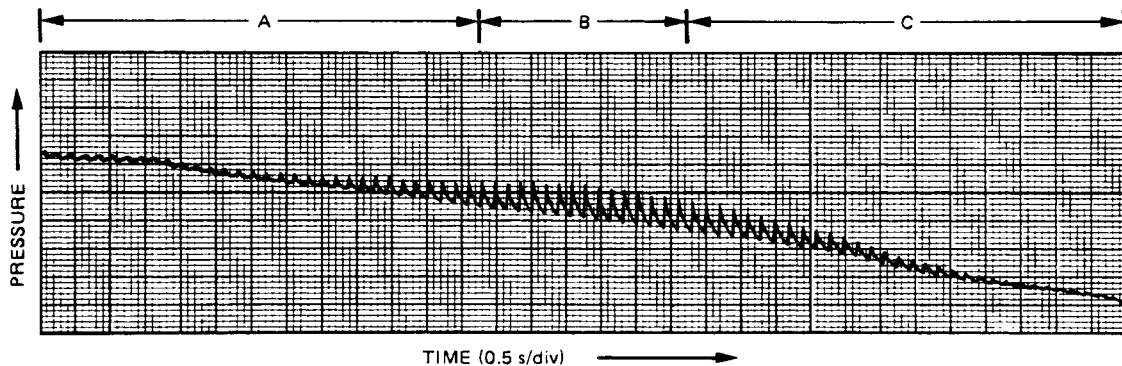


Figure 6. Effect of hold-down force on a tonometer output signal.

reported (9) that substantially achieved these goals. The silicon-and-glass sensor die included 31 tonometer sensor elements in a linear array ~ 6 mm long.

The first single-element tonometer sensors were constructed of aluminum. Strain gages were attached to miniature beams supporting the arterial rider (3). This approach made effective sensors, but they suffered from the positioning problems previously discussed. Subsequently, various workers devised alternative single-element tonometric sensors. Bigliano et al. (22) reported a pressure sensor with a thin membrane that controlled air flow through a narrow passage. Stein and Blick (23) used a modified myographic force transducer to record blood pressure waveforms from the radial artery. Bahr and Petzke (17,24) used a semiconductor pressure transducer placed over the radial artery. Kelly et al. (6) used a pressure sensor intended for insertion via a catheter. This sensor was mounted in the tip of a pencil-shaped probe, which was then applied to the skin. Borkat et al. (25) placed a pressure capsule, consisting of a rubber bladder attached to a pressure transducer, over the radial artery. Borkat's device is probably the most inexpensive and simple of these alternatives, but it fails to meet the size and stiffness requirements that apply for accurate tonometric measurements.

Multiple-element tonometer sensors have been fabricated from a monolithic silicon substrate using anisotropic etching to define pressure-sensing diaphragms about $10 \mu\text{m}$ thick in the silicon (26). The IC processing techniques are then used to create piezoresistive strain gages in the diaphragms. External circuitry measures the resistance of the strain gages to determine the pressure exerted on each sensor element. Figure 7 is a photomicrograph of an eight-element sensor fabricated in this way. The diaphragms are square and are arranged in two staggered rows of four each. Note the scale in the figure. The arterial riders in this device are 0.75×0.75 mm. Figure 8 is a further magnified view of one element of the Figure 7 sensor. Two radial and two tangential piezoresistive strain gages can be seen together with aluminum metallization used for connection to external circuitry. The performance of these sensors is representative of the best that had been achieved (for tonometers) as of 1984 (21) and is summarized in Table 1.

Achieving the size and accuracy requirements for multiple-element tonometer sensors noted above (21) is difficult, even using the latest advances in IC sensor fabrication. One fundamental problem involves the difficulty of placing independent pressure sensors side by side while minimizing interaction between them. In 1990, Terry et al. (9) reported a clever configuration to address this problem. Briefly, a multiplicity of independent pressure transducers shared a single, long, narrow silicon diaphragm. Some performance parameters of this sensor are also shown in Table 1. This sensor was the first to substantially achieve the size and precision requirements proposed by Eckerle et al. (21).

While the single-diaphragm sensor of Terry et al. (9) has superior performance, relative to the Fig. 7 device, it requires a more complex manufacturing process, and is therefore more expensive. For cost-sensitive applications, a sensor such as Fig. 7 may be preferred. Other groups are

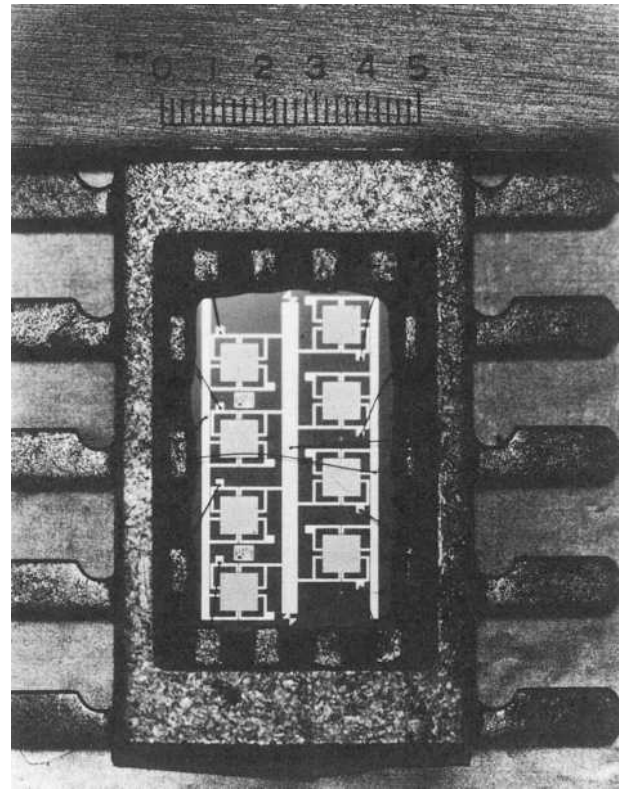


Figure 7. Eight-element tonometer sensor.

investigating fiber-optic transducers (27) and capacitive transducers (28,29) for use in multiple-element tonometer sensors.

Proper mounting of an IC tonometer sensor is very important. The mounting arrangement must protect the fragile sensor while faithfully transmitting the pressure of the patient's skin to the sensor. Consideration should be given to measurement drift caused by thermal effects or material creep. Finally, the shape of the mounted sensor can be chosen to conform to the local anatomy (e.g., the nearby radius in the case of a radial artery tonometer). Fujikawa and Harada (30) and others (31,32) developed

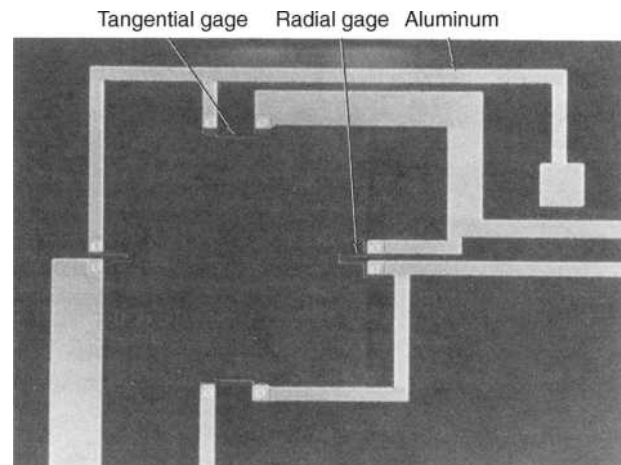


Figure 8. Pressure-sensitive diaphragm of tonometer sensor.

Table 1. Typical Performance of Silicon Tonometer Sensors

Parameter	Typical Value ^a		Units
	Ref. 26	Ref. 9	
Number of elements	8	31	Each
Element spacing	0.75	0.2	mm
Sensitivity	25	25–50	$\mu\text{V}/\text{V mmHg}$
Nonlinearity	3	n.r.	mmHg
Temperature coefficient of sensitivity	-0.25	n.r.	$\% \cdot ^\circ\text{C}^{-1}$
Temperature coefficient of offset (uncompensated)	1	n.r.	$\text{mmHg} \cdot ^\circ\text{C}^{-1}$
Offset error due to temperature (with temperature compensation)	± 1 (estimate)	n.r.	mmHg
Frequency response (flat to <1.0 dB)	>50	n.r.	Hz
Noise	<0.5	n.r.	mmHg

^an.r. = not reported.

methods for mounting IC sensors for tonometry in clinical environments.

A clinical tonometer instrument, incorporating many of the features described above, is shown in Fig. 9. Comparable instruments may be obtained from suppliers such as Colin Medical Technology Corp., Komaki, Japan; Hypertension Diagnostics, Inc., Eagan, Minnesota; and AtCor Medical, Sydney, Australia.

MISCELLANEOUS CONSIDERATIONS

One significant advantage of the arterial tonometer is its ability to make noninvasive, nonpainful, continuous measurements for long periods of time. There are several reasons for this superiority over the sphygmomanometer:

1. The sensor is adjusted to partly flatten, but not occlude the radial artery. In contrast, sphygmomanometric systems typically occlude the artery in order to determine systolic pressure.

2. The part of the tonometer sensor representing the arterial riders and side plates of Fig. 3 is ~ 20 mm in diameter. There may be occlusion of some veins beneath this part of the sensor, but numerous parallel venous paths in the remainder of the wrist allow return venous flow. [The reader may wish to try the following experiment: Place a dime (a U.S. coin ~ 18 mm in diameter) over your radial artery pulse. Holding your wrist with your free hand, press down on the dime with the thumb to occlude the radial artery. You are now experiencing greater venous occlusion than a tonometer sensor would cause.]
3. There are several venous paths through the wrist. Some of these pass between bones that apparently help protect them from pressure that may be applied externally to the wrist by the mounting strap and other parts of the tonometer sensor.

These factors all help to minimize development of edema distal to a tonometer sensor. Discomfort and development of edema will depend on factors, such as subject-to-subject

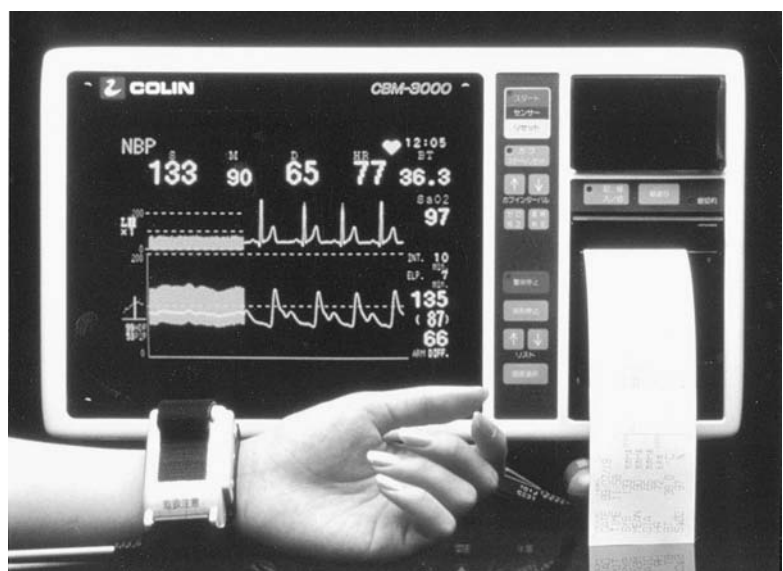


Figure 9. Clinical tonometer instrument.

variations in wrist anatomy, measurement duration, and the hold-down force used. In a study involving 20 conscious, healthy subjects (33) the blood pressure waveform was monitored continuously by tonometry for more than 30 min. The subjects reported no significant discomfort due to the tonometer sensor.

The tonometer's ability to measure blood pressure continuously with minimal distraction to the subject makes it attractive for ambulatory blood pressure monitoring. This application should be feasible in the future, but present instruments do not have sufficient artifact-rejection capability for reliable measurements on ambulatory subjects (21). This is not to say that the arterial tonometer is unusually sensitive to movement artifacts. In a study involving 20 subjects (34), the tonometer was found to be less sensitive to subject movement than photoplethysmographic, quadrupolar impedance plethysmographic, and sphygmomanometric sensors.

Eckerle et al. (35) have taken the first step toward an ambulatory tonometer for blood pressure by developing an ambulatory pulse rate sensor using tonometry. By use of a curved spring, the sensor can be mounted unobtrusively in a watchband <8 mm thick.

Taking an analog signal-processing viewpoint, the blood pressure can be described as the sum of an alternating current (ac) component and a direct current (dc) component. Briefly, the ac component is more easily measured by tonometry than the dc component. More specifically, under clinical conditions, the six conditions listed above may not always be satisfied. When this happens, the tonometer sensor often continues to measure the ac component with good fidelity, while the measurement accuracy of the dc component (on which the familiar systolic and diastolic depend) becomes degraded. Shinoda and others (36–38) have used sphygmomanometric measurements of blood pressure to correct errors in the dc component and thereby produce an output with good waveform fidelity (ac component) and accurate measurement of systolic and diastolic (dc component).

The radial artery is not the only site at which a tonometer may be applied, but most research to date has used this site. Other sites suitable for tonometric measurements include the brachial artery at the inner elbow (the antecubital fossa), the temporal artery in front of the ear, the carotid, and the dorsalis pedis artery on the upper foot.

Anatomical variations of the wrist can make the location and support of the radial artery unsuitable for tonometric

pressure measurement. Unsuitability occurs in only a small fraction of the population and such difficult subjects can be recognized by the computer used with multiple-element sensors. In one study involving 6 subjects (7,8) there was one difficult subject, while in another involving 20 subjects (33) there were none.

As described above, a single-element tonometer sensor, such as Fig. 3, can be used for tonometric blood pressure measurement if six conditions are met. The multiple-element tonometer was developed to help simplify tonometric measurements for the clinician (relative to a single-element instrument). Many clinicians (39–45) have reported using multiple-element instruments. Many other clinicians (6,46–54) have used single-element instruments. Chen et al. (44) have used both types. They observe that "Although [the hand-held, single-element tonometer]...was probably adequate for brief steady-state data, manual recording was too unstable for accurate pressure tracking during hemodynamic transients, and it introduced an element of user dependence and thus potential bias to the data. The automated [multiple-element tonometer] system circumvented these limitations." We may also observe that the two groups are generally pursuing different applications, suggesting that the choice of optimum sensor type is application dependent.

APPLICATIONS

Perhaps the first clinical application of tonometry was for blood pressure monitoring in surgery and other procedures. Stein and Blick (23) used radial artery tonometry during a catheterization procedure. Kemmotsu et al. (39) used a multiple-element tonometer for surgical monitoring. Several others (40–43,55) have also evaluated the instrument under various surgical and postsurgical situations.

Figure 10 shows a typical blood pressure waveform, obtained with an arterial tonometer from the radial artery of an adult male. Note that systolic and diastolic pressures, as well as pulse rate and dicrotic notch information, can be obtained from the waveform. The subject performed a Valsalva maneuver at the time indicated. The attendant changes in blood pressure and pulse rate are quite clearly indicated. Of course the tonometer measures only the pressure in the underlying artery, not central arterial

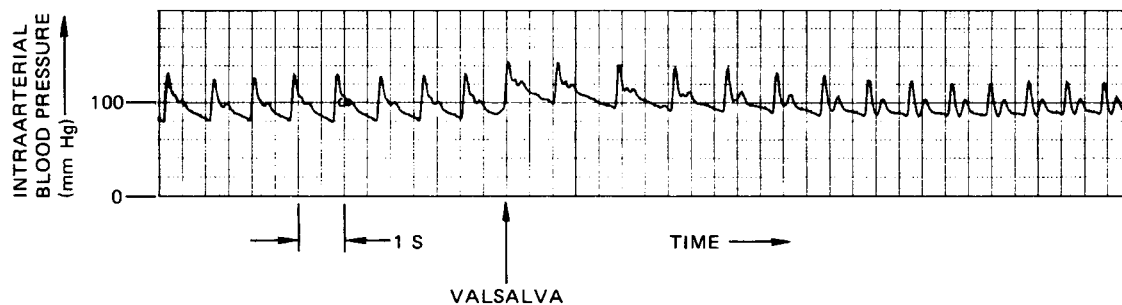


Figure 10. Intraarterial blood pressure waveform.

(aortic) pressure. Central aortic pressure is useful for clinical applications, so there has arisen a desire to calculate central aortic pressure based on a radial pressure obtained with a tonometer.

A brief discussion of circulation dynamics may now help reveal the motivation for some other tonometer applications. Physiologists have long struggled to devise a useful physical model for the circulatory system. Briefly, two complementary models have arisen, a windkessel model and a pulse wave velocity (PWV) model (56,57). In the windkessel model, the circulatory system is represented (often using an electronic analogy) as a small number of lumped elements such as capacitors, inductors, and resistors. In the PWV model, the circulatory system is represented as one or more pipes in which a pulse wave, generated by the heart, travels at a certain velocity. Physiologists and clinicians seek to determine the values of the lumped elements and the magnitude of the PWV by means of noninvasive measurements. Various aspects of the ac component of the pressure waveform of Fig. 10 can be used to estimate the windkessel parameters and PWV.

Focusing on the windkessel model, one can devise a transfer function that relates the pressure at a peripheral site (e.g., the radial) to an input pressure waveform in the aorta (58,44). It is then straightforward (in theory) to compute the inverse of this transfer function. Using this inverse transfer function, one may then calculate the pressure waveform in the aorta based on the peripheral waveform. This technique is potentially very powerful because it allows making a previously invasive measurement (the aortic pressure waveform) by using a noninvasive instrument: a radial artery tonometer.

Several investigators (44,46,58) have evaluated the above technique to determine aortic pressure by mathematical manipulation of a radial artery pressure determined by tonometry. Others (47,59) have evaluated a similar technique to determine aortic pressure from a tonometric carotid pressure measurement.

Numerous investigators (48–51) have used tonometry to estimate windkessel parameters, particularly arterial elasticity. Other investigators (45,52–54) have used it to estimate PWV.

For certain applications, it is sufficient to measure relative, rather than absolute, blood pressure. These include polygraph (lie detector) tests and studies of the physiological effects of various cognitive and physical stressors (e.g., weightlessness and high $-g$ aircraft maneuvers). A multiple-element tonometer designed specifically for relative blood pressure measurement has been developed by Eckerle et al. (33) and tested extensively on 20 subjects of both sexes. Subjects were subjected to stressors such as mental arithmetic and cold pressor, and their blood pressures were recorded continuously for periods in excess of 30 min.

MEASUREMENT ACCURACY

Because the accuracy of tonometric blood pressure measurements depends on the size and positioning of

the sensor, the value of hold-down force, and the accuracy of the sensor itself, some variability in reported tonometer accuracy can be expected when these factors are not well controlled. In nearly all of the measurements reported prior to 1987 hold-down force was adjusted manually, and in many the tonometer was manually positioned as well. Some improvement in tonometer accuracy and repeatability can be expected from automatic sensor positioning and hold-down force adjustment.

An early validation of the principles of tonometry (60) involved comparison of pressures measured in an exposed canine femoral artery by both a tonometer and an intraarterial catheter. One series of tests involved 15 animals with blood pressure changes being induced by drug injection and vagal stimulation. For this entire series of tests, the difference between pressures measured by the two instruments was never $> 5\%$.

Stein and Blick (23) compared tonometric measurements with direct arterial blood pressures on 20 patients undergoing cardiac catheterization. They found that the tonometer reproduced both the intraarterial waveform and any abrupt changes in pressure caused by various interventions with remarkable fidelity. Specifically, during interventions that increased or decreased systolic pressures by as much as 30%, the tonometric pressure measurement followed the direct arterial pressure within 10% in 92% of the observations.

Weaver et al. (8) compared sphygmomanometric blood pressures with those obtained with a multiple-element tonometer sensor on five subjects of both sexes ranging in weight from 135 to 225 lb (61–102 kg). The standard deviation between the two measurements of systolic and diastolic pressures was 6.5 mmHg (870 Pa), indicating that the tonometer was at least as accurate as the sphygmomanometer.

In 1989, Kemmotsu et al. (39) reported the accuracy of an automated, multiple-element tonometer instrument used for monitoring during surgery. The tonometric measurements showed good correlation ($r = 0.94$, $P < 0.001$ for systolic) with invasive measurements on the contralateral radial artery. Kemmotsu subsequently performed additional studies (40,41) involving surgical monitoring. In the later study (41), involving 60 patients, the mean absolute value of error ranged from 3.6 to 6.6 mmHg (480 to 880 Pa). Standard deviations ranged from 4.5 to 6.2 mmHg (600 to 830 Pa).

Other investigators (42,43,55) have evaluated multiple-element tonometers for surgical and postsurgical monitoring. They found standard deviations ranging from 1.7 (for a single patient) to 14.2 mmHg (230 to 1960 Pa).

Sato et al. (61) evaluated the accuracy of a multiple-element tonometer instrument on conscious subjects subjected to a Valsalva maneuver and a tilting test. This study is of particular interest because they directly addressed the question of frequency response of the tonometer/tissue/artery system. Good frequency response is important for computation of windkessel parameters or PWV as described above. Sato et al. found the frequency response of the tonometer/patient system to be essentially flat from 0 to 5 Hz.

BIBLIOGRAPHY

1. Mackay RS, Marg E. Fast, automatic electronic tonometers based on an exact theory. *Acta Ophthalmol* 1959;37:495–507.
2. Mackay RS. Fast automatic ocular pressure measurement based on an exact theory. *IRE Trans Med Electron* 1960;ME-7:61–67.
3. Pressman G, Newgard P. A transducer for the continuous external measurement of arterial blood pressure. *IEEE Trans Bio-Med Electron* 1963;BME-10:73–81.
4. Drzewiecki GM, Melbin J, Noordergraaf A. Arterial tonometry: Review and analysis. *J Biomech* 1983;16:141–152.
5. Drzewiecki GM, Melbin J, Noordergraaf A. Deformational forces in arterial tonometry. *Proc Annu Conf IEEE/Eng Med Biol Soc*, 6th 1984;26:642–645.
6. Kelly R, Hayward C, Avolio A, O'Rourke M. Noninvasive determination of age-related changes in the human arterial pulse. *Circulation* 1989;80:1652–1659.
7. Eckerle JS, Newgard PM. A non-invasive transducer for the continuous measurement of arterial blood pressure. *Proc Annu Conf Eng Med Biol* 1976;29:98.
8. Weaver CS, et al. A study of non-invasive blood pressure measurement techniques. *Noninvasive Cardiovasc Meas* 1978;167:89–105. Prepared for a conference sponsored by NASA Office of Technology Utilization and Cardiology Division, Stanford University School of Medicine. Society of Photo-Optical Instrumentation Engineers, Bellingham (WA).
9. Terry S, Eckerle JS, Kornbluh RD, Ablow CM. Silicon pressure transducer arrays for blood-pressure measurement. *Sensors Actuators* 1990;A 21–A 23:1070–1079.
10. Eckerle JS. Noninvasive blood pressure monitoring transducer. US Patent 4,269,193. 1981.
11. Niwa M. Pulse wave detecting apparatus. US patent 5,103,831. 1992.
12. Kobayashi I. Pulse wave detecting apparatus. US patent 5,170,796. 1992.
13. Takaya M. Continuous blood-pressure monitor apparatus. US patent 6,394,959 B1. 2002.
14. Shinoda M, Ogletree WA. Pulse wave detecting apparatus. US patent 4,784,152. 1988.
15. Kaida N, et al. Pulse wave detecting apparatus. US patent 4,901,733. 1990.
16. Eckerle JS. SRI Tonometer Blood Pressure Measurement System, Operator's Instruction Manual. Final Rep., SRI Proj. No. 4908, Washington (DC): Lab. Div., Federal Bureau of Investigation; 1984.
17. Petzke JC, Bahr DE. Blood pressure measuring apparatus. US patent 3,926,179. 1975 Dec 16.
18. Eckerle JS. Blood pressure monitoring method and apparatus. US patent 4,799,491. 1989.
19. Niwa M. Pulse wave detecting apparatus. US patent 5,119,822. 1992.
20. Ohmori K, Narimatsu K, Kobayashi I. Blood pressure monitoring apparatus. US patent 5,590,661. 1997.
21. Eckerle JS, Fredrick J, Jeuck P. Toward a practical tonometric blood pressure instrument. In: Semmlow JL, Welkowitz W, editors. *Proceedings of the Annual Conference IEEE/English Medical Biology Society* 6th; 1984. p. 635–641.
22. Bigliano RP, Molner SF, Sweeney LM. A new physiological pressure sensor. *Proc Ann Conf Eng Med Biol* 1964;6:82.
23. Stein PD, Blick EF. Arterial tonometry for the atraumatic measurement of arterial blood pressure. *J Appl-Physiol* 1971;30:593–596.
24. Bahr DE, Petzke JC. The automatic arterial tonometer. *Proc Annu Conf Eng Med Biol* 1973;15:259.
25. Borkat FR, Kataoka RW, Silva J. An approach to the continuous non-invasive measurement of blood pressure. *Proceeding of the San Diego Biomedical Symposium*; 1976.
26. Weaver CS, et al. 1976: Wearable Blood Pressure and ECG Recording System. (Grant HL 17604-01A1), Interim Report to the National Heart and Lung Institute, Bethesda (MD); 1976.
27. Drzewiecki GM, Butterfield RD, and Ciaccio EJ, Pressure waveform monitor. US patent 5,363,855. 1994.
28. Guckel H, Burns DW. Planar processed polysilicon sealed cavities for pressure transducer arrays. *Proc IEEE Int Electron Devices Meet* 1984.
29. Corcuera M, Aravamudhan S, Bhansali S. (No Date). A non-invasive microsystem for blood pressure measurement, Research Experience for Undergraduates, Spring 2004 Symposium, Poster No. EE.3. University of South Florida. [Online]. Available at www.eng.usf.edu/~schlaf/REU/Symposium/Spring2004/Corcuera2004.pdf Accessed 2005April 8.
30. Fujikawa K, Harada C. Semiconductor pressure pulse wave system. US patent 5,101,829. 1992.
31. Harada C, et al. Contact pressure sensor. US patent 5,179,956. 1993.
32. Narimatsu K, Kawamura N. Pressure wave sensor. US patent 5,467,771. 1995.
33. Eckerle JS, et al. Cardiovascular Activity Monitoring for Polygraph Examination, SRI Final Rep., Contract J-FBI-82-108, Washington (DC): Federal Bureau of Investigation; 1984.
34. Eckerle JS, The arterial tonometer: A non-invasive blood pressure instrument for trauma victims. *Non-Invasive Neurologic Evaluation of the Combat Casualty Victim*, Bethesda (MD) Naval Medical Research and Development Command; 1985.
35. Eckerle JS, et al. Pulse rate sensor system. US patent 5,243,992. 1993.
36. Shinoda M, Lippincott HW. Continuous blood pressure monitoring system having a digital cuff calibration system and method. US patent 5,165,416. 1992.
37. Aung Y, Takaya M, Nishibayashi H. Blood pressure monitor system. US patent 5,261,414. 1993.
38. Kawamura N, Nakagawa T, Aung Y. Blood pressure monitor system. US patent 5,279,303. 1994.
39. Kemmotsu O, et al. A non-invasive blood pressure monitor based on arterial tonometry. *Anesthes Analges (Suppl.)* 1989;68:S145.
40. Kemmotsu O, et al. Blood pressure measurement by arterial tonometry in controlled hypotension. *Anesthes Analges* 1991; 73:54–58.
41. Kemmotsu O, et al. Arterial tonometry for non-invasive, continuous blood pressure monitoring during anesthesia. *Anesthesiology* 1991;75(2):333–340.
42. Searle NR, Perrault J, Ste-Marie H, Dupont C. Assessment of the arterial tonometer (N-CAT) for the continuous blood pressure monitoring in atrial fibrillation. *Can J Anaesthesiol* 1993;40(4):388–393.
43. Siegel LC, Brock-Utne JG, Brodsky JB. Comparison of arterial tonometry with radial artery catheter measurements of blood pressure in anesthetized patients. *Anesthesiology* 1994;81(3): 578–584.
44. Chen C, et al. Estimation of central aortic pressure waveform by mathematical transformation of radial tonometry pressure validation of generalized transfer function. *Circulation* 1997; 95:1827–1836.
45. Narimatsu K, Takatani S, Ohmori K. A multi-element carotid tonometry sensor for non-invasive measurement of pulse wave velocity. *Frontiers Med Biol Eng* 2001;11(1):45–58.

46. Adji A, O'Rourke MF. Determination of central aortic systolic and pulse pressure from the radial artery pressure waveform. *Blood Pressure Monitoring* 2004;9:115–121.
47. Chen C, et al. Validation of carotid artery tonometry as a means of estimating augmentation index of ascending aortic pressure. *Hypertension* 1996;27:168–175.
48. Collins VR, Finkelstein SM, Cohn JN. Evaluation of pulse contour technique for measuring arterial elasticity. *Circulation* 1980;62(Suppl. II):1111–1120.
49. Tanaka H, et al. Aging, habitual exercise, and dynamic arterial compliance. *Circulation* 2000;102:1270–1275.
50. Zimmerman A, et al. Loss of oscillatory arterial compliance is detectable in young patients by radial artery pulse contour analysis. *Am J Hypertension* April 2000;13: (4, Part 2) Abstract No. B026.
51. Zimlichman R, et al. The seven European sites study of arterial elasticity—using the blood pressure waveform analysis—reliability, repeatability and establishment of normal values for healthy European population with comparison to healthy US population. *Am J Hypertension* May 2003;16: (No. 5, Part 2):P–315.
52. Blacher J, et al. Aortic pulse wave velocity as a marker of cardiovascular risk in hypertensive patients. *Hypertension* 1999;33:1111–1117.
53. Lacy PS, et al. Increased pulse wave velocity is not associated with elevated augmentation index in patients with diabetes. *J Hypertension* 2004;22:1937–1944.
54. Salvi P, et al. Validation of a new non-invasive portable tonometer for determining arterial blood pressure wave and pulse wave velocity: The PulsePen device. *J Hypertension* 2004;22:2285–2293.
55. Weiss BM, et al. Radial artery tonometry: Moderately accurate but unpredictable technique of continuous non-invasive arterial pressure measurement. *Br J Anaesthesiol* 1996; 76(3):405–411.
56. Toy S, Melbin J, Noordergraaf A. Reduced models of arterial systems. *IEEE Tran. Biomed Eng* 1985;BME-32(2).
57. Quick CM, Berger DS, Noordergraaf A. Apparent arterial compliance. *Am J Physiol—Heart* 1998;274:1393–1403.
58. Karamanoglu M, O'Rourke MF, Avolio AP, Kelley RP. An analysis of the relationship between central aortic and peripheral upper limb pressure waves in man. *Eur Heart J* 1993; 14:160–167.
59. Stergiopoulos N, Westerhof BF, Westerhof N. Physical basis of pressure transfer from periphery to aorta: A model-based study. *Am J Physiol* 1998;274(Heart Circ. Physiol. 43): H1386–H1392.
60. Pressman G, Newgard P. Development of a Blood-Pressure Transducer for the Temporal Artery NASA CR-293, Contract NAS 2-1332, Menlo Park (CA): Stanford Research Institute; 1965.
61. Sato T, Nishinaga M, Ozawa T, Takatsuji H. Accuracy of continuous blood pressure monitor based on arterial tonometry. *Hypertension* 1993;21:866–874.

Further Reading

- Geddes LA. *The Direct and Indirect Measurement of Blood Pressure*. Chicago: Year Book Medical Publishers, Inc.; 1970.
- O'Rourke MF, Kelly RP, Avolio AP. *The Arterial Pulse*. London: Lea & Febiger; 1992.
- Nichols WF, O'Rourke MF, Hartley C. *McDonald's Blood Flow in Arteries*. 4th ed. New York: Oxford University Press; 1998.

See also BLOOD PRESSURE MEASUREMENT; IMPEDANCE PLETHYSMOGRAPHY; PERIPHERAL VASCULAR NONINVASIVE MEASUREMENTS.

TOOTH AND JAW, BIOMECHANICS OF

SINAN MÜFTÜ
Northeastern University
Boston, Massachusetts

ALI MÜFTÜ
Tufts University
Boston, Massachusetts

INTRODUCTION

The relations among forces, motion, and deformation are studied in the field of mechanics. Biomechanics seeks to understand the mechanics of living systems (1). The teeth and jaw perform the critical function of initiating the digestion by breaking the food into smaller sizes. This increases the surface area of the food and improves the effectiveness of the enzymes involved in digestion. *Mastication* is defined as the action of chewing foods. The masticatory system is composed of

1. The dentition
2. The bones (the maxilla the mandible, and the temporal bone),
3. The ligaments
4. The muscles
5. The temporomandibular joint (TMJ):

The periodontal ligament (PDL), which attaches the teeth to the bones, the articular disk, and the cartilage, which are located on the articulating surfaces of the TMJ, and the other soft tissues such as the blood vessels and the nerves are also parts of this complex. The three main functions of the masticatory system are chewing, swallowing, and speech. This system also helps in expression of emotions and respiration. The temporomandibular system is controlled by the nervous system, and its successful function requires a harmonious relation of these components. Biomechanical investigations of the mastication system aim to understand the fundamental relations between anatomy and function and thus either aid the available treatment modes or help design new ones.

During function (mastication) or rest, various components of the temporomandibular complex work together. Explanation of the complex interrelations between the components of the masticatory system, requires knowledge of the functions of each component. A comprehensive review of the biomechanics of tooth and jaw was given in Reference 2 up to 1988. The aim of the current article is to review the recent developments in this field. In the last few years, the cost and the speed of performing analysis using computational techniques has improved dramatically (3). Thus, many new investigations were enabled particularly by, but not limited to, using the finite element method. This article reflects some of these new analyses. In general, this article is divided into three sections related to the biomechanics aspects of the anatomy, function, and treatments of the masticatory system.

The article starts with an overview of the functional anatomy of the mastication system. Brief anatomical

descriptions of the dentition, the skeletal components, the musculature, the temporomandibular joint/disk, and the connective tissues, such as the ligaments, are given. The monographs given in References 4 and 5 can be consulted for more detailed information.

Next, the biomechanical models of mastication are introduced. This involves investigating the relations between the forces applied by the muscles, and the reaction forces that develop on the teeth, the TMJ, and the ligaments, as well as the deformations of the mandible. In-depth understanding of these relations help treatment of TMJ disorders (such as bruxism), design of endosseous implants, reconstructive prosthetics, and various tooth replacement modalities. Definitions of key engineering concepts such as external and internal forces, internal stresses and strains, and elastic material (stress-strain) behavior are provided. The material properties of all components of the masticatory system are briefly reviewed. The biomechanical fundamentals of mastication are introduced by using force balance relationships, where the mandible is treated as a rigid body. In addition, investigating the deformations of the mandible is briefly covered. The monographs given in References 3,4 and (6–11) are only some of the many references that review these topics in more detail.

Many treatment modalities involving the masticatory system benefit significantly from the studies performed in the fields of biomechanics and biomaterials. These include,

but are not limited to, prosthodontic and orthodontic treatments, reconstructive surgery, and reconstructive prosthetics. In this article, the biomechanical considerations in prosthetic dentistry, including dental implant treatments, intracoronary and extracoronary restorations, and fixed and partial dentures, are reviewed.

DENTITION AND SUPPORTIVE STRUCTURES

Human dentition is composed of a total of 32 teeth (Figs. 1 and 2). Each tooth has a part that is visible, called the *crown*, and a part located inside the bone called the *root* as shown in Fig. 3. The teeth are attached to the bone by a soft tissue called the PDL, which serves to connect the tooth to the surrounding bone and distributes the occlusal loads to the bony tissue. On the superior aspect (top portion) of the masticatory system, 16 teeth are attached to the maxillary arch. Likewise, on the inferior aspect (bottom portion), 16 teeth are attached to the mandibular arch. As the sizes of the maxillary teeth are larger, the teeth in the maxilla are situated on a larger arch.

The teeth on each jaw can be classified in four groups, according to their function as shown in Fig. 2. The *incisors* are used to cut the food into smaller pieces. They are located on the anterior (frontal) section of the mouth. The incisors have a relatively small cross-section; there-

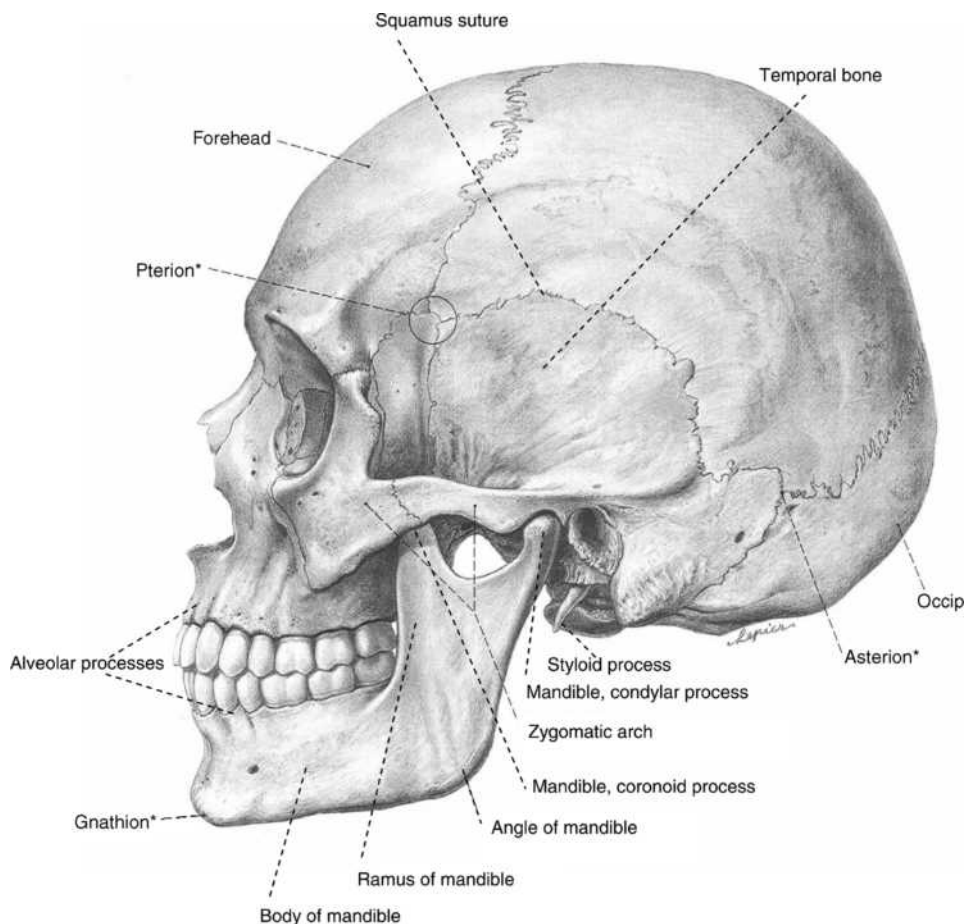


Figure 1. Lateral view of the skull showing the skeletal components of the masticatory system. (From Sobotta Atlas of Human Anatomy, Vol. 1, 13th ed.).

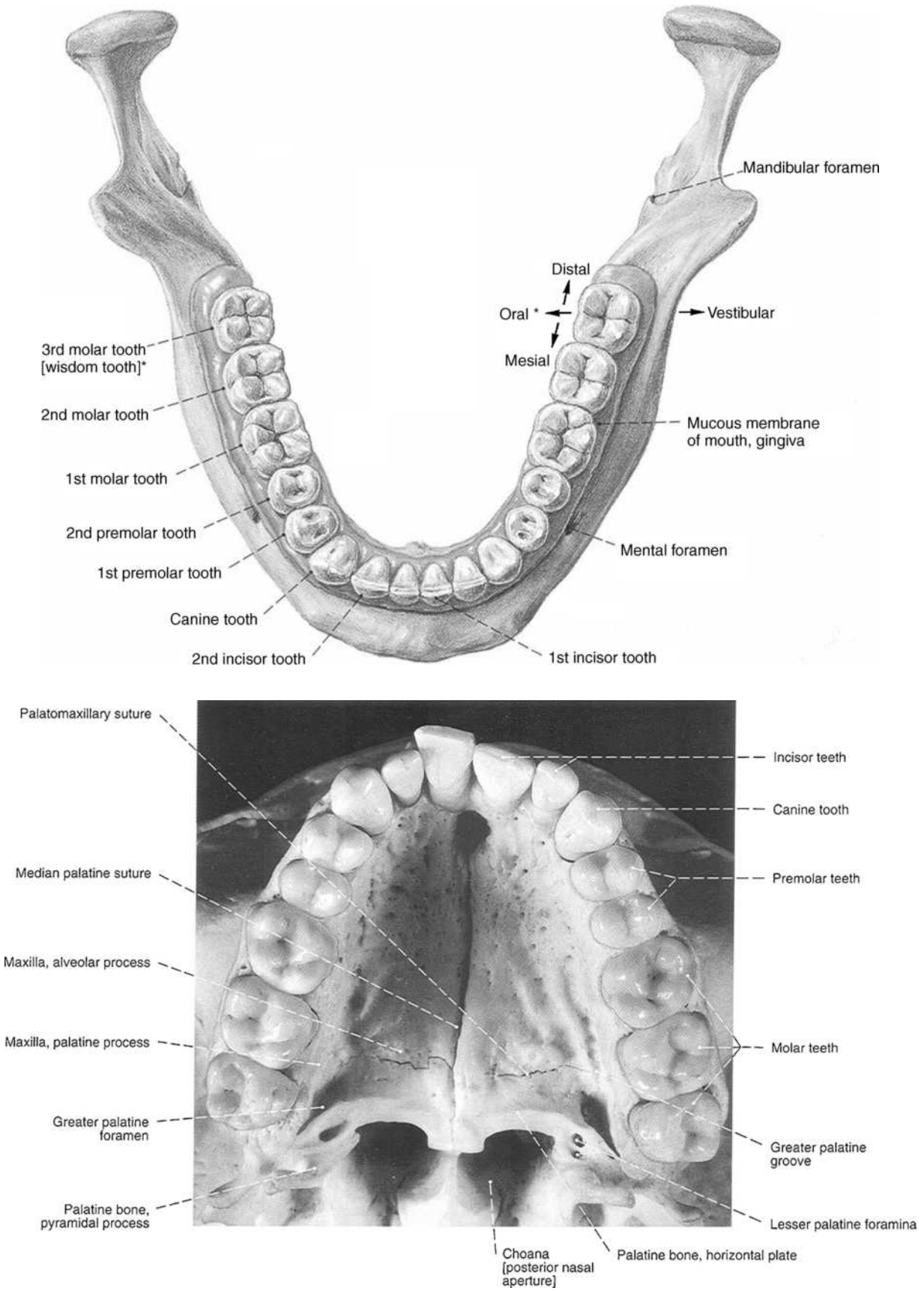


Figure 2. (a) The mandibular dental arch and (b) the maxillary dental arch. (From Sobotta Atlas of Human Anatomy Vol. 1, 13th ed.).

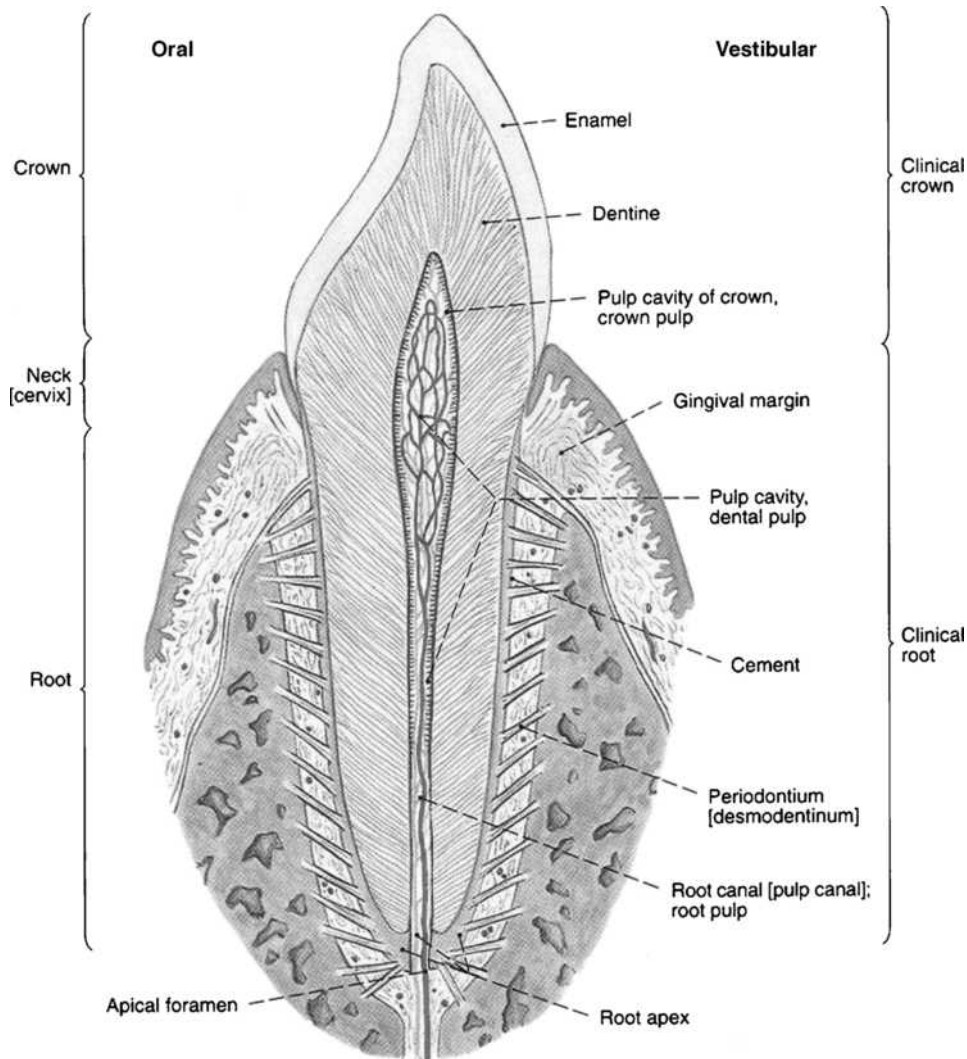


Figure 3. Schematic representation of an incisor tooth showing tooth socket and periodontium. (From Sobotta Atlas of Human Anatomy, Vol. 1, 13th ed.).

fore, they can apply large pressures on the food to provide effective cutting. Four incisors are located on the maxilla and the mandible, each. The *canines* are the long teeth whose function is to tear the food. In humans they mostly function as the incisors and cut the food. There are two incisors on the mandible and two on the maxilla. The *premolars* crush the food to smaller sizes during rhythmic, repetitive phases of mastication. Six maxillary and six mandibular *molars* are located at the ends of lower and upper dental arches. Molar and premolar teeth have multiple cusps and groves, which provide a relatively large surface area, enabling more effective crushing of the food.

The Teeth

Each tooth has two main sections, a *crown* and a *root* or roots as shown in Fig. 3. Each section is further subdivided to crown, cervix (neck), root, and apex. The tooth is solid except for the *pulp cavity* centered within it. The major portion of the tooth is made of *dentin*. A layer of *enamel* covers the crown portion of the tooth mostly above the gum line. A thin layer of *cementum* covers the root set in the bony tissue. The *cementum* is true bone. The pulp cavity may be divided into

two portions; the *pulp chamber*, which is mostly in the crown, and the *pulp canal* traversing the interior of the root, ending in a constricted opening at the root apex. The pulp cavity contains the *dental pulp*, a soft tissue containing connective tissue, blood vessels, and nerves (12).

Both the dentin and the enamel contain collagen, hydroxyapatite (HAP) (an inorganic molecule), and water. However, the distributions of these constituents are different in these structures: The dentin contains 47% HAP, 30% collagen, and 23% water; and; the enamel contains; 92% HAP, 2% organic material, and 6% water, by volume. The microstructure of the enamel is characterized by parallel rods of HAP. In the dentin the tubules are connected by organic material. Because of the difference of HAP content and the microstructure, the modulus of the enamel can be as much as four to six times as high as that of dentin (4). See Table 1 for a summary of the elastic properties of the tooth and the periodontium.

The PDL

The general name given to the attachment mechanism of the teeth is the *periodontum*, which consists of the

Table 1. Elastic Properties of the Enamel, Dentin Layers of a Tooth and the PDL

	Elastic Modulus E [GPa]	Poisson's ratio ν	Tensile Strength [MPa]	Compressive Strength [MPa]	Shear Strength [MPa]
Enamel	80 ^{b,c}	0.3 ^b	10 ^a	288–400 ^a	8 ^a
Dentin	14 ^d , 15 ^b , 18 ^a , 20 ^b	0.15 ^d , 0.31 ^{b,c}	48 ^a	232–297 ^a	20 ^a
PDL	0.002 ^c , 0.003 ^a , 0.05 ^b , 10 ^d	0.45 ^c , 0.49 ^{b,d}			

^aFrom Toparli et al.(13)^bRees and Jacobsen (14)^cArola et al.(15)^dImanishi et al.(16)

cementum, the PDL, the alveolar bone, and a portion of the gingiva (Fig. 3) (4). The PDL is a connective tissue present between the root of the tooth and the alveolar bone, whose function is (1) to provide support for the teeth and (2) to control the distribution of the occlusal loads on the bony tissue (5). Its thickness varies between 0.15 and 0.38 mm, in humans. Perhaps a good indication of its function can be observed when it is realized that the magnitude of occlusal loads affects the thickness of the PDL, where increased load results in thickening of the PDL.

Among other cellular components of the PDL are the osteoblast and osteoclast cells associated with the alveolar bone; cementoblast and cementoclast cells that associated with the cementum; and fibroblast cells that are responsible for collagen generation. The extracellular components of the of the PDL are the collagen fibers, oxytalan fibers, nerves, vessels, and the ground substance, which is composed of hyaluronic acid, glycoproteins, proteoglycans, and water (4).

Approximately 65% of the PDL's volume is occupied by dentoalveolar fiber bundles. The collagen molecules (type-I and III) of PDL are wrapped into collagen *fibrils* (55 nm diameter), which are wrapped into *fibers*. The fiber bundles are arranged into networks having a complex three-dimensional overlapping arrangement (4). These fiber bundles provide the load-bearing capacity to the PDL.

It has been mentioned that one of the primary functions of the PDL is to distribute the forces acting on the teeth to the bony tissue. The force–displacement relationship of the PDL is nonlinear. When a tooth is subjected to an external force, its initial displacements are caused by relatively small forces until the force magnitude reaches 1 N. Thereafter, increasingly higher forces are required to displace the PDL. The PDL is stiffer under axial loads acting on the tooth as compared with tangential loads. The same level of load applied tangentially causes a larger displacement of the PDL, as compared with the axial load. Like similar connective tissues, the PDL also shows viscoelastic material behavior.

THE SKELATAL COMPONENTS

The major bones of the human skull that are involved in mastication are the *mandible*, the *maxilla*, and the *temporal* bone. The *maxilla* is composed of two parts that are connected at the midpalatal suture, as shown in Fig. 2b. A large portion of the facial bone is composed of the maxillary bones. On the frontal plane, the maxilla is connected to the

zygomatic bone and the *nasal bone*, and on the sides of the skull, it is connected to the temporal bone as shown in Fig. 1. The teeth on the maxillary arch are connected to the maxilla on the alveolar ridges. These teeth are considered to be the fixed part of the masticatory system, as the maxilla is fixed to the skull.

The *mandible* is the arch-shaped bone that forms the lower part of the facial skeleton and the masticatory system as shown in Fig. 2a. On the posterior (back) sides, the mandible extends vertically. This vertical (ascending) extension of the mandible is the *ramus*. The *mandibular angle* located at the posterior (back) part of the ramus is shown in Fig. 1. The superior (top) extension of ramus forms two processes: The anterior (frontal) one is the *coronoid process*, and the posterior (back) one is the *condyle*. These two processes make critical attachments to the rest of the masticatory system. The mandible is connected to the temporal bone at the condyle through the TMJ. The TMJ allows pivoting and sliding of the mandible with respect to the fixed part of the masticatory system, and it consists of various ligaments, the articular disk, the synovial capsules, and the synovial fluid. The articulating surfaces of the bones in the TMJ, namely, the condyle and the fossa, are covered with a fibrous tissue called the *articular cartilage*. The coronoid process serves as one of the endpoints of the temporal muscle. Various muscles and ligaments are attached to the mandible, enabling the mastication function. On the front part of the mandible, the teeth are connected through the *alveolar ridges*.

Two *temporal* bones are located on each lateral side of the human skull as shown in Fig. 1. The temporal muscles are connected to the wide area called the squamus part of the temporal bone. The mandible articulates in the concave part of the temporal bone called the *mandibular fossa*, located below the squamus part and near the zygomatic process as shown in Fig. 1. The mandibular fossa is also called the *articular fossa*; or the *glenoid fossa*. The *articular eminence* is located immediately in front of the mandibular fossa. The condyle of the mandible articulates on the maxilla through the TMJ. In the initial phases, of the opening of the jaw, the condyle rotates in the mandibular fossa; however, in the later phases; it moves forward and slides along the articular eminence. The thickness of the bone in the posterior part of the mandibular fossa is relatively thin, whereas thicker bone is found in the anterior part and in the articular eminence. This is an indication of the load-bearing nature of these surfaces, where bone thickness is larger in sections subjected to larger stresses.

THE MUSCULATURE OF THE MASTICATION SYSTEM

The masticatory muscles are divided into *depressor* and *elevator* groups (6). The depressors are the *digastric*, *suprahyoid*, and *infrahyoid* muscles, and the elevators are the *temporal*, the *masseter*, the *medial pterygoid*, and the *lateral pterygoid* muscles (5).

The depressor muscles are located in the floor of the mouth. The supra- and infrahyoid muscles connect the hyoid bone to the mandible. The digastric muscle connects the mastoid process of the skull with the mandible, and it is attached to the hyoid bone through a tendon. These muscles are primarily involved during jaw opening and swallowing.

The three muscles, which form the fan-shaped muscle, shown in Fig. 4a, are collectively called the *temporal muscle*. The *anterior (front) temporal muscle* is composed of vertically oriented muscle fibers. The fibers orientation turns gradually toward the horizontal direction in the *middle temporal muscle* and the *posterior (back) temporal muscle*. The muscle fibers are attached to the temporal bone on the open part called the squamus of the temporal bone. These fibers come together as they descend downward through the zygomatic arch and form a tendon. This tendon attaches to the coronoid process and anterior border of the ascending ramus of the mandible.

The temporal muscle is both an elevator and a positioner. It can function unilaterally, bilaterally, or in sections to position and elevate the mandible. In bilateral closure, this muscle moves the condyle into the mandibular fossa. In unilateral action, the posterior temporal muscle moves the mandible toward the active side.

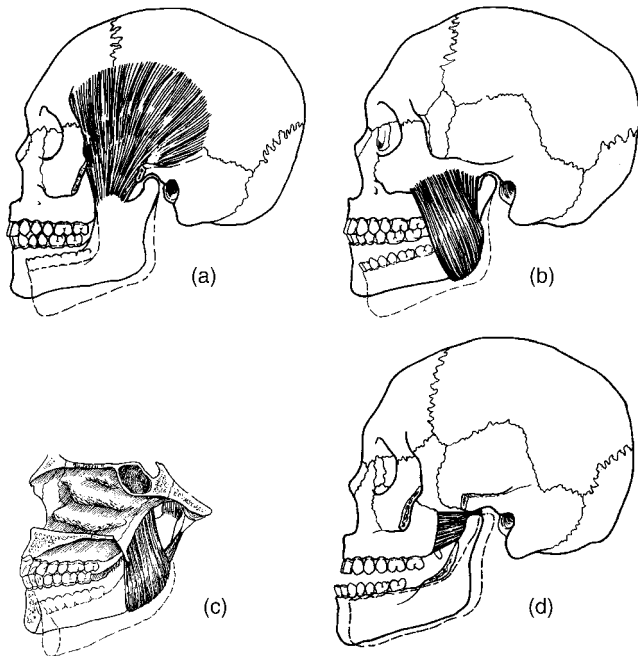


Figure 4. The elevator muscles of mastication: (a) the temporalis, (b) the masseter, (c) the medial pterygoid, and (d) the lateral pterygoid. (Modified from Temporomandibular Disorders and Occlusion by Okeson JP, 5th ed., Mosby, 2003.)

The *masseter* is located on the distal (outer) lateral sides of the ramus of the mandible, as shown in Fig. 4b. This muscle extends from the angle of the mandible upward and attaches to the zygomatic arch. The masseter is the most powerful elevator muscle in the masticatory system, and it is responsible for the high loads in the molar area. The crushing forces on the molar area could become high when the masseter acts together with the internal pterygoid and the anterior temporal muscles.

The *medial (internal) pterygoid* muscle is the internal counterpart of the masseter as shown in Fig. 4c. The fibers of this muscle originate from the *pterygoid fossa* of the *sphenoid bone*, located internally in the mouth, extend downward, internally, and connect to the internal face of the mandibular angle, as shown in Fig. 4c. The medial pterygoid and the masseter function in a coordinated manner, and together they can generate high loads. Although the primary function of the medial pterygoid is to elevate the mandible, it can also move the mandible medially (toward the center).

The *lateral pterygoid* muscle: is the collective name of the two muscles: the *upper (superior)* and *lower (inferior) lateral pterygoid*, as shown in Fig. 4d. The upper lateral pterygoid muscle originates from the sphenoid bone located internally in the mouth, extends horizontally, and attaches to the condylar neck, the capsular ligament, and the articular disk. This muscle is active along with the elevator muscles during closing the teeth together for chewing or clenching. The upper lateral pterygoid muscle is a weaker muscle as compared with the lower.

The lower lateral pterygoid muscle originates at the pterygoid plate (Fig. 4d). It extends backward and upward connecting to the neck of the condyle. Bilateral contraction of the right and left lower pterygoids pulls the condyles down, out of the articular eminences, and the mandible moves forward, as shown in Fig. 4d. Unilateral contraction causes movement of the mandible to the opposite side.

THE TMJ

The relative motion of the mandible with respect to the temporal bone takes place at the TMJ, as shown in Fig. 5a. This is where the condyle of the mandible sits against the temporal articulating surfaces of mandibular fossa and the articular eminence. The TMJ is composed of the condyle of the mandible, the articulating surfaces of the temporal bone, the articular cartilage and disk, the ligaments, and the muscles. The articular disk is considered to act as a non-ossified bone, between the mandibular fossa and the condyle; hence, the TMJ is considered a compound joint of the human body.

The articular disk deserves special attention, as it facilitates the relative motion to take place without direct bone-to-bone contact. The disk is located between the condyle and the temporal fossa. This is a fibrous tissue, which does not have blood vessels and nerves for the most part. The articular disk is mainly a mesh of *collagen fibers* whose interstices are filled with *proteoglycans*. In the TMJ disk, the collagen fibers help maintain its shape during loading, whereas the elastin fibers function to recover the form after

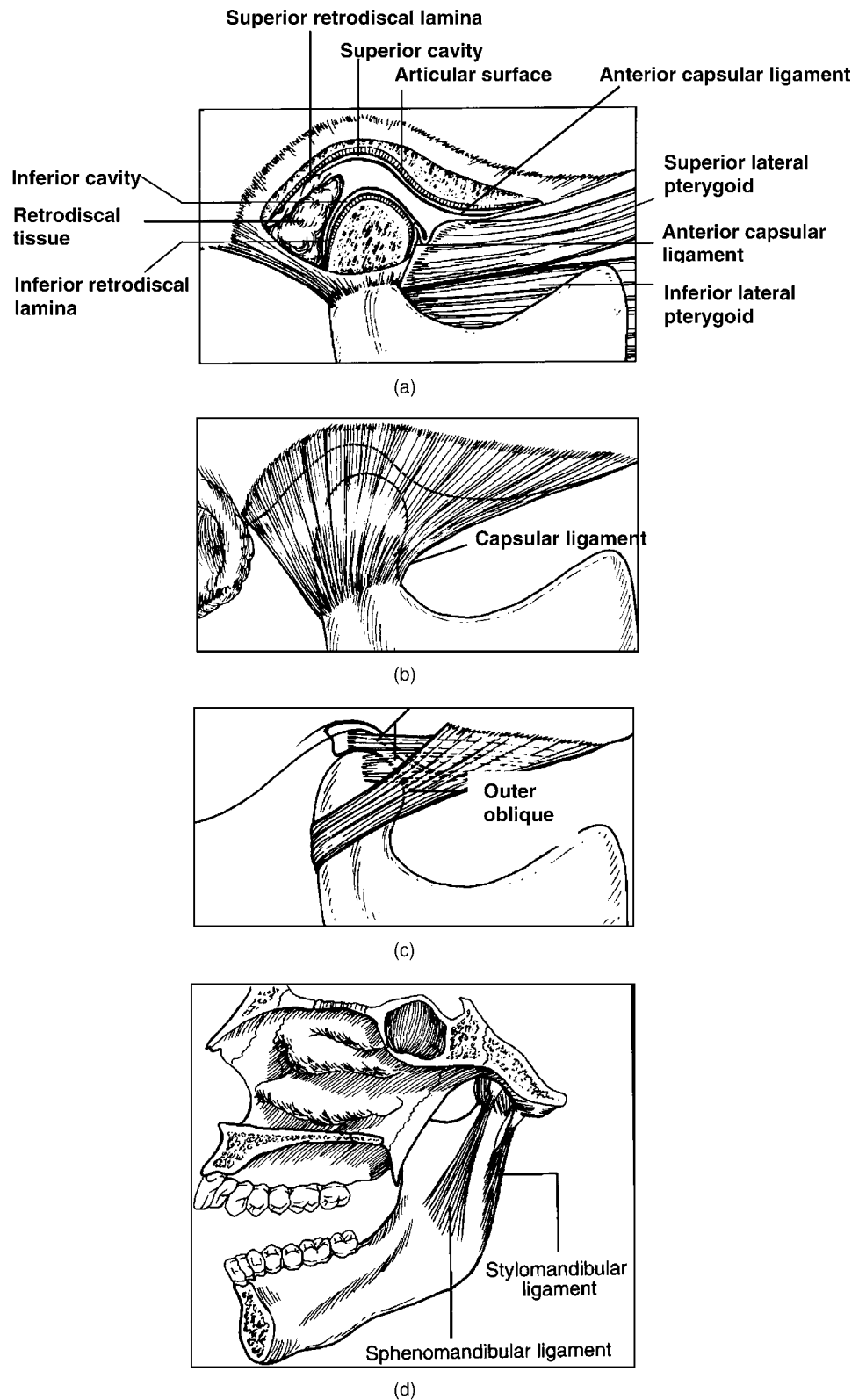


Figure 5. Schematic representation of the articular disk and the ligaments of the masticatory system: (a) the articular disk and its connecting tissues, including the collateral ligaments; (b) the capsular ligament; (c) the tempo-romandibular ligament; and (d) stylomandibular and sphenomandibular ligaments. (Modified from *Temporomandibular Disorders and Occlusion* by Okeson JP, 5th ed., Mosby, 2003.)

unloading. The mechanical properties of the articular disk strongly depend on the collagen fiber and the proteoglycan composition and organization (17). Experimental evidence suggests that the elastic modulus of the articular disk increases with age (18). The load-displacement (stress-strain) behavior of the articular disk is nonlinear and

viscoelastic. A poroelastic material behavior model where the solid matrix behavior was hyperelastic showed several similarities with dynamic indentation tests of articular disks (19).

The articular disk is mainly supported by ligaments, as shown in Fig. 5a. In the posterior (back) region of the TMJ,

the articular disk is attached to a loose connective tissue named the *retrodiscal tissue* (RT). The superior (top) portion of the RT is attached to the tympanic plate by the *superior retrodiscal lamina*, and its inferior (bottom) part is attached to the condyle by the *inferior retrodiscal lamina*. The medial, lateral, anterior, and posterior surfaces of the articular disk are attached to the *capsular ligament*, which surrounds most of the surfaces of the condyle and the articular surface of the temporal bone. In the front, the articular disk is also attached to the superior (upper) lateral pterygoid muscle (5).

The capsular ligaments divide the joint into the *upper* and *lower cavities*. The internal surfaces of the cavities are covered with endothelial cells. The cavities are filled with *synovial fluid* produced by these cells. The synovial fluid serves as a lubricant in the TMJ and reduces the frictional forces between articulating surfaces. Depending on the joint speed and load, different lubrication regimes are thought to be responsible for this effect (7). These include *boundary lubrication*, *elastohydrodynamic lubrication*, and *hydrodynamic lubrication* modes (20). Another effect, which possibly contributes to joint lubrication, is the *weeping lubrication*, which takes place as the synovial fluid, retained in the articular cartilage, is forced out of the cartilage into the synovial cavity, under sufficient normal pressure. The weeping lubrication is thought to be more prevalent during elevation and clenching (5).

Two- (21) and three- (22,23) dimensional finite element models of the quasi-static opening of the jaw that included the contact relations among the articular disk, the condyle, and the articular eminence showed that the disk moves together with the condyle. These models predicted that the superior lateral pterygoid muscle and the ligaments attached to the disk do not play a significant role in the disk movement during jaw opening (21,23). The biconcave shape of the disk is sufficient to move the disk with the condyle. These studies also showed that the articular disk is primarily loaded in its intermediary (central) region.

THE LIGAMENTS

Ligaments are the connective tissues between the bones. The main function of the ligaments of the masticatory system is to prevent the mandible from undergoing extreme relative motion. Ligaments also protect the nerves and the vessels that connect to the mandible. In the masticatory complex, there are five main ligaments as shown in Fig. 5: the *collateral (discal) ligament*, *capsular ligament*, the *temporomandibular ligament*, the *sphenomandibular ligament*, and the *stylomandibular ligament*.

The *collateral ligaments* attach the medial (inner) and distal (outer) surfaces of the articular disk to the condyle of the mandible. These attachments along with the anterior (front) and posterior (back) attachments of the articular disk to the capsular ligament create the *synovial cavities*. The attachment of these ligaments permit the motion of the articular disk front-to-back, or in the anterior–posterior direction. Therefore, this ligament allows the disk to travel with the condyle.

The *capsular ligament* encloses the entire TMJ, as shown in Fig. 5b, and thus it provides a sealing function for the synovial fluid. The entire circumference of the articular disk is also attached to this ligament. The capsular ligament is attached superiorly (top) to the temporal bone and inferiorly (bottom) to the neck of the condyle. This ligament resists lateral or inferior (downward) forces.

The *temporomandibular (TM) ligament* consists of an inner *horizontal* and an *outer oblique part*, as shown in Fig. 5c. The TM ligament plays an important role in the pivoting action of the TMJ. Both parts of the TM ligament originate from the zygomatic arch. The inner horizontal part extends from the zygomatic arch horizontally and attaches to the anterior (frontal) neck of the condyle. The outer oblique part extends from the zygomatic arch and attaches to the posterior (back) part of the neck of the condyle, as shown in Fig. 5c. The outer oblique part of this ligament resists excessive dropping of the mandible. During the initial phase of the mouth opening, the condyle can pivot around a fixed point, while this ligament is becoming stretched. When the stretching of this ligament reaches its limit, then the condyle moves downward and forward across the articular eminence to continue opening. It is believed that the *inner horizontal part* of this ligament limits the backward (posterior) movement of the condyle and the articular disk.

The full effect of this ligament in limiting the motion of the articular disk has been debated (9); biomechanical models of the disk movement during mandibular opening and closing have shown that the disk could move in the anterior and posterior directions due to the favorable contact conditions provided by its bicuspal shape (21,24).

The *sphenomandibular ligament* extends from the sphenoid bone and attaches to the *lingula* on the inner (medial) surface of the ramus as shown in Fig. 5d. The function of this ligament is not well understood. However, its main function could be protection of nerves and blood vessels from dislocation and trauma, and it can prevent extreme anterior and lateral dislocations (4).

The *stylomandibular ligament* is another accessory ligament. It is located between the styloid process and the back of the ramus as shown in Fig. 5d. It limits excessive protrusive movements of the mandible (4).

MASTICATION AND ITS BIOMECHANICAL MODELS

Mastication is the action of chewing foods. Mastication involves rhythmic and repetitive motion of the mandible with respect to maxilla. The mastication cycle starts with the *opening phase*, followed by the *crushing phase* and *grinding phase*, which occur during closure, as shown in Fig. 6a. The motion of the mandible during mastication is a three-dimensional, complex motion, which has been described as having the shape of a teardrop or a pear. When the motion of the incisors in an idealized mastication cycle is viewed from the frontal plane (Fig. 6a), this analogy becomes clear. During the opening phase, the mandible drops vertically for about 15–18 mm; thereafter, it moves

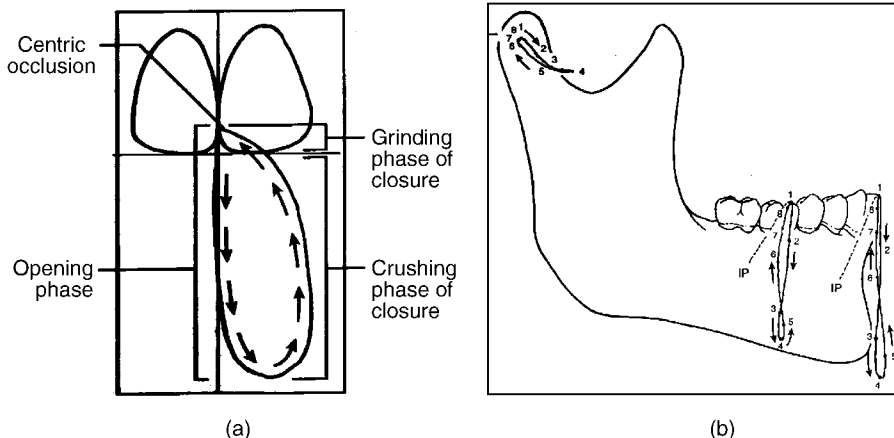


Figure 6. The mastication cycle viewed from (a) the frontal plane; and (b) the sagittal plane. (Modified from Temporomandibular Disorders and Occlusion by Okeson JP, 5th ed., Mosby, 2003.)

laterally and the closure phase begins. During the initial phase of closure, the incisors move 4–5 mm laterally while crushing the food. This phase lasts until incisors are located 3–4 mm laterally and 3 mm vertically with respect to their initial (intercuspal) condition. During the grinding phase, which follows, the food is sheared between the cusps of the incisors. The view of the motion of the incisors from the side (sagittal) plane on the working side, shown in Fig. 6b, shows that during the opening phase, the incisors move slightly in the frontal (anterior) direction, followed by a posterior motion during closing. The TMJ and the premolars also follow a path similar to that of the incisors in the sagittal (side) plane on the working side.

The amplitude of the anterior and the lateral movement of the mandible depend on the stage of mastication. During the initial phases, the incisors are used to cut the bolus (food) and the anterior and lateral movement are relatively large. During the later phases, the posterior teeth are used more, and the lateral movement of the teeth is reduced. The consistency of the food also affects the lateral movement, where harder foods require larger lateral movements and more chewing cycles, as shown in Fig. 7.

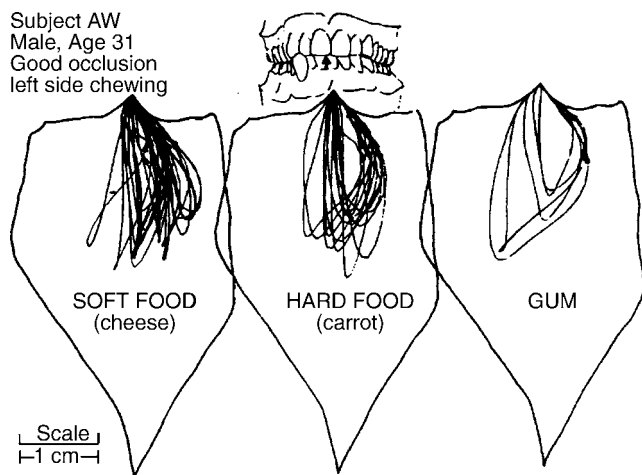


Figure 7. Frontal view of the chewing stroke from a 31-year-old man with good occlusion, for soft food, hard food, and chewing gum. (From Advances in Occlusion, by Lundeen HC, Gibbs CH, Boston, 1982).

In biomechanical analysis of mastication, static analysis are carried out to determine maximum clenching forces that can be applied by the muscles; dynamic analysis, on the other hand, provides information about the muscle TMJ interactions during the open and close cycles, as well as laterodeviations. In these analyses, the mandible is modeled as a rigid body, and the muscle forces are modeled as concentrated forces, as shown in Fig. 8. The attachment points and the three-dimensional vectorial orientations of the muscles are typically, carefully, measured from cadaver specimens. The biomechanical analysis of mastication also considers the deformations of the mandible due to external forces.

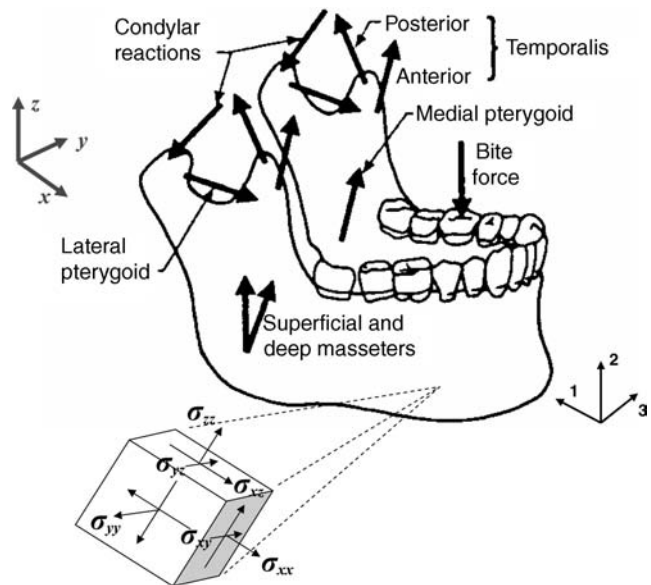


Figure 8. The locations and directions of the muscle force vectors, the bite force vector, and the condylar reaction force vectors are shown schematically. The x -, y -, and z -coordinate system indicates the coordinate system in which the forces are measured. The 1-, 2-, and 3- coordinate system indicates the transversely isotropy directions for material properties of the mandible. The small stress cube indicates the internal stresses generated at a point inside the mandible in response to external loading. (Modified from Faulkner et al., "A three dimensional investigation of temporomandibular joint loading," Journal of Biomechanics 1987;20:997–1002.)

In the next section, a classification of the internal and external forces acting on a deformable body will be described. The material property definitions and deformation behavior of the tissues involved in the masticatory system is described next. Finally, some static, dynamic, and elastic models of mastication developed in the last 20 years will be described.

In general, the forces acting on a structure can be classified as *external forces* and *internal forces*. The external forces can be further classified as *surface tractions* (external pressure), *concentrated forces*, and *body forces*.

External Forces

Surface tractions are external forces distributed over a finite area, on the surface of the structure. For example, the contact pressure between the condyle of the mandible and the articular disk takes place over a relatively large area. This contact pressure is a normal traction acting on the mandible (and the disk). Note that distribution of this traction can vary from point-to-point. In the metric system, the unit of traction is Pascal (Pa), where $1 \text{ Pa} = 1 \text{ N/m}^2$.

Concentrated force is a term used for an idealized traction acting over an infinitesimally small area. In the metric system, the unit of the concentrated force is Newton (N), where $1 \text{ N} = 1 \text{ kg}\cdot\text{m/s}^2$. Use of concentrated forces is common in mechanics as they simplify the analysis in many instances. For example, in many mechanical models of the mandible, the forces exerted by the muscles, the forces experienced by the teeth during clenching or mastication, and the reactions on the condyle are idealized as concentrated forces as shown in Fig. 8. While idealizing the muscle forces, investigators spend a great deal of effort to ensure that the idealized forces represent the (actual) tractions in a mechanically equivalent manner.

Both surface tractions and concentrated forces are transferred from one body onto another through physical contact. Forces exerted by muscles on the bones are in this category. On the other hand, a *body force*, such as that due to gravity, is a force that acts over a distance, without requiring direct physical contact. For example, if it was not for the slight state of contraction of the clenching muscles, the mandible would stay open due to the effect of gravity.

Internal Forces and Internal Stresses

The internal forces develop inside a structure in response to external forces and enable the structure to stay together. The internal forces can only be "visualized" by taking a virtual cross-section of the structure. Depending on the location and orientation of the cross-section, the internal forces will vary in magnitude and direction. The resultant of the internal forces acting on an infinitesimally small cross-sectional area (dA) of the structure can be decomposed into a *normal force* (dF^n) acting perpendicular to the cross-section and a *shear force* (dF_s) acting in the plane of the cross-section. *Stress* at a point inside the structure can then be defined as the limit of the internal forces acting on the area dA as it becomes infinitesimally small. In general, the internal stress state of a structure can be expressed by six independent stress components σ_{xx} , σ_{yy} , σ_{zz} , σ_{xy} , σ_{xz} , σ_{yz} as shown in Fig. 8. (Note that six indepen-

dent stress components assume that no internal twisting moments exist in the structure. In case these exist, then nine stresses are necessary for this description) Note that in this representation, the first subscript refers to the direction of the normal of the plane on which the stress is acting, and the second subscript refers to the direction of action of the internal force. Thus, σ_{xx} , σ_{yy} , and σ_{zz} are the *normal stress* components, and σ_{xy} , σ_{xz} , and σ_{yz} are the *shear stress* components. It is important to remember that internal stresses are defined at a point inside the structure, and in general, they can vary from point-to-point. Whether they do vary depends on many factors such as the shape of the structure, external force distribution, and material properties. In the metric system, like the tractions, the unit of stress is Pascal.

Deformation and Strain

In general, a structure that is properly fixed on its boundary will deform in response to external forces. The deformation of a point on the structure can be characterized by its displacements. In general, a point P^- before deformation will be displaced to a new location P^+ after deformation. For small displacements, the displacement vector d^- can be expressed in a Cartesian coordinate system. The components of d^- along the x -, y -, and z -axes are u , v , and w , respectively.

Relations governing the mechanics of deformable bodies are expressed in terms of strain, rather than the displacements. In general, *strain* is a nondimensional measure of deformation at a point. The *normal strains* ϵ_x , ϵ_y , and ϵ_z represent the change in length per unit of initial length, along the x -, y -, and z -axes, respectively. The *shear strains* γ_{xy} , γ_{xz} , and γ_{yz} represent the decrease in the right angle initially formed by the sides parallel to the x - y , x - z , and y - z axes, respectively (26).

Material Properties

The load-displacement behavior of a structure depends on the type of material involved. In general, the material properties of a structure are determined by uniaxial tension test, shear test, and hardness test. A uniaxial tension test and a pure-torsion test establish the relations between, for example, σ_{xx} and ϵ_x and σ_{xy} and γ_{xy} by Hooke's law:

$$\sigma_{xx} = E\epsilon_x \quad \text{and} \quad \sigma_{xy} = G\gamma_{xy} \quad (1)$$

where the proportionality constants are the elastic (or Young's) modulus E and the shear modulus G . In general, the three-dimensional stress-strain relations are expressed with the generalized Hooke's law, which can be given in matrix form as

$$\{\sigma\} = [E]\{\epsilon\} \quad \text{or} \quad \{\epsilon\} = [C]\{\sigma\} \quad (2)$$

where the *stress vector* is $\{\sigma\} = \{\sigma_{xx} \ \sigma_{yy} \ \sigma_{zz} \ \sigma_{xy} \ \sigma_{yz} \ \sigma_{xz}\}^T$, the *strain vector* is $\{\epsilon\} = \{\epsilon_x \ \epsilon_y \ \epsilon_z \ \gamma_{xy} \ \gamma_{yz} \ \gamma_{xz}\}^T$, the *elasticity matrix* is $[E]$, and the *compliance matrix* is $[C] (= [E]^{-1})$. The elasticity matrix of an *anisotropic* material has 21 independent components. In case the material has three orthogonal planes of symmetry, the material is said to be *orthotropic*. The compliance matrix for an orthotropic

material is defined by nine independent material properties (27):

$$[C] = \begin{bmatrix} 1/E_1 & -v_{21}/E_2 & -v_{31}/E_3 & 0 & 0 & 0 \\ -v_{12}/E_1 & 1/E_2 & -v_{32}/E_3 & 0 & 0 & 0 \\ -v_{13}/E_1 & -v_{23}/E_2 & 1/E_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/G_{23} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/G_{31} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/G_{12} \end{bmatrix} \quad (3)$$

where E_i are the Young's moduli, G_{ij} are the shear moduli, and v_{ij} are the Poisson's ratios in the respective directions, with the restriction that $v_{12}/E_1 = v_{21}/E_2$, $v_{13}/E_1 = v_{31}/E_3$, and $v_{32}/E_3 = v_{23}/E_2$. The number of independent elastic constants is reduced to five for a *transversely isotropic* material, such as the bone, where each plane through a longitudinal axis is a plane of elastic symmetry. For fully *isotropic* materials, the material properties are the same in all directions, and two independent properties (E , ν) are sufficient to describe the stress-strain relationship.

Mechanical Properties of Bone

Bone is composed of collagen, water, hydroxyapatite mineral, and small amounts of proteoglycans and noncollagenous proteins. Collagen is the structural protein that gives the flexibility and tensile strength, also found in ligaments and the articular disk. Hydroxyapatite, $\text{Ca}_{10}(\text{PO}_4)(\text{OH})_2$, is a crystal with hexagonal symmetry located within and between collagen fibers, found in the form of needles, plates, and rods. The internal structure of bone is porous. Two distinct porosity ranges give rise to the *cortical* (compact) bone with low porosity (5–10%) and the *trabecular bone* with 75–95% porosity. The interstices are filled with marrow, a tissue composed of blood vessels, nerves and various types of cells (28).

Among these cells are the *osteoclasts*, which are responsible for formation of new bone, and *osteoblasts*, which are responsible for resorption of bone. *Remodeling* of bone involves repair of internal cracks, and so on, due to coupled activity of osteoclasts and osteoblasts. In addition to self-healing, the bone also has the ability to *adapt* to variations in imposed stresses (Wolff's law). It has been hypothesized that the strain in the bone tissue stimulates the biological response resulting in adaptation (28). The physiological loading zone has been reported to be in the 1000–3000 microstrain range (28). Bone adaptation has significant biomechanical consequences, in design of endosseous implants, and prosthetics. The loading conditions created in the bone by these should be carefully considered to prevent excessive or insufficient loading.

The cortical bone is a transversely isotropic material, as two of its principal material directions have similar mechanical constants. The stiffest direction of the mandibular cortical bone is along a tangent to the parabolic curve of the mandible, i.e., direction -3 in Fig. 8 (29). The mechanical properties of the *cortical bone* depend on several fac-

Table 2. Elastic Properties of Cortical and Trabecular Bones Measured for a Transversely Isotropic Material Model

	Transversely Isotropic	
	Cortical Bone ^a [GPa] (30)	Trabecular Bone ^b [Gpa] (30)
$E_1 = E_2$	13.0	0.27
E_3	19.0	0.82
G_{12}	5.3	0.12
$G_{23} = G_{31}$	5.9	0.12
$\nu_{12} = \nu_{21}$	0.22	0.19
$\nu_{31} = \nu_{32}$	0.42	0.34
$\nu_{13} = \nu_{23}$	0.29	0.11

^aMeasured for mandible using ultrasonic techniques.

^bMeasured for human tibia.

tors, including the porosity of the bone, mineralization level, bone density, collagen fiber organization, and rate of deformation (28). The ultimate stress of the cortical bone has been reported to be higher in compression (170 MPa) than in tension (100 MPa) (28).

The mechanical properties of the *trabecular (cancellous) bone* depend on the porosity, the orientation of the trabecular architecture, and the material properties of the tissue in the individual trabeculae (28). The strength of the trabecular bone has been reported to be the same in tension and compression, and it is approximately 2–5 MPa (28).

Experimental data for the transversely isotropic material properties of the cortical bone of the mandible were obtained by Carter as reported by Hart et al. (30). These values are reported in Table 2. This table also contains the trabecular bone properties measured for human tibia. In other studies, many investigators assumed that the jaw bone deforms in an isotropic manner. The elastic modulus values used in these finite element analysis models ranged between 11 and 20 GPa for the cortical bone and between 0.2 and 7.9 GPa for the trabecular bone. The Poisson ratio was typically taken in the range of 0.3–0.33 (31–35).

Mechanical Properties of the Cartilaginous Tissue of the Masticatory System

The cartilagenous tissue of the masticatory system is the *articular cartilage* on the articular surfaces of the TMJ and the *articular disk*. In general, this type of tissue is composed of cells (chondrocytes) imbedded in intercellular matrix, permeated by a system of fibers (1). Depending on the function of the cartilage tissue in the body, the appearance and the nature of the intercellular matrix can be different. For example, the *yellow elastic fibrocartilage* is found in external ears and larynx; the *hyaline cartilage* is found in nasal, tracheo-bronchial, and articular surfaces; and the *white fibrocartilage* is found in intervertebral and articular disks (1).

In indentation experiments, the cartilage tissue shows an instantaneous recovery, followed by a time-dependent creep deformation, when it is released after compression. Full recovery to the initial state takes place in a finite amount of time. Such material behavior is termed *viscoelastic*. The cartilage tissue is porous, and the synovial

fluid fills the interstices of the tissue. Under compressive stress, the fluid can move out of the tissue, and upon relaxation, it will return to interstitial openings. This fluid flow provides lubrication to the joints as well as nutrients to the cartilage cells (1). The viscoelastic behavior is due to (1) the inherent viscoelastic nature of the solid matrix and (2) the velocity differences between the solid matrix and the liquid (11).

The mechanical response, biomechanical function, and biological integrity of the cartilaginous tissue are governed by the movement of the interstitial fluid (1) *within* cartilage and (2) *across* the articular surface. The synovial fluid is transported through the porous membrane due to a pressure gradient across the tissue. Flow also occurs due to the deformation of the cartilage matrix; the pressure of the fluid internal to the matrix rises in response to deformation. The resulting pressure gradient with respect to the articular surface causes the fluid flow. The biphasic model for the mechanical behavior of the articular cartilage was formulated by Mow et al. (36). In this model, the tissues are modeled as consisting of a soft, permeable, elastic solid, mixed with water. The fluid and the solid phases are coupled through the pressure. Experimentally observed viscoelastic behavior of the articular cartilage can be explained by this biphasic model (36).

Although the articular cartilage behaves as a viscoelastic material due to its poroelastic nature, it can undergo large deformations. Experimental evidence further suggests that the material properties are nonhomogeneous and anisotropic. The experimental measurements of the load-displacement behavior of the articular cartilage of the rabbit TMJ have indicated that the elastic modulus (E) depends on the location with respect to the joint, varying between 0.95 and 2.34 MPa (37). These values are on the same order of magnitude as that reported by Woo et al. obtained for bovine humeral articular cartilage (11). The elastic modulus and Poisson's ratio of the articular cartilage of the temporal joint from these references are given in Table 3.

The articular cartilage is a connective tissue that provides very small frictional resistance to the sliding motions of the joints, in general. Experimental studies to determine the coefficient of friction (COF) show that the presence of the synovial fluid plays a significant role in its low level. Although it has been found that the COF can be as low as 0.005, it has also been shown that it depends on the level of the normal stress and the motion of the joint. In general, under large normal stress values, the COF is found to increase. Moreover, the static friction coefficient is larger

than the dynamic friction coefficient (1,18,38). This behavior is in contrast to the behavior of the COF between most other materials (20).

Tension and compression tests of human *articular disk* show nonlinear stress-strain behavior (14,39). Chin et al. measured the viscoelastic properties of human TMJ disks (40). Chen et al. (39) assumed that the solid matrix of the articular disk can be modeled as a hyperelastic material of the Mooney–Rivlin type (41), where the strain energy of an incompressible rubber-like material is expressed as

$$U = c_1(I_1 - 3) + c_2(I_2 - 3) \quad (4)$$

where I_1 and I_2 are the first and second invariants of the Cauchy–Green deformation tensor and c_1 and c_2 are empirically determined constants (41). Chen et al. used $c_1 = 27.91$ MPa and $c_2 = -20.81$ MPa, based on experimental data obtained from dog articular disks (39).

More recently, Tanaka et al. obtained experimental results for human articular disks (18). Although their data show the same trend predicted by the Mooney–Rivlin type of hyperelastic behavior, they approximated the curve by a single elastic modulus calculated at 2% strain level, and used a linear FEA. Chen and Xu (42) and Hu et al. (37) used a linear piecewise continuous material model. Other investigators assumed that the articular disk behaves strictly in a linear-elastic manner (19,21,43). The data for elastic modulus (E) used in these work are reported in Table 4.

Mechanical Properties of the Ligaments

The physiological function of a joint ligament is to provide stability to joints and to limit their range of motion (45). Ligaments can resist only tensile forces. The structural component responsible for the tensile strength of ligaments is the protein, collagen. Other components of this tissue are elastin, proteoglycans, and glycoproteins. The basic molecular collagen unit is a left-handed molecule. Collagen molecules are wound into super-helices, microfibrils, and fibrils in a hierarchical manner. At each level, the orientation of the helix is reversed. This alternating helix direction is useful in converting the axial tension to circumferential compression, and it is a contributing factor to the strength of the ligaments (45).

The load-displacement curve of the ligament under tensile loading initially displays nonlinear behavior with tangent modulus increasing with increasing strain. Increasing the strain even further, eventually a linear

Table 3. Elastic Properties of the Articular Cartilage

Reference	Location	Elastic Modulus E [MPa]	Poisson's Ratio ν
Hu (37)	Anterior	2.34 ^a	0.46
	Central	1.48 ^a	0.39
	Posterior	1.51 ^a	0.41
	Medial	1.11 ^a	0.38
	Lateral	0.95 ^a	0.31
Woo (11)	–	0.79	0.4

^aMeasured from rabbit mandibular condyle.

Table 4. Elastic Properties Used for the Articular Disk Assuming That It Behaves in a Linear Elastic Manner, Used in the FEA of the TM

Reference	Elastic modulus E [MPa]	Poisson's Ratio ν
DeVocht (21)	1.8	0.4
Beek (1)	6.0	0.4
Beek (1)	6.8	0.4
Tanaka (18)	47.1 ^a	
Hu (44), Chen(42)	44.1 ^b and 92.4 ^b	0.4

^aThis value is measured at 2% strain, and it is the average of seven human specimens.

^b $E = 44.1$ MPa if $\sigma < 1.5$ MPa and $E = 92.4$ MPa if $\sigma > 1.5$ MPa. These data are based on specimens obtained from dogs.

portion is reached. The post-yielding region shows a decreasing tangent modulus with increasing strain. The elastic modulus (E) of the nuchal (neck) ligament is 7.5 MPa, the ultimate strength, and the ultimate strain are 2.4 and 1.25 MPa, respectively (45). Chen and Xu used nonlinear springs, which can only carry tension, to model the upper, and lower posterior ligaments and the anterior ligament (42). They used data from the talofibular ligament of human ankle and assumed that the stiffness of the ligament would be proportional to its cross-sectional area. Thus, they calculated spring stiffness values that are on the order of 10.9–16.35 kN/m.

Static Models of Mastication

Static models of mastication aim to develop an understanding of the maximum forces applied by muscles of mastication. These models typically treat the mandible as a rigid body. The motion of a rigid body is fully described by the displacements and rotations of its center of mass. During occlusion, the external loads acting on the masticatory system are the forces on the TMJ, the ligaments, the teeth, and the muscles, as shown in Fig. 8.

The static equilibrium, for example, in clenching, requires that the sum of the external force and moment vectors be in balance. Considering the subdivisions of the muscles of mastication (i.e., anterior, medial, posterior temporalis; superficial and deep masseter; superior and inferior lateral pterygoid; and, the medial pterygoid), it can be observed that there will be $N_m = 16$ muscle force vectors ($\vec{F}_i^{(m)}$) acting on the mandible, where N_m is the total number of muscles. Note that the subscript i ($1 \leq i \leq N_m$) is an integer counter used to identify each muscle. Similarly, the reaction force vectors due to the ligaments, due to the contacting teeth, and at the joints of the TMJ are indicated by $\vec{F}_i^{(l)}$, $\vec{F}_i^{(t)}$, and $\vec{F}_i^{(j)}$, respectively. Note that these forces can only attain positive values; muscles only apply tensile forces; the ligaments provide no appreciable resistance when they are compressed; and when the TMJ and the teeth are in contact, the reaction forces are considered to be positive. The static equilibrium of external forces is ensured by the following vector equation:

$$\sum_{i=1}^{N_m} \vec{F}_i^{(m)} = \sum_{i=1}^{N_l} \vec{F}_i^{(l)} + \sum_{i=1}^{N_t} \vec{F}_i^{(t)} + \sum_{i=1}^{N_j} \vec{F}_i^{(j)} = 0 \quad (5)$$

where N_l , N_t , and N_j are the total number of force vectors related to the ligaments, teeth, and joint reaction forces. The location of each force vector is represented by a location vector $\vec{r}_i^{(l)}$. Each force vector causes a moment with respect to the origin of the location vector represented by the vector cross-product $\vec{r} \times \vec{F}$. The moment balance with respect to the origin is expressed by the moment equilibrium equation:

$$\sum_{i=1}^{N_m} \vec{r}_i^{(m)} \times \vec{F}_i^{(m)} + \sum_{i=1}^{N_l} \vec{r}_i^{(l)} \times \vec{F}_i^{(l)} + \sum_{i=1}^{N_t} \vec{r}_i^{(t)} \times \vec{F}_i^{(t)} + \sum_{i=1}^{N_j} \vec{r}_i^{(j)} \times \vec{F}_i^{(j)} = 0 \quad (6)$$

Equations 5 and 6 are a compact representation of the force and moment vectors. In a Cartesian coordinate system, each vector is represented as the vector sum of its components acting along the x -, y -, and z -axes. Thus, equations 5 and 6 each represent three equations of equilibrium along these axes, resulting in a total of six equations.

As mentioned, there are four elevator muscles on each side of the skull. Some investigators consider only these eight muscles during function; therefore, they use $N_m = 8$ (46). In other studies, depending on the level of detail, the total number of muscle forces used can vary up to $N_m = 26$ (47). Similar comments apply for the number of forces on the ligaments, teeth, and joints. Thus, it can be easily recognized that the six equations of equilibrium, represented by equations 5 and 6, are not sufficient in solving for the magnitudes of the numerous unknown forces. Such systems are statically indeterminate.

To find a solution for a statically indeterminate system, additional equations and assumptions are necessary. Osborn and Baragar (47) assumed that the *strain sensors* in the muscle tissue and pressure sensors in the TMJ are activated at a rate proportional to the magnitude of the muscle tension and joint reaction forces. The total output f of these sensors then becomes

$$f = \sum_{i=1}^{N_m} c_i^{(m)} |\vec{F}_i^{(m)}| + \sum_{i=1}^{N_j} c_i^{(j)} |\vec{F}_i^{(j)}| \quad (7)$$

where $c_i^{(m)}$ and $c_i^{(j)}$ are the sensor output rates specific to the muscles and joints, respectively. They further postulated that the central nervous system minimizes the output of the (cost) function f . Thus, the problem of finding the reaction forces in mastication is transformed into a problem where the cost function in equation 7 is minimized subject to satisfaction of equations 5 and 6 and conditions $|\vec{F}_i^{(m)}| \geq 0$ and $|\vec{F}_i^{(j)}| \geq 0$. This defines a *linear programming* problem, typically encountered in economics, and it can be handled with various available approaches (48). Osborn and Baragar investigated the effects of minimizing only the joint forces ($c_i^{(j)} = 1$, $c_i^{(m)} = 0$) or only the muscle forces ($c_i^{(m)} = 1$, $c_i^{(j)} = 0$).

The following muscle groups were considered in their study: the masseter (superficial, deep, anterior, and posterior), the medial pterygoid (anterior and posterior), the temporalis (anterior, medial, and posterior), the lateral pterygoid (superior, upper, inferior), and the digastric. Thus, they had $N_m = 23$. They neglected the effect of ligaments ($N_l = 0$), and assumed a single reaction force ($N_j = 1$) acts on each TMJ. The bite force on different teeth were treated as known external forces. They showed that the model where the cost function involves only the minimization of muscle forces results in more realistic mastication force scenarios. Their model predicted that muscle elements with longer moment arms relative to the joint are activated first. As the bite force increases, a ripple activity spreads into muscles with shorter moment arms.

Hatcher et al. investigated the muscle forces involved in unilateral clenching by using a model that is symmetrical around the mid-sagittal plane, including six muscles: the posterior and anterior temporalis, the deep and superficial masseter, the medial pterygoid, and the lateral pterygoid (49).

Due to the assumption of symmetry, $N_m = 6$. They assumed that the left and right TMJ forces were identical ($N_j = 1$). They compared the mathematical results with the results obtained from an *in vitro* model of a skull. The muscle action in the *in vitro* model was simulated by applying external forces, and occlusal and TMJ forces were measured. The study concluded that more realistic results are obtained in the case where the applied force magnitudes are based on the cross-sectional area of the muscles and the electromyogram (EMG) data, in contrast to when they are solely based on the cross-sectional area of the individual muscles. In a different study, the same authors predicted that the balancing side condylar reaction force is larger than the working side, when occlusion occurs on either one of the molars (25). They also showed that the ipsilateral (same side) condylar reaction force varies considerably as the occlusion direction varies in the parasagittal plane.

Smith et al. used a total of six muscles: the right and left lateral pterygoid, the right and left temporalis, and the right and left masseter/medial pterygoid muscle complex (46). They considered the TMJ load and one bite force. They used a minimization technique in which their cost function was the root-mean-square of the TMJ reaction force. This analysis predicted that the TMJ is loaded over the normal functional range of bite force positions and angles. The magnitudes of the reaction forces on the TMJ were found to vary between 5% and 60% of the occlusal force depending on which teeth the occlusion takes place.

Koolstra et al. (50) considered the effects of the eight clenching muscles: the deep and superficial masseter; the medial pterygoid; the anterior, posterior, and deep temporalis; and the superior and inferior lateral pterygoid. They treated the left and the right sides as being independent; hence, a total of 16 muscles were included. The magnitudes of the TMJ reaction forces and the muscle forces were unknown, but their directions of application were measured from the skull. The bite force, on the other hand, was treated as a known quantity. The total number of unknowns were 18, and the linear programming technique was used. The cost function f was based on the postulation that the relative activity of the most active muscle among all muscles should be as small as possible. The relative activity $|\vec{F}_i^{(m)}|$ of a given muscle i was defined as

$$|\vec{F}_i^{(m)}| \leq \mu |\vec{F}_i^{(m)}|_{\text{Max}} \quad (8)$$

where $|\vec{F}_i^{(m)}|_{\text{Max}}$ is the maximum possible for of the muscle and μ is a coefficient defined below. The maximum muscle force of a muscle element is assumed to be proportional to its physiological cross-section, and it is a known quantity. The objective function was given as

$$f(|\vec{F}_i^{(m)}|) = \mu \quad (9)$$

subject to N constraints defined by equation 8. Using this model, the maximum possible bite forces, for which the masticatory system can remain stable, were predicted. Thus, biteforce and joint-force envelopes for mastication at different teeth were evaluated. *In vivo* validation of these results were carried out by Koolstra and van Eijden (51). An overall good agreement between measured and predicted values was observed.

Dynamic Modeling of Jaw Opening and Closing

Biomechanical analysis of the jaw opening and closing was modeled by Koolstra and van Eijden (24,52–54). The dynamic modeling of the jaw opening and closing involves equations of dynamic equilibrium, where the jaw is considered rigid. Thus, the motion of the mandible can be determined by six degrees of freedom of its center of mass; the three rigid-body displacements $\{u\} = \{u_x, u_y, u_z\}^T$ in the x -, y -, and z -directions, and the three rotations $\{\theta\} = \{\theta_x, \theta_y, \theta_z\}^T$ about the x -, y -, and z -axes. The six equations of motion are given in the general matrix form as follows:

$$[M]\{\ddot{u}\} + [C_u]\{\dot{u}\} = \sum \vec{F} \quad (10)$$

$$[\bar{I}]\{\ddot{\theta}\} + [C_\theta]\{\dot{\theta}\} = \sum \vec{M}_G \quad (11)$$

where $[M]$ is the mass-matrix, $[C_u]$ is the damping matrix for the displacements, $\{\dot{u}\}$ is the velocity vector, $\{\ddot{u}\}$ is the acceleration vector, $[\bar{I}]$ is the rotational inertia matrix calculated with respect to the center of gravity of the mandible, $[C_\theta]$ is the damping matrix for the rotations, $\{\dot{\theta}\}$ is the angular speed vector, and $\{\ddot{\theta}\}$ is the angular acceleration vector. The sum of the external forces ($\sum \vec{F}$) and the sum of the moments ($\sum \vec{M}_G$) with respect to the center of mass are calculated using the same relations given in equations 5 and 6 (55).

In Reference 52 the muscle forces are treated as known external forces. This restriction is removed in later studies. The forces on the TMJs and the teeth were calculated based on contact constraints. This aspect of the model makes it nonlinear. The condyle was approximated as a sphere, and the shape of the articular fossa and articular eminence were approximated by a third-order polynomial, based on the dimensions measured from a skull. The cartilage was modeled as a nonlinear spring. Damping of the jaw motion due to friction, which originates from surrounding soft tissues, was included in the model. The forces on the ligaments were neglected.

Using this model, Koolstra and van Eijden performed simulations of *jaw-closing* as a result of isotonic forces (10 N) generated by various pairs of masticatory muscles (52). This model demonstrated that the normally observed swing-slide condylar movement along the articular eminence can be generated by various masticatory muscles. However, different parts of the masseter and the medial pterygoid muscle seemed to be most suitable for this motion (52).

The mastication forces are controlled by the central nervous system, and they are modulated by the physical limitations of the sarcomeres, the force generating units of the muscles (53). The force generated by the sarcomeres depends on the length and contraction velocity of the sarcomeres. In addition, a passive elastic force is generated depending on the amount of stretching. These are determined empirically and expressed as the force-length (F_L), force-velocity (F_v), and stretching (F_p) factors (53). The magnitude of the instantaneous muscle force $F(t)$ can then be formulated as (54):

$$F(t) = F_{\text{max}}[A(t)F_L(t)F_v(t) + F_p(p)] \quad (12)$$

where $A(t)$ is the activation level of the muscle.

The jaw-opening and closing simulations were performed by Koolstra and van Eijens using equation 12 in their dynamic model (54). In this work, they included the jaw-opening muscles, in addition to the closing muscles. The activation level $A(t)$ was kept constant, and all muscles were activated simultaneously. It was found that the level of activation of the temporalis muscle parts was critical in jaw-closing movements. The amount of jaw opening was limited by the passive forces of the jaw-closing muscles. However, the amount of jaw closing was not significantly influenced by the passive forces of the jaw opening (54). The TMJ remained loaded throughout the jaw movements. They also concluded that the average moments generated by the jaw-closing muscles, with respect to the center of mass of the mandible, are responsible for stable operation of the TMJ (52,54).

The lateral deviation of the jaw from the closed position was also modeled by Koolstra and van Eijens (24). They found that laterodeviations that conform to the naturally observed ones could be generated by unilateral muscle contractions. Their analysis concluded that movements of the jaw predominantly depend on the orientation of the contributing muscles with respect to the center of mass, and not on the TMJ ligaments or passive elastic muscle properties (24).

Deformations of the Mandible

Forces applied to the mandible cause internal stress response within the mandible. Highest masticatory loads are experienced when the maxillary and mandibular teeth are in contact. During typical chewing, such high load contacts last on the order of 0.1 s for chewing strokes and longer for swallowing. At the end of the day, 15–30 min of high loads are experienced by the masticatory system. Persons with parafunctional habits, such as bruxism, may experience longer durations of high levels of loading (4).

The stress and strain distribution in the mandible change in response to the loss of teeth (30), mandibular reconstruction (56,57) TMJ reconstruction, presence of dental implants (58), or prosthesis (59); consequently, the mandible changes its geometry and material property distribution (30). Analysis of the internal stresses in the mandible has, therefore, a potential to aid treatment. The deformations and internal stresses of the masticatory system could be evaluated by modeling the mandible as a curved beam (60), using the photoelastic method (4) or by strain measurements. The nonuniformity of the cross-sectional area and that of the trabeculae of the mandibular bone, and the presence of the teeth, prevent the beam theory from being an accurate tool of analysis. On the other hand, the photoelastic method provides a good visualization of the internal stress distribution (4). But the trabecular architecture is generally not included in the photoelastic models. The strain measurements *in vivo* or *in vitro* can only be performed on the surface of the bone.

The finite element method offers a suitable alternative to model the geometric and material variations of the mandible. Detailed analysis using FEA comes at a high computational expense. However, combined with experimental strain measurements and the photoelastic method,

for verification purposes, the FEA can be a useful tool; and, it has been used to analyze the internal stress distribution in the mandible (30,61–63). The FEA of a partially edentulous mandible, by Hart et al., showed that the strain distribution in the mandible is extremely complex (30). This work showed that an asymmetrically edentulous mandible could experience different condylar reactions even in bilateral loading. The anterior portion of the ramus, the coronoid process, and the attachment locations of the muscles experience relatively high tensile stresses under various loading conditions. The condyle experiences relatively high compressive stresses. In general, these results were found to be in agreement with photoelastic stress analysis (4,30).

BIOMECHANICAL PROPERTIES OF TOOTH REPLACEMENT MODALITIES

The treatment objectives of prosthodontic services include maintaining and enhancing quality of life by providing the function, comfort, and aesthetics that is compromised or lost due to oral disease. Amount of damage to oral structures and properties of materials are two important factors that dictate treatment techniques. *Restoration* is a broad term applied to any material or prosthesis that restores or replaces lost tooth structure, teeth, or oral tissues (17). Patient desires and expectations aside, depending on the extent of damage or loss, treatment options may be classified as (1) intracoronal restorations, (2) extracoronal restorations, (3) fixed partial dentures, and (4) removable partial dentures. Each of these treatment modalities has several different choices that can be used by the clinician. In addition to these treatment options, dental implants offer an anchoring function for single tooth replacement or support for various denture scenarios.

To meet functional, biological, and esthetic requirements, a “fixed” restoration (filling, crown, bridge) must remain firmly attached to the tooth. In general, *retention form* indicates the feature of a tooth preparation that resists dislodgement of a crown in a vertical direction or along the path of placement, whereas *resistance form* is the feature to resist dislodgement in all other directions. Tooth preparation also involves removal of sufficient tooth material (reduction) so that the “replacement material” can have enough bulk for structural durability.

Intracoronal Restorations

Absence of tooth structure due to caries (decayed tissue), trauma, or developmental defects is the main indication for *intracoronal* restorations. Before placement of such a restoration, a damaged tooth is “prepared” by the dentist to a certain geometric form to remove all caries, to protect the remaining tooth structure, and to minimize the chances of dislodgement of the restorative material during function (i.e., eating and swallowing) and parafunction (i.e., tooth grinding and clenching).

Metal direct filling materials are gold foil and silver amalgam, which are supplied in many different compositions. *Esthetic direct filling* materials, composite resins, are BISGMA acrylic resins filled with inorganic materials.

The preparation of the teeth for metal direct filling restorations is different than for composite resins. Mechanical locks and undercuts provide retention and resistance to gold foil and amalgam restorations. Preparations are usually extended into the dentin layer to have at least 1–1.5 mm thickness of the material and to take advantage of the elasticity difference between dentin and enamel. Composite restorations, on the other hand, require less invasive preparations. Analyses of preparation designs for various clinical scenarios via photoelastic or finite element studies, as well as bench tests on extracted teeth, have been conducted to provide insight to the optimum preparation form for a specific material to protect the tooth from further damage. Strength (15), wear characteristics (64,65), and reaction to temperature changes are examples of important factors that have been investigated.

Inlays are alternatives to direct filling materials. Preparations for this type of restoration require occlusally diverging walls because these restorations are made indirect and require a path of insertion. Gold alloys, composite resins, or porcelain can be used as inlay materials. This type of restoration is indicated in non-stress-bearing, two-surface preparations (mesio-occlusal, disto-occlusal). Their use in stress-bearing, three-surface (mesio-occluso-distal) situations has been associated with a “wedging” effect, risking cusp fractures.

Increased tooth loss or extensive caries necessitates modification of preparations from ideal forms. At this point, the clinician has to either employ auxiliary retention methods for direct fillings via retentive pins or choose cuspal coverage with indirect restorations (onlay, crown).

Extracoronary Restorations

As the damage to a tooth gets more extensive, an extracoronary type restoration (crown and dowel) is indicated. Major factors that affect retention of a crown include, but are not limited to, the degree of taper of the walls of the prepared tooth, total surface area of the luting cement, area of the cement under shear forces, and roughness of the tooth surface. The resistance form of a crown includes consideration of length of the preparation, tooth width, and taper. Crowns can be made of gold alloys, porcelain, or a combination of both (porcelain fused to metal).

Very often an endodontically treated tooth has a very limited amount of coronal dentin. This is in part due to the required access opening to perform the root canal, and in part to previous tooth loss, i.e., existing fillings. In such situations, foundation restorations are indicated. Frequently, these restorations require utilization of *dowels* (posts). A dowel is fitted into a prepared root canal of a natural tooth. When combined with an artificial crown or core, it provides retention and resistance for the restoration. Many different geometric configurations (e.g., custom cast, prefabricated tapered, parallel sided, threaded), materials of different strength and fatigue characteristics, and corrosion resistance properties (gold alloy, stainless steel, titanium alloy, carbon, ceramic, fiber) have been manufactured for this purpose. The effects of post length and diameter on retention have been widely studied.

Fixed Partial Dentures (Bridges)

Fixed partial dentures (FPDs) can be used to replace missing teeth. Whether these FPDs are supported by natural teeth or dental implants presents different biomechanical issues. On natural teeth, variables such as pericemental support of the abutments (teeth that support the FPD), crown-to-root ratio, configuration of the roots (i.e., single vs. multiple and converging vs. diverging) are among the biomechanical factors to be considered. Available space could become an issue, as the span of the restoration, which is dictated by dimensions of the missing tooth/teeth, is inversely proportional to the strength of the restoration. Similarly, height of the abutment teeth influence height of the connector (part of the bridge where missing tooth replacement, pontic(s), connect(s) to the retainers) and its strength. Hence, material selection is very important. Difference in mobility of abutment teeth may affect seating of the restoration. Another clinical scenario involves the presence of an intermediate abutment when a natural tooth located between terminal abutments serves to support a fixed prosthesis. Under occlusal loads, this “pier” abutment acts as an undesired fulcrum, and the abutment with the least retention fails. To overcome this issue, segmental bridges have been used.

Horizontal components of bite forces are transmitted in part from posterior to anterior directions, because of proximal contact of teeth within the same arch. These forces are well tolerated by the bridge system. However, forces that create movement toward buccal or lingual side (cheek or tongue side) may be detrimental.

When implants are used to support and retain a fixed prosthesis, the difference in mobility of implants is less of a concern, because successful osseointegration is assumed and mobility is more likely related to deformation of bone. However, passive fit of the restoration is a difficult task to achieve due to limitations of materials and techniques. Misfit of restorations is reported to cause more maintenance problems, rather than jeopardizing osseointegration. Unlike natural teeth-supported FPDs, the dentist has the control over the number, size, and location of implant placement when an implant-supported FPD is planned (Fig. 9).

Removable Dentures

Removable dentures are used in complete and partial edentulism. Indications for removable partial dentures include, but are not limited to, situations where the edentulous span is very long, periodontal support is reduced, posterior abutments are absent, and tissue loss is excessive. Typically, a removable partial denture has a cast metal framework, which extends from one side of the dental arch to the other, regardless of the location of the missing teeth. This “cross arch stabilization” by the rigid framework helps reduce detrimental horizontal forces on the abutment teeth.

Components of a removable partial denture (RPD) framework include major connectors, rests, direct and indirect retainers, and minor connectors. One of the most important aspects of removable partial denture fabrication involves design of the framework to achieve retention,

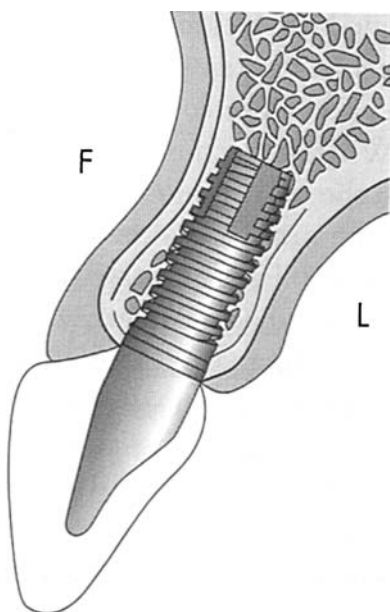


Figure 9. Schematic representation of an osseointegrated, endosseous implant with the abutment and the restorative crown. (From “Single tooth implants,” by Misch CE, in *Implant Dentistry*, 2nd ed., Mosby, 1999).

support, stability, comfort, function, and aesthetics. Removable partial dentures are subject to a combination of forces arising from three different fulcrums on horizontal, sagittal, and vertical planes. As described for other treatment modalities, planning is aimed at directing forces along the long axes of the teeth as much as possible. Factors that influence the amount of stress transmitted to abutment teeth include length of the edentulous span, quality of the ridge support, clasp design flexibility and material, and occlusal harmony.

Tooth loss in an arch varies significantly in number and location; therefore, removable partial dentures are often designed based on available support (tooth supported, or tooth–tissue supported). When the tooth loss pattern allows a tooth-supported RPD, design issues are less complex. However, posterior “free end” scenarios present specific problems because the denture will be supported both by teeth and tissues. Under function, healthy teeth may be displaced as much as 0.2 mm, whereas tissue may be displaced 1.0 mm or more. Thus, prevention of moments on the terminal abutment teeth (teeth that are next to the edentulous space) has been a major concern for tooth–tissue-supported prosthesis. To minimize moments, various clasp assembly designs, resilient precision attachments, and “stress breaking” mechanisms have been proposed.

Recently dental implants have been used to assist RPDs in retention and support.

Complete dentures are indicated in total edentulism. For patients that cannot adapt to complete dentures, use of dental implants is indicated. Retentive mechanisms for implant-assisted dentures vary from individual attachments to bar systems, joining multiple implants. There are many variations to individual attachments as well as to

bar systems. These result in different levels of retention, support and freedom, and lack of movement in a specific plane.

Dental Implants

A dental implant is a prosthetic device of alloplastic material implanted into the oral tissues beneath the mucosa and periosteal tissues and into the jaw bone to support a fixed or removable prosthesis. A dental implant system serves as the anchor for the prosthetic reconstruction of missing teeth by supporting a fixed or removable prosthesis. The system mainly consists of an implant and an abutment. A *prosthetic attachment* is typically fixed on the *abutment* by one of the following methods: cementation, use of an occlusal screw, or a socket arrangement that allows retention of a removable prosthesis.

The *implant* is the component implanted into the bone tissue and serves the function of the root. Upon surgical placement of the implant, a healing period of 2–6 months is allowed during which osseointegration takes place. *From the patient's point of view*, “a fixture is considered osseointegrated if it provides a stable and apparently immobile support of a prosthesis under functional loads, without pain, inflammation or loosening” (66). During osseointegration, new bone forms in contact with the implant and a direct structural and functional connection is established, without initiating an immune response (rejection). Mineralized tissue is found to be in contact with the implant surface over most of the surface within nanometers (66). Relative movement (micromotion) between the implant and the bone at the time of placement is more likely to favor development of a fibroosseous interface (28). The presence of the implant modifies the mechanical environment in its surrounding, by altering the normal physiology, distribution of the fluids, and force transmission (28). Nevertheless, dental implants have high long-term success rates (67,68) due to careful bioengineering of the choice of materials (69), surface topography and coatings, overall size and shape of the implant body, and thread shape.

The *abutment* is the component that supports and/or retains the prosthesis (70). The abutment is secured to the implant with a mechanical attachment method, and ideally, it should stay fixed with respect to the implant throughout the life of the implant. Currently, three methods are used to attach an abutment onto an implant. In the most common mechanical attachment method, an abutment retaining-screw is used to fix the abutment with respect to the implant (70). The mechanics of this type of screwattachment are analyzed by classic methods (71) and the FEA (72). Another approach is to use a screw with a relatively large tapered end (73). Finally, in some implant systems, a tapered interference fit between the abutment and the implant is also used to provide the connection (74,75).

From a bioengineering perspective, an important issue is to design the implant with a geometry that will minimize the peak bone stress caused by standard loading (76). The complex geometry of the implants prevents the use of closed-form solutions in stress analysis, where simple formulas relate the effect of external loads to internal stresses

and deformation. The FE method has been applied to the dental implant field to predict stress distribution patterns in the implant–bone interface not only by comparison of various root-form implant designs (76–80), but also by modeling various clinical scenarios (1,58,81,82) and prosthesis designs (32,83–85). This method offers the advantage of solving complex structural problems by dividing them into smaller and simpler interrelated sections by using mathematical techniques (41).

FEAs, which investigate the relation between implant design and stress distribution, have addressed the overall shape and size of the implant body, implant neck geometry, and thread geometry for threaded implants. Rieger et al. showed that a tapered design made of a material with high elastic modulus would be most suitable to serve as a free-standing implant (78). Holmgren et al. suggested considering application of oblique load to FEAs, indicating that these were more realistic occlusion directions capable of causing the highest localized stress in the cortical bone (77). These authors found the stepped implant design to exhibit a more even stress pattern than a straight cylindrical design.

Threaded implants exhibit geometric variations in terms of thread pitch, shape, and depth. Threads are used to increase the surface area of the implant (86). Use of different thread configurations for different bone qualities has been proposed as thread geometry may play an important role in the type of force transmitted (70,87–89). Chun et al. showed that the maximum stress in compact bone is higher for the plateau design compared with the triangular or square designs and their variations. According to these authors, screw pitch had a significant impact on the stress distribution (90).

The transosteal region of the implant body has been defined as the “crest module” (86). For most systems, this neck portion of the implant is smooth. Different designs include parallel, converging, and diverging sides. One particular implant investigated by Hansson using the FE method included both taper and retention elements up to the crest of the implant and was found to have much lower interfacial shear stresses compared with a smooth neck design (76).

BIBLIOGRAPHY

1. Fung YC. *Biomechanics, Mechanical Properties of Living Tissues*. New York: Springer-Verlag; 1981.
2. Brunski JB. *Biomechanics of tooth and jaw*. In: Webster JG, editor. *Encyclopedia of Medical Devices and Instrumentation*. New York: Wiley; 1988. 2776–2788.
3. Koolstra JH. Number crunching with the human masticatory system. *J Dent Res* 2003;82:672–676.
4. Caputo AA, Standlee JP. *Biomechanics in Clinical Dentistry*. Chicago: Quintessence Books; 1987.
5. Okeson JP. *Management of Temporomandibular Disorders and Occlusion*. St. Louis: Mosby; 2003.
6. Koolstra JH. Dynamics of the human masticatory system. *Crit Rev Oral Biol Med* 2002;13:366–376.
7. Mow VC, Ateshian GA, Spiker RL. *Biomechanics of diarthroidal joints: A review of twenty years of progress*. *J Biomechan Eng* 1993;115:460–467.
8. Mow VC, Mak AF. *Lubrication of diarthroidal joints*. In: Skalak R, Chien S, editors. *Handbook of Bioengineering*. New York: McGraw-Hill; 1988. 1–34.
9. Osborn JW. The disk of the human temporomandibular joint: Design, function and failure. *J Oral Rehabil* 1985;12:279–293.
10. van Eijden TMGJ. *Biomechanics of the mandible*. *Crit Rev Oral Biol Med* 2000;11:123–136.
11. Woo SL-Y, Mow VC, Lai WM. *Biomechanical properties of articular cartilage*. In: Skalak R, Chien S, editors. *Handbook of Bioengineering*. New York: McGraw-Hill; 1988;4:1–44.
12. Anonymous. *Dorland's Illustrated Medical Dictionary*. 25th ed. Philadelphia: Saunders; 1974.
13. Toparli M, Gökay N, Aksoy T. Analysis of a restored maxillary second premolar tooth by using three-dimensional finite element method. *J Oral Rehabil* 2001;28:157–164.
14. Rees JS, Jacobsen PH. The effect of cuspal flexure on a buccal Class V restoration: A finite element study. *J Dentistry* 1998; 26:361–367.
15. Arola D, Galles LA, Sarubin MF. A comparison of the mechanical behavior of posterior teeth with amalgam and composite MOD restorations. *J Dentistry* 2001;29:63–73.
16. Imanishi A, Nakamura T, Ohyama T, Nakamura T. 3-D finite element analysis of all-ceramic posterior crowns. *J Oral Rehabil* 2003;30:818–822.
17. Anonymous. *The glossary of prosthodontic terms*. *J Prosthet Dent*. 1999;81:39–110.
18. Tanaka E, Sasaki A, Tahmina K, Yamaguchi K, Mori Y, Tanne K. Mechanical properties of human articular disk and its influence on temporomandibular joint loading studied with finite element method. *J Oral Rehabil* 2001;28:273–279.
19. Beek M, Koolstra JH, van Eijden TMGJ. Human temporomandibular joint disk cartilage as a poroelastic material. *Clin Biomech* 2002;18:69–76.
20. Williams JA. *Engineering Tribology*. Oxford: Oxford University Press; 2000.
21. DeVocht JW, Goel VK, Zeitler DH, Lew D. A study of the control of disk movement within the temporomandibular joint using the finite element technique. *J Oral Maxillofacial Surg* 1996;54:1431–1437.
22. Akca K. H. Iplikcioglu H. Finite element stress analysis of the influence of staggered versus straight placement of dental implants. *Int J Oral Maxillofacial Implants* 2001;16:722–730.
23. Tanaka E, Rodrigo dP, Tanaka M, Kawaguchi A, Shibazaki T, Tanne K. Stress analysis in the TMJ during jaw opening by use of a three-dimensional finite element model based on magnetic resonance images. *Int J Oral Maxillofacial Surg* 2001;30:421–430.
24. Koolstra JH, van Eijden TMGJ. Three dimensional dynamical capabilities of the human masticatory muscles. *J Biomech* 1999;32:145–152.
25. Faulkner MG, Hatcher DC, Hay H. A three-dimensional investigation of temporomandibular joint loading. *J Biomech* 1987;20:997–1002.
26. Malvern LE. *Introduction to the Mechanics of a Continuous Medium*. Englewood Cliffs, NJ: Prentice-Hall; 1969.
27. van Rietbergen B, Huiskes R. *Elastic constants of cancellous bone*. In: Cowin SC, editor. *Bone Mechanics Handbook*. 2001.
28. Martin RB, Burr DB, Sharkey NA. *Skeletal Tissue Mechanics*. New York: Springer; 1998.
29. Rigsby DF, Bidez MW, Misch CE. *Bone response to mechanical loads*. In: Misch CE, editor. *Contemporary Implant Dentistry*. 2nd ed. St. Louis: Mosby; 1999. 317–328.
30. Hart RT, Henebber V, Thongpreda N, van Buskirk WC, Anderson RC. Modeling the biomechanics of the mandible: A three dimensional finite element study. *J Biomech* 1992;25:261–286.
31. Baimonte T, Abbate MF, Pizzarello F, Lozada J, James R. The experimental verification of the efficacy of finite element modeling to dental implant systems. *J Oral Implantol* 1996;12:104–110.

32. Papavasiliou G, Kamposiora P, Bayne SC, Felton DA. Three-dimensional finite element analysis of stress-distribution around single tooth implants as a function of bony support, prosthesis type, and loading during function. *J Prosthet Dent* 1996;76:633–640.
33. Sakaguchi RL, Borgersen SE. Nonlinear finite element contact analysis of dental implant components. *Int J Oral Maxillo Facial Implant* 1993;8:655–661.
34. Teixeira ER, Sato Y, Akagawa Y, Shindo N. A comparative evaluation of mandibular finite element models with different lengths and elements for implant biomechanics. *J Oral Rehab* 1998;25:299–303.
35. Van Oosterwyck H, Duyck J, Vander Sloten J, Van Der Perre G, De Cooman M, Lievens S, Puers R, Naert I. The influence of bone mechanical properties and implant fixation upon bone loading around oral implants. *Clin Oral Impl Res* 1998;9:407–418.
36. Mow V, Kuei SC, Lai WM, Armstrong CG. Biphasic creep and stress relaxation of articular cartilage in compression: Theory and experiments. *J Biomech Eng* 1980;102:73–84.
37. Hu K, Radhakrishnan P, Patel RV, Mao JJ. Nanomechanical and topographic properties of the articular fibrocartilage of the rabbit mandibular condyle. *J Struct Biol* 2001;136:281–288.
38. Linn FC. Lubrication of animal joints. I, The arthrotriometer. *J Bone Joint Surg* 1967;49A:1079.
39. Chen J, Akyuz U, Xu L, Pidaparti RMV. Stress analysis of the human temporomandibular joint. *Med Eng Phys* 1998;20:565–572.
40. Chin LPY, Aker FD, Zarrinnia K. The viscoelastic properties of the human temporomandibular joint disk. *J Oral Maxillofacial Surg* 1996;54:315–318.
41. Belytschko T, Liu WK, Moran B. *Nonlinear Finite Elements for Continua and Structures*. Chichester: Wiley; 2004.
42. Chen J, Xu L. A finite element analysis of the human temporomandibular joint. *J Biomech Eng* 1994;116:401–407.
43. Tanaka E, Kazuo T, Sakuda M. A three-dimensional finite element model of the mandible including the TMJ and its application to stress analysis in the TMJ during clenching. *Med Eng Phys* 1994;16:316–322.
44. Hu K, Qiguo R, Fang J, Mao JJ. Effects of condoylar fibrocartilage on the biomechanical loading of the human temporomandibular joint in a three-dimensional, non-linear finite element model. *Med Eng Phys* 2003;25:107–113.
45. Viidik A. Properties of tendons and ligaments. In: Skalak R, Chien S, editors. *Handbook of Bioengineering*. New York: McGraw-Hill; 1988. 1–19.
46. Smith DM, McLachlan KR, McCall WD. A numerical model of temporomandibular joint loading. *J Dent Res* 1986;65:1046–1052.
47. Osborn JW, Baragar FA. Predicted pattern of human muscle activity during clenching derived from a computer assisted model: Symmetric vertical bite forces. *J Biomech* 1985;18:599–612.
48. Strang G. *Introduction to Linear Algebra*. Cambridge, MA: Wellesley-Cambridge Press; 1998.
49. Hatcher DC, Faulkner MG, Hay H. Development of mechanical and mathematic models to study temporomandibular joint loading. *J Prosthetic Dentistry* 1986;55:377–384.
50. Koolstra JH, van Eijden TMGJ, Weijs WA, Naeije M. A three-dimensional mathematical model of the human masticatory system predicting maximum possible bite forces. *J Biomech* 1988;21:563–576.
51. Koolstra JH, van Eijden TMGJ. Application and validation of a three-dimensional mathematical model of the human masticatory system in vivo. *J Biomech* 1992;25:175–187.
52. Koolstra JH, van Eijden TMGJ. Biomechanical analysis of jaw-closing movements. *J Dental Res* 1995;74:1564–1570.
53. Koolstra JH, van Eijden. Dynamics of the human masticatory muscles during a jaw open-close movement. *J Biomech* 1997;30:883–889.
54. Koolstra JH, van Eijden TMGJ. The jaw open-close movements predicted by biomechanical modeling. *J Biomech* 1997;30:943–950.
55. Tongue BH, Shepard SD. *Dynamics*. New York: Wiley;
56. Cox T, Kohn MW, Impelluso T. Computerized analysis of resorbable polymer plates and screws for the rigid fixation of mandibular angle fractures. *J Oral Maxillofacial Surg* 2003;61:481–487.
57. Fernandez JR, Gallas M, Burguera M, Viano JM. A three-dimensional numerical simulation of mandible fracture reduction with screwed miniplates. *J Biomech* 2003;36:329–337.
58. Ishigaki S, Nakano T, Yamada S, Nakamura T, Takashima F. Biomechanical stress in bone surrounding an implant under simulated chewing. *Clin Oral Impl Res* 2003;14:97–102.
59. Zarone F, Apicella A, Nicolais L, Aversa R, Sorrentino R. Mandibular flexure and stress build-up in mandibular full-arch fixed prostheses supported by osseointegrated implants. *Clin Oral Impl Res* 2003;14:103–114.
60. Hylander WL. Stress and strain in the mandibular symphysis of primates: A test of competeing hypothesis. *Am J Phys Anthro* 1984;64:1–46.
61. T.Koriioth WP, Hannam AG. Mandibular forces during simulated tooth clenching. *J Orofacial Pain* 1994;8:178–189.
62. Hirayabashi M, Motoyoshi M, Ishimaru T, Kasai K, Namura S. Stress in mandibular cortical bone during mastication: Biomechanical considerations using a three dimensional finite element method. *J Oral Sci* 2002;44 1–6.
63. Vollmer D, Meyer U, Joos U, Vegh A, Piffko J. Experimental and finite element study of a human mandible. *J Cranio-Maxillofacial Surg* 2000;28:91–96.
64. Wassell RW, McCabe JF, Walls AW. A two-body frictional wear test. *J Dent Res* 1994;73:1546–1553.
65. Wassell RW, McCabe JF, Walls AW. Wear characteristics in a two-body wear test. *Dent Mater* 1994;10 269–274.
66. Skalak R, Branemark P-I. Definition of osseointegration. In: Branemark P-I, Rydevik BL, Skalak R, editors. *Osseointegration in Skelatal Reconstruction and Joint Replacement*. Chicago, IL: Quintessence Books; 1994.
67. Behneke A, Behneke N, d'Hoeedt B. The longitudinal clinical effectiveness of ITI solid-screw implants in partially edentulous patients: A 5-year follow-up report. *Int J Oral Maxillofacial Implants* 2002;15:633–645.
68. Haas R, Polak C, Furhauser R, Mailath-Pokorny G, Dortbudak O, Watzek G. A long-term follow-up of 76 Branemark single-tooth implants. *Clin Oral Implants Res* 2002;13 38–43.
69. Lemons JE, Dietsch-Misch F. Biomaterials for dental implants. In: Misch CE, editor. *Contemporary Implant Dentistry*. 2nd ed., St. Louis: Mosby; 1999. 271–302.
70. Misch CE, Hoar J, Beck G, Hazen R, Misch CM. A bone quality-based implant system: A preliminary report of stage I & stage II. *Implant Dent* 1998;7:35–42.
71. Shigley JE, Mischke CR. *Mechanical Engineering Design*. 5th ed. Cambridge, MA: McGraw Hill; 1989. 450–457.
72. Lang LA, Kang B, Wang RF, Lang BR. Finite element analysis to determine implant preload. *J Prosthet Dent* 90:539–546. 2003.
73. Bozkaya D, Müftü S. Mechanics of the taper integrated screwed-In (TIS) abutments used in dental implants. *J Biomech* 2005;38:87–97.
74. Bozkaya D, Müftü S. Mechanics of the tapered interference fit in dental implants. *J Biomech* 2003;36:1649–1658.

75. Bozkaya D, Müftü S. Efficiency considerations for the purely tapered interference fit (TIF) abutments used in dental implants. *J Biomech Eng* 2004;126:393–401.
76. Hansson S. The implant neck: smooth or provided with retention elements: A biomechanical approach. *Clin Oral Implants Res* 1999;10:394–405.
77. Holmgren EP, Seckinger RJ, Kilgren LM, Mante F. Evaluating parameters of osseointegrated dental implants using finite element analysis—a two-dimensional comparative study examining the effects of implant diameter, implant shape, and load direction. *J Oral Implantol* 1998;24:80–88.
78. Rieger MR. Finite element stress analysis of root-form implants. *J Oral Implantol* 1988; 14:472–484.
79. Rieger MR, Adams WK, Kinzel GL. A finite element survey of eleven endosseous implants. *J Prosthet Dent* 1990;63:457–465.
80. Siegele D, Soltesz U. Numerical investigations of the influence of implant shape on stress distribution in the jaw bone. *Int J Oral Maxillofacial Implants* 1989;4:333–340.
81. Pierrisnard L, Hure G, Barquins M, Chappard D. Two dental implants designed for immediate loading: A finite element analysis. *Int J Oral Maxillofacial Implants* 2002;17:353–362.
82. Van Oosterwyck H, Duyck J, Vander Sloten J, Van Der Perre G, Naert I. Peri-implant bone tissue strains in cases of dehiscence: A finite element study. *Clin Oral Implants Res* 2002; 13:327–333.
83. Papavasiliou G, Tripodakis AP, Kamposiora P, Strub JR, Bayne S. Finite element analysis of ceramic abutment-restoration combinations for osseointegrated implants. *Int J Prosthodont* 1996;9:254–260.
84. Stegaroiu R, Kusakari H, Nishiyama S, Miyakawa O. Influence of prosthesis material on stress distribution in bone and implant: A 3-dimensional finite element analysis. *Int J Oral Maxillofacial Implants* 1998;13:781–790.
85. Stegaroiu R, Sato T, Kusakari H, Miyakawa O. Influence of restoration type on stress distribution in bone around implants: A three-dimensional finite element analysis. *Int J Oral Maxillofacial Implants* 1998;13:82–90.
86. Misch CE, Waren Bidez M. A scientific rationale for dental implant design. In: Misch CE, editor. *Contemporary Implant Dentistry*. St. Louis: Mosby; 1999. 329–343.
87. Misch CE, Waren Bidez M, Sharawy M. A bioengineered implant for a predetermined bone cellular response to loading forces. A literature review and case report. *J Periodontol* 2001; 72:1276–1286.
88. Misch CE, Dietsch-Misch F, Hoar J, Beck G, Hazen R. A bone quality-based implant system: first year of prosthetic loading. *J Oral Implantol* 1999;25:185–197.
89. Misch CE. Implant design considerations for the posterior regions of the mouth. *Implant Dent* 1999;8:376–386.
90. Chun HJ, Cheong SY, Han JH, Heo SJ, Chung JP. Evaluation of design parameters of osseointegrated dental implants using finite element analysis. *J Oral Rehabil* 2002;29:565–574.

See also BONE AND TEETH, PROPERTIES OF; BIOMATERIALS FOR DENTISTRY; JOINTS, BIOMECHANICS OF.

TOTAL ARTIFICIAL HEART. See HEART, ARTIFICIAL.

TOTAL JOINT PROSTHESES. See MATERIALS AND DESIGN FOR ORTHOPEDIC DEVICES.

TOTAL PARENTERAL NUTRITION. See NUTRITION, PARENTERAL.

TRACER KINETICS

EDWARD V. R. DI BELLA
University of Utah
Salt Lake City, Utah

INTRODUCTION

Tracer kinetics refers to the changing distribution of a tracer that provides information regarding a biological system. The tracer may be introduced by injection into the bloodstream, or inhaled, or in some cases be native. By definition, the tracer is present in relatively small (trace) amounts so that it has little or no effect on the system being studied.

A large variety of tracers, along with methods for their measurement, are employed daily in numerous medical applications. Imaging function with tracers can provide complementary or more relevant information than anatomical imaging. For example, positron emission tomography (PET) and single photon emission computed tomography (SPECT) imaging can track radiolabeled tracers to detect and stage cancer, to assess neural function, to follow response to therapy, and to detect heart disease. Magnetic resonance imaging (MRI) with paramagnetic contrast acting as a tracer is a rapidly growing area, as is ultrasound with tracer-like microbubbles. Over 1 million patients are imaged annually in the United States for oncology applications. Over 7 million/year have a cardiac SPECT scan, mostly for the diagnosis of coronary artery disease.

How do these tracers work? Two brief examples will help to illustrate the process. The first example is of tracer applications that are suitable for analysis at equilibrium. That is, the kinetics portion of Tracer Kinetics is not used directly, but rather serves only to bring the tracer to some equilibrium state. The second example is of a radiolabeled glucose analogue, which is most often analyzed as if it were at equilibrium, but offers more information when its changing distribution over time is considered as well. This will provide a feel for the field of tracer kinetics and for the controversy sometimes surrounding the use of tracer kinetics with quantitative modeling methods, compared to visualization of static images of tracer distribution.

“Static” Tracers

By far, the bulk of clinical radionuclide studies use static images acquired while the tracer distribution is changing slowly or is fixed. Microsphere-type tracers are the simplest case of this type of tracer. These tracers provide a static distribution after their first pass and may be, for example, ^{99m}Tc-labeled albumin, or in studies with animals that will be sacrificed, 10–15 μm diameter carbonized microspheres tagged with radioactive or fluorescent labels. Such tracers lodge in capillaries in proportion to flow. The microspheres must be infused into the left atrium or left ventricle or will all be trapped in the lungs. Absolute flow values in milliliters per minute per gram (mL/min/g) can be computed using these tracers. However, the amount of tracer in the region of interest must be known. For animal studies with microspheres, this is done by excising the

regions of interest after sacrifice. The small tissue regions are counted in a type of well counter if radiolabeled microspheres were used, or digested and read in a fluorimeter if fluorescent microspheres were used.

Absolute flow calculation with microspheres also requires knowledge of the flow to a “virtual” organ. This virtual organ can be at the aortic valve at the output of the left ventricle, in which case it is known that the full amount of tracer enters the virtual organ at a flow rate equal to the cardiac output. Thus if cardiac output is known and the number of microspheres is known, then flow to any region of interest can be computed as cardiac output x (number of microspheres in region)/(number of microspheres injected). The units will be milliliters per minute (mL/min), which can be made into mL/min/g after dividing by the weight of the region of interest.

The virtual organ could instead, for example, be in the femoral artery if a catheter was inserted into the artery and a blood sample withdrawn at a known flow rate during the first pass of the microsphere injection. This knowledge of number of microspheres in the virtual organ given a known flow rate allows computation of the flow rates in other regions.

Other tracers that lodge in position in proportion to flow can provide relative flow measures though it is difficult to provide absolute flow measures. One example of a widely used study is ^{99m}Tc -sestamibi SPECT. Sestamibi is injected after exercise or pharmacological vasodilation, and rapidly equilibrates with deposition in proportion to flow. The distribution then changes very slowly, as it is “stuck” in the myocardium. The timing of the SPECT scan is then not critical. The half-life of radioactive decay of ^{99m}Tc is ~ 6 h, so decay is not a central issue in the imaging. These myocardial perfusion SPECT scans are done at exercise or with a vasodilator in order to be sensitive to stenosed coronary vessels. An excellent reference for the use of tracers in Nuclear Cardiology, including a chapter on tracer kinetics, is found in Ref. 1.

Dynamic Tracers

In general, a thorough understanding of the kinetics of the particular tracer of interest leads to the best method for processing the measurements of the tracer. Consider fluoro-deoxyglucose (^{18}F FDG); ^{18}F FDG is a positron emitting tracer of glucose. However, ^{18}F FDG is trapped at a certain stage in the glycolysis cycle (Fig. 1). As with glucose, hexokinase catalyzes its conversion from ^{18}F FDG to phosphorylated ^{18}F FDG, ^{18}F FDG-6- PO_4 . However, ^{18}F FDG-6- PO_4 does not continue along the glycolysis pathway as does glucose-6- PO_4 , and very little of it leaks back into the interstitial space and vasculature. The changes in ^{18}F FDG concentration over time in an area can be represented with a physiological three compartment model as seen in Fig. 1. Compartment models are defined in detail in the next section. The use of dynamic (time series) images of ^{18}F FDG acquired with PET requires the application of tracer kinetic principles to provide semiquantitative or quantitative absolute measurements of regional metabolic rate of glucose usage. A mathematical model that can predict or fit the PET data and provides estimates of

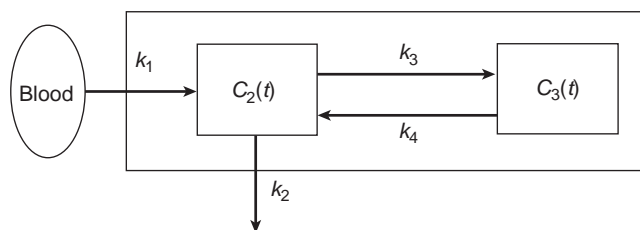


Figure 1. Three-compartment model. For FDG, k_1 represents uptake into a free compartment outside the vasculature. The parameter k_2 is the washout of free FDG back into the vasculature; k_3 represents the rate constant of phosphorylation of FDG to FDG-6- PO_4 ; and k_4 represents the rate constant of dephosphorylation and is relatively small. Note the concentration measured in the tissue, $C_t(t)$ in the text, is $C_t(t) = C_2(t) + C_3(t)$ (illustrated by the largest box).

meaningful physiological parameters is needed. Such models require the arterial input function (AIF), which is the time–activity function the tissue of interest sees as supplying itself.

Yet, in typical clinical ^{18}F FDG PET scans, one tomographic volume image is acquired over ~ 5 – 30 min, at a time point when the distribution of FDG is changing only slowly. Thus the dynamic data is not used. The images are simply analyzed visually by an experienced reader. Thus at least to date, the complexities of acquiring the data dynamically and measuring an accurate input function and performing the required processing seemingly outweigh the gain of having a quantitative measure of the metabolic rate of glucose usage. This is not the case with some other dynamic tracers such as radiolabeled water, where visual analysis is not useful and kinetic analysis is nearly essential.

Nomenclature

The study of tracer kinetics began prior to 1950, and, due to its myriad applications in a breadth of fields, has had a number of different terms used to describe the same things. Some unified nomenclature was proposed for use with tracer kinetics for biomedical research in 1990 (2). The paper includes a 170 term glossary, but has not prevented the field from having a rich array of discrepant notation and terminology.

Tracer kinetics have been rapidly growing in use with MRI applications. Sufficient research has been published using dynamic contrast enhanced (DCE) MRI with confusing terminology to motivate a paper by MRI researchers regarding terminology for the popular two compartment Kety–Schmidt model. The paper proposes using K^{trans} and k_{ep} for the washin and washout rate constants in the two compartment model (see Fig. 2 and the section on two compartment models below).

Pharmacokinetics is a large related field of study that uses tracer kinetic techniques, sometimes with different nomenclature. The pharmacokinetics field generally seeks to understand drug and biochemical delivery patterns. Drugs of interest may be radiolabeled or tagged with fluorescent or other substances so that they can be tracked and their spatiotemporal behavior identified.

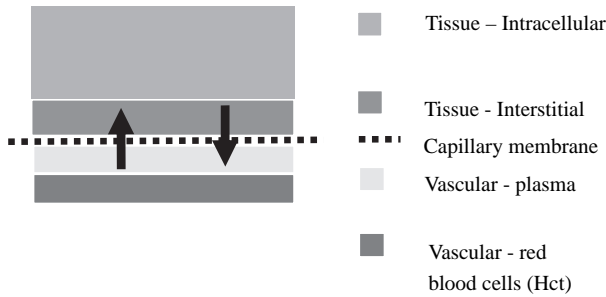


Figure 2. A compartment model type view of a section of capillary (lower two segments) and a section of tissue. The arrows represent the exchange that would occur with an extracellular tracer such as gadolinium. Water would distribute throughout all four regions shown.

COMPARTMENT MODELING

By kinetic modeling we refer to using any of a number of quantitative approaches to extracting information from a time series of data. All of these approaches invoke some sort of mathematical model in order to parameterize the data. The idea is that the parameters can provide more information than processing without models or simple visual examination of the data can provide.

Depending on the type of tracer or contrast agent used, and on the fidelity and resolution of the signal, appropriate models can be selected. That is, the models should not include detail that cannot be discriminated by the measured data, even if it is known to exist physiologically. In practice, almost all models will be required to be simplified versions. The models also need to have useful parameters for the task at hand. Compartment models are widely used for a number of applications due to their relative simplicity and physiological relevance.

A compartment refers to a well-mixed volume in a system of interest that can be considered to have a homogeneous concentration of the tracer. A widely used compartment model of interest is the two compartment model (Fig. 2). This case is often an extremely useful approximation to the system of interest. For example, one compartment is considered to be the vasculature, and one compartment is the tissue of interest. The washin and washout rate constants k_1 and k_2 (or K^{trans} and k_{ep}) govern the concentration in the tissue compartment. The driving compartment, the arterial input function, is usually considered to be measured or otherwise known. For this reason, some authors call this a one compartment model.

Note that a compartment model is often used even in the analysis of multiple heterogeneous regions. Typically, it is assumed that all are driven by the same input function. That is, the blood “compartment” is the same for all of the regions, but the tissue kinetic parameters of uptake and washout of each region may vary. The model remains the same, but the parameters are different for each region.

Kety Model

The roots of compartment modeling approaches are found in Kety’s seminal work in the 1940s (3,4). Kety and

Schmidt were the first to successfully determine the cerebral blood flow (CBF) and oxygen consumption in a relatively noninvasive way. They had normal volunteers breathe air with a ~5–10% concentration of nitrous oxide. The nitrous oxide that was used as a tracer is an inert gas that is soluble in the brain tissue. They then applied the Fick principle to determine CBF. The Fick principle is simply mass balance for convective systems. Use of the Fick principle gives

$$Q_b(t) = CBF \int_0^t C_a(\tau) - C_v(\tau) d\tau \quad (1)$$

where Q_b is the mass quantity of tracer per 100 g in the brain, CBF is the flow of the tracer (and of the blood) in mL/(min·100 g), C_a is the concentration of the tracer in g/mL, and C_v is the concentration of the tracer in the venous blood. Alternatively, for experiments with radiolabeled tracers, Q could instead be in terms of activity in Bequerels and concentration in terms of activity/milliliters.

Rearranging and evaluating at a time $t \sim 15$ min when the system is at equilibrium gives the equation Kety used for global CBF:

$$CBF = \frac{Q_b}{\int_0^t C_a(\tau) - C_v(\tau) d\tau} = \frac{\lambda C_v(t)}{\int_0^t C_a(\tau) - C_v(\tau) d\tau} \quad (2)$$

Here the venous concentration times λ , which is the solubility constant or partition coefficient for nitrous oxide, is used as a measure of the amount of tracer in the brain at equilibrium. The partition coefficient refers to the concentration of the tracer in the tissue divided by the concentration in the blood at equilibrium (at equilibrium arterial and venous blood ideally have the same concentration). Another way to consider the partition coefficient is as a virtual volume, relative to the volume the tracer distributes in a unit of blood. Thus λ is also called the distribution volume. For water, $\lambda = 1$. For nitrous oxide $\lambda = 0.48$ mL/mL. The parameter λ is also sometimes given in units of g/g or mL/g when the density of blood is ~1.06 g/mL or the specific gravity of gray or white matter ~1.05 are used.

For an extracellular tracer such as gadolinium-diethylenetriamine penta-acetic acid (DTPA), $\lambda \sim 0.3$. Figure 2 shows a graphic interpretation of the distribution volume for gadolinium.

In practice, venous measurements of nitrous oxide concentration are taken from an internal jugular vein, and arterial concentrations from a peripheral arterial line. This gives an accurate measure of whole brain CBF, ~50 mL/(min·100 g) in normal humans.

PET H₂¹⁵O Compartment Model

Interestingly, at first glance Kety’s use of the Fick principle does not appear to link to most of today’s compartment models. The approach today is often to write a system of differential equations to describe the behavior of the tracer, and then solve the equations to get a form such as:

$$C_t(t) = C_a(t) \otimes k_1 e^{-k_2 t} \quad (3)$$

where \otimes denotes convolution, and the arterial input function $C_a(t)$ convolved with one or more exponential terms is

fit to the measured image data $C_t(t)$. Though the Kety form has no derivatives or convolutions, Kety's Eq. 2 can be rearranged in a similar form. This rearrangement is illustrated by describing the use of a two compartment model for PET $H_2^{15}O$. Water rapidly diffuses across capillary and cellular membranes, so one can postulate based on Fig. 2 that the rate of change of concentration in the tissue compartment is equal to the input to the compartment (the rate constant k_1 times the arterial concentration) minus the output from the compartment (C_t times the rate constant k_2).

$$\frac{dC_t}{dt} = k_1 C_a - k_2 C_t \quad (4)$$

For the tracer water, $k_2 = k_1$ since the capillary membrane appears symmetrical to this tracer and the distribution volume is 1. Solving this first-order differential equation gives Eq. 3. Note that integrating both sides of Eq. 4 gives the form seen in Kety's Eqs. 1 and 2.

The model can be made more general for tracers that are not freely diffusible everywhere (as water is) by including the partition coefficient. That is, for the more general case of diffusible tracers, $k_1 = F$ and $k_2 = F/\lambda$. This formulation allows estimation of both flow to the region of interest (perfusion) and the volume of distribution of the tracer. Both of these measures can be useful clinically.

Tracer Kinetics of Gadolinium Measured with MRI

Less straightforward tracers, such as the paramagnetic moiety Gd-DTPA used in dynamic contrast MRI, can also be approximated with the two compartment model. Gadolinium diffuses out of the vasculature into the interstitial space, but does not enter cells. Thus the distribution volume reflects the size of the interstitial space. This lack of distribution into cells also creates complications when water exchange is considered. That is, Gd-DTPA is not detected directly with MRI as is the case with radionuclide detection. It is detected by how it changes the relaxation rates of hydrogen in its local microenvironment by interaction of its electrons and the hydrogen protons in blood or tissue (5). So, if water (hydrogen) is moving between where the Gd-DTPA is in the interstitial space and the intracellular space free of gadolinium, it experiences different microenvironments and thus different relaxation rates. Assuming the rate of exchange of water molecules across the cell membrane is rapid, one can neglect this effect on the image signal intensity. Otherwise, more complicated modeling including the water exchange is needed.

Another interesting example of tracer kinetics is the use of gadolinium contrast enhanced MRI for determining if an area of cardiac muscle is dead or alive. A static scan ~15 min after contrast injection can show that gadolinium is present at a higher concentration in infarcted tissue compared to normal tissue. This delayed enhancement is most easily understood through a tracer kinetics approach. Figure 3 shows the time curves from a normal and abnormal region and one can see the abnormal region washes in more slowly, but the washout is a slower process. Another way of interpreting this is the distribution volume is larger. This corresponds intuitively to the agent having more

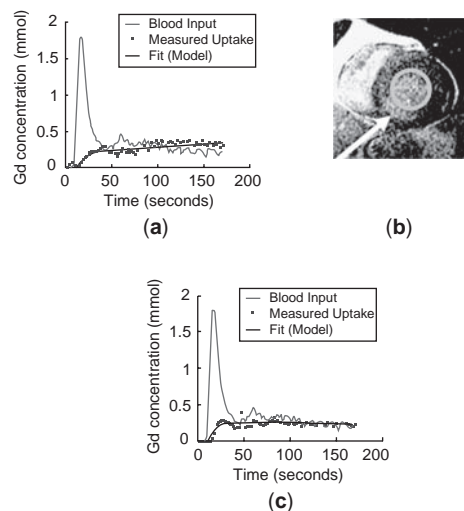


Figure 3. (a) Time curves from measured uptake in region corresponding to arrow in the middle image, distribution volume $v_e = 0.58$ mL/g. The blood input is from a region of interest drawn on the left ventricle cavity. The fit is from a two compartment model for gadolinium-DTPA. (b) “Delayed” image of a short axis slice of the heart (right ventricle on left, left ventricle the large doughnut structure with its endocardial wall traced in green) from MRI acquired ~15 min after contrast injection, showing subendocardial inferior infarct (arrow). (c) Curve from a remote normal region, $v_e = 0.37$. The magnitude and spatial distribution of v_e may help characterize the infarct better than delayed enhancement alone.

volume to distribute in within regions where cell membranes have been destroyed, since gadolinium is an extracellular tracer.

The volume/g of tissue that is occupied by cells in the heart muscle is estimated as ~70%. The interstitial space, or distribution volume for Gd-DTPA, has been reported as $v_e \sim 0.3$ mL/g in healthy tissue, where v_e stands for volume of extracellular extravascular space and is used in place of the symbol λ . This distribution volume increases if cell membranes are disrupted as in acute myocardial infarction (heart attack). Scar tissue, composed of a collagen matrix, will still accumulate gadolinium if it is still perfused. This is likely because the collagen matrix has more interstitial space, but is still not completely understood (6).

Another issue for most tracers is that tissue regions of interest contain on the order of 10% of their volume as vasculature. Thus the time curve $C_t(t)$ is biased by a contribution from something similar to $C_a(t)$. [Due to hematocrit changes in smaller vessels and delay and dispersion, it may not be exactly $C_a(t)$.] One method of accounting for this is to include another parameter in the fit to estimate the blood volume in the region of interest:

$$C_t(t) = C_a(t) \otimes k_1 e^{-k_2 t} + v_b C_a(t) \quad (5)$$

The v_b parameter can also be used to model blood signal in tissue that arises from partial volume effects. Figure 4 illustrates the use of v_b to account for blood spillover into the tissue in a dynamic cardiac MRI study.

Intravascular tracers that do not leave the vasculature may be used to obtain measures of blood volume and to

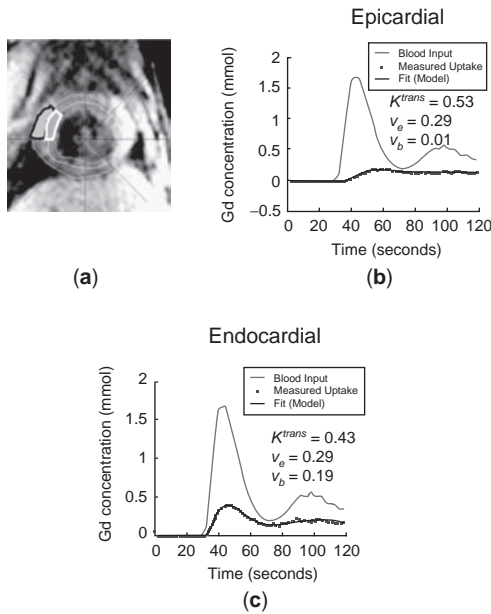


Figure 4. Short axis dynamic cardiac MRI time series study. (a) MRI image precontrast. Epicardial (light blue) and endocardial (orange) regions that give the tissue curves in (b) and (c) are marked. Parts (b) and (c) use the same blood input function and should have similar flows as implied by the K^{trans} ($K^{trans} = k_1$) values given. The point here is that due to partial volume effects, typical upslope or percent enhancement semiquantitative measures would incorrectly give very different flow indices for the two regions. The model used Eq. 5 to obtain blood volume v_b values to compensate for partial volume effects.

measure flow and can be analyzed as

$$C_t(t) = C_a(t) \otimes h(t) \quad (6)$$

where in this case $h(t)$ is not parameterized and the peak value of $h(t)$ can be shown to reflect flow (7).

Impulse Response

The response of a system to the input of an impulse, or a delta function $\delta()$, completely characterizes the system if it is a linear time-invariant system. Equation 3 shows that the impulse response for a two compartment model with rate constants k_1 and k_2 is $k_1 * \exp(-k_2 t)$.

While many physiological systems are not linear time invariant, linear time-invariant systems often serve as a good approximation. Examples of nonlinear systems include enzyme kinetics, saturable receptors, or any process that saturates (as the input continues to double the output does not). Too much of any tracer may force a system to nonlinearity, or at least to some abnormal physiological state. For example, at some point there could be so much FDG in the blood stream that doubling the amount only changes the uptake in the tissue by a small amount. More complicated nonlinear and time-varying models are needed for processes such as insulin kinetics (8).

Three or More Compartments

More complex systems, for example, when a tracer diffuses out of the vasculature and may either diffuse back, or

remain freely available in the interstitial space, or undergoes a chemical trapping such as phosphorylation, requires a more complex model. The two compartment model may not be sufficient since the tissue now has two different pools of nonuniform concentrations. Note that these pools generally are not measured separately (Fig. 1). The two compartment model can still be used as an approximation, but a three-compartment model (Fig. 1) is likely more appropriate. A more complex model with four compartments has also been proposed for the case of FDG uptake in skeletal muscle (9).

Order of Model

How does one determine if the lower order model or the higher order model is best to use? This is certainly very task dependent, and is an open question. One approach is the Akaike Information Criteria, or AIC (10). This measure is based on the intuitive notion that as the number of parameters increases, the fit to the data should improve in proportion ($AIC = \text{residual sum of squares} + 2 * P$, where P is the number of parameters being fit).

If the tracer kinetics are not well understood and the order of model to use is unclear, more exploratory methods such as spectral analysis are appropriate. Spectral analysis models the impulse response as a summation of a number of exponentials (11). That is, spectral analysis is a type of generalized compartment model where the impulse response is a linear combination of N decaying exponentials: $h(t) = \sum_{i=1}^N \alpha_i \exp(-\beta_i t)$. For example, if the data truly is from a two compartment model, then only one of the exponential terms will have a nonzero coefficient. The drawback, or advantage, depending on one's viewpoint for a particular problem, is that the coefficients are fixed so that the problem becomes linear as only α 's need be estimated. The selection of the spectrum of exponential decay rates can thus be quite important. A logarithmic distribution of α has been used in PET (11).

Even if the model is well-known, spectral analysis may be a desirable approach since the linearized fitting can be more rapid than standard nonlinear compartment modeling. Several published studies in PET and SPECT have found high correlations between compartment modeling and spectral analysis methods. Such studies are reviewed in Ref. 12.

A higher level of realism can be obtained by using a distributed model of multiple parallel pathways of capillaries feeding the tissue region of interest. XSIM is freely available software useful for modeling such distributed systems where blood-tissue exchange units form basic building blocks that can have differing delays or flows or permeability (13). Most often the measurement system, particularly medical imaging, does not have the spatial and temporal resolution and is too noisy to robustly use such a complex model. Such models are, however, of great importance in understanding the underlying physiology (8,13).

Model-Free Methods

It is not required to use a specific model. Some "model-free" methods are easier to use or have the advantage of making fewer assumptions regarding the tracer distributions. The

downside is that the parameters may not offer as much insight into physiological mechanisms underlying the imaging results.

The mean transit time is a measure of how long the tracer requires to pass through the capillary bed from the arterial to the venous side. One can consider the impulse response as a histogram of transit times through the region. The expected value of the impulse response provides the mean transit time if there is no recirculation [$\tau = \int th(t)dt$]. Or, for a short bolus, the area divided by the height at time zero of the tissue uptake time curve gives τ .

$$\tau = \int_0^{\infty} C_t(t)dt/C_t(0)$$

The mean transit time is on the order of 5 s for nondiffusible intravascular tracers in the brain, and on the order of minutes for diffusible tracers (14). Or one can define the mean transit time as the inverse of the washin rate constant, $1/k_1$ (14).

Other easily computable parameters such as maximum upslope, percent signal enhancement, area under the curve, or the time until peak signal have been suggested and used as model-free approaches.

ESTIMATING KINETIC PARAMETERS

Weighted Least Squares

The estimation of kinetic parameters from noisy data is a well-studied problem. The errors in the parameters can be estimated based on the variance in the data. The model can be written as:

$$C_t(t) = h(t) \otimes C_a(t) + \varepsilon \quad (7)$$

where ε is an additive noise process. The impulse response can be cast in a toeplitz matrix H and the equations expressed

$$\mathbf{y} = H\mathbf{b} + \varepsilon \quad (8)$$

where $\mathbf{y} = C_t(t)$, possibly for multiple regions, and $\mathbf{b} = C_a(t)$, and the unknowns are the parameters \mathbf{p} that form H . The weighted least-squares fitting process minimizes $\mathbf{y} - H\mathbf{b}$, weighted by the data covariance matrix $\Phi = \text{cov}(\mathbf{y})$:

$$\min_{\mathbf{p}} (\mathbf{y} - H\mathbf{b})^T \Phi^{-1} (\mathbf{y} - H\mathbf{b}) \quad (9)$$

If the measured data \mathbf{y} have uniform Gaussian noise then Φ^{-1} will be a diagonal matrix with on the diagonal and will not be any different from an unweighted fit. Technically, Φ should be the covariance matrix of the residuals $\mathbf{y} - H\mathbf{b}$, since \mathbf{b} is typically a measured quantity with noise. Huesman and Mazoyer compared the results with the use of $\text{cov}(\mathbf{y})$ and $\text{cov}(\mathbf{y} - H\mathbf{b})$ as the weighting function in the fitting of dynamic PET data with a noisy arterial input function (15).

One can consider the system of equations (Eq. 8) in the more "standard" form of $\mathbf{y} = A\mathbf{p}$, where \mathbf{p} are the unknown parameters. If we assume the model is linear in the para-

eters (i.e., $d\mathbf{y}/d\mathbf{p} = A$) then the best linear unbiased estimate, the estimate with minimum variance (16), is given by

$$\mathbf{p} = (A^T \Phi^{-1} A)^{-1} A^T \Phi^{-1} \mathbf{y} \quad (10)$$

This gives the parameters that minimize the $\text{cov}(\mathbf{p})$ matrix $E[(\mathbf{p} - \hat{\mathbf{p}})(\mathbf{p} - \hat{\mathbf{p}})^T]$. The point is that given this form, one can show that the resulting cov matrix of the parameter estimates is the Fisher information matrix

$$\text{cov}(\mathbf{p}) = (A^T \Phi^{-1} A) \quad (11)$$

where A are the partial derivatives with respect to the parameters of the model, also called the sensitivity matrix.

More commonly, and for the case of a compartment model, the matrix A will be a function of the parameters. So the model will be nonlinear in the parameters. Equation 11 still holds if A is evaluated at the true parameters \mathbf{p} , and the measured data are Gaussian (17).

The $\text{cov}(\mathbf{p})$ is the combination of the sensitivity matrix A and the data covariance matrix Φ . In addition to providing a way to estimate the error in the fitted parameters, one can also compare data acquisition strategies in this manner. Mazoyer and Huesman compared input function shapes and temporal sampling strategies by comparing the determinants of their information matrices (17).

Note that Eq. 8 need not only represent a single measured region. The measured curves \mathbf{y} and impulse responses can be stacked to form a larger system of equations using the same AIF \mathbf{b} and different parameters for each region. (see, e.g., Ref. 18).

Bayesian Approaches

Two quite different Bayesian approaches have been published. One class of methods combines tomographic reconstruction with compartment models (19,20). The method in (19) assumes the AIF is known a priori and then encourages the reconstructed time courses of each voxel to fit a two compartment model. The reconstruction is performed simultaneously to satisfy the tomographic reconstruction criteria that the projections of the estimated image match the measured projections. The method was applied to dynamic teboroxime SPECT data and resulted in improved washin estimates.

Another type of Bayesian approach is to use multiple (normal) regions to estimate the kinetic parameters using a two step process (21). The first iteration calculates the parameters for the regions independently. These are then averaged and their covariance matrix computed and used in the second step in a penalized least squares, or Bayesian, formulation:

$$\min_{\mathbf{p}} (\mathbf{y} - H\mathbf{b})^T \Phi^{-1} (\mathbf{y} - H\mathbf{b}) + (\mathbf{p} - \mu)^T \Omega^{-1} (\mathbf{p} - \mu) \quad (12)$$

where \mathbf{p} are the parameters to be estimated, μ is the average of the parameters estimated in the standard fashion, and

$$\Omega = 1/M \sum_{m=1}^M (\mathbf{p}^m - \mu)(\mathbf{p}^m - \mu) + 1/M \sum_{m=1}^M (A^m \Phi^{m-1} A^{mT})^{-1} \quad (13)$$

That is, Ω is the standard deviation of the parameters estimated in the first step and the estimated uncertainty

of the parameter estimates. The parameter M is the number of regions (assumed homogeneous) and m is used to index the M regions. This approach is equivalent to a specific case of ridge regression (22). These Bayesian methods represent a new class of analysis techniques for tracer kinetics.

IMPROVING AND AUTOMATING ANALYSIS OF THE KINETIC TRACER CURVES

While some applications have sufficient signal-to-noise ratio (SNR) to perform pixelwise analysis of the kinetic tracer curves, other applications can benefit greatly from spatial averages of the data. Methods such as factor analysis and clustering can be useful for automatically identifying blood and tissue curves from the image series data (23). This approach may give better kinetic parameter estimates and be more automated than manual delineation of regions.

Factor Analysis

Factor analysis, also termed FADS, which stands for Factor Analysis of Dynamic Structures, has been somewhat widely used in nuclear medicine applications (24–27). The FADS is essentially principal components analysis followed by a “rotation” to satisfy nonnegativity conditions. While this does not provide a unique solution, several different FADS methods have been validated as clinically useful for certain tasks. As well, a number of modifications to the basic method to incorporate additional information have been proposed, for example, (27).

Clustering

Clustering is becoming a popular method for a data-driven approach to choosing the regions. K-means clustering, also termed c-means clustering, groups alike time–activity curves (or time-concentration curves) by computing a distance between each pair of curves and grouping curves into the cluster that they are closest to. The cluster centers are updated (continuously or at discrete intervals). Most often a Euclidean distance is used to determine to which cluster a voxel belongs.

It is still an open issue to make factor analysis, clustering, and automated approaches robust and clinically practical. Quantitative imaging finds particular importance for longitudinal studies. These studies also need robust registration and processing methods to perform well.

Since a number of the analysis methods are nonlinear, it is typically left to empirical task-specific studies to determine if specific image processing methods can significantly impact the accuracy or robustness of the results.

ISSUES WITH THE ARTERIAL INPUT FUNCTION, AIF

Importance of an Accurate AIF

In order to obtain accurate absolute measures of kinetic parameters, it is important to use an accurate arterial input function. This is critical for any of the models or model-free approaches that use the AIF. Accurate input

functions, or at least some measure of what the tissue “sees” is essential for obtaining quantitative parameters and is intrinsic to model-based methods. Many “model-free” methods also rely on measures of input and output to obtain the mean transit time from the images. As other issues key to acquiring good quality dynamic data (scanners, computer speed, and memory) are surmounted, more effort is being focused on estimating the AIF more accurately and automatically.

In many situations, it is difficult to obtain an accurate AIF. If the tracer of interest binds to red blood cells or plasma proteins, then the AIF obtained from the images does not reflect the concentration of the tracer that is available for uptake by the tissues. Or the contrast agent in MRI may be saturated and not provide a true measure of gadolinium concentration when the concentration is high, or there may be flow effects that change the MRI signal. Arterial blood samples and subsequent processing are required to obtain accurate AIFs in these settings.

Methods for Improving the AIF Measurement

A number of techniques for improving the AIF have been proposed. For example, in contrast MRI studies of the brain, either a cluster or a single voxel that meets certain criteria can be automatically obtained and used as the AIF. In MRI cardiac studies, either a small bolus given before the main bolus to give the shape of the AIF curve without saturation, or a pulse sequence modification to obtain a nonsaturated AIF are current methods to improve AIF estimation.

Blind Deconvolution—Methods for Estimating an Unmeasured AIF

There have recently been efforts to estimate the arterial input function jointly with the parameters of interest. That is, given the measured data y , estimate both H and b in the equation:

$$y = Hb + \epsilon \tag{14}$$

In the field of telecommunications, this is termed blind estimation or blind deconvolution, since the input is not known (28). While this is not possible given a single region of interest, multiple regions with differing kinetic parameters, driven by the same input, can provide a system of equations with sufficient measurements to estimate the AIF within a global scale factor. Then, Eq. 14 is composed of a number of stacked matrices (18). Another way to see this is to consider two regions without using the matrix formulation:

$$\begin{aligned} y_1(t) &= b(t) \otimes h_1(t) \\ y_2(t) &= b(t) \otimes h_2(t) \end{aligned} \tag{15}$$

Then convolving both sides of Eq. 16 with $h_2(t)$ and substituting from Eq. 12 gives the “cross-relation” expression (29):

$$\begin{aligned} h_2(t) \otimes y_1(t) &= h_2(t) \otimes b(t) \otimes h_1(t) \\ y_1(t) \otimes h_2(t) &= y_2(t) \otimes h_1(t) \end{aligned}$$

This last expression can be solved for \mathbf{p} (recall h is a function of \mathbf{p}) with no knowledge of the AIF! The solution is unique with a two compartment model and the use of two different regions, to within a global scale factor (18). A three-compartment model requires more than two regions (or the use of other constraints) to obtain a unique solution (30).

The blind deconvolution method is sensitive to noise (18). Future hybrid approaches incorporating some measurements and other a priori information such as population expectations with the blind deconvolution technique will likely result in less sensitivity to noise and artifact.

FUTURE ISSUES

The last few topics—the Bayesian fitting approaches, factor analysis, clustering, and blind deconvolution represent cutting-edge directions for analysis of tracer kinetics data. These techniques need more research and validation before being widely used in clinical applications.

More and more dynamic studies of tracer kinetics will be performed as the equipment (in particular computer speed and disk space) allows dynamic acquisition and processing with smaller penalties. While the analysis of the studies will be task dependent, insight into the fundamentals of acquiring and analyzing tracer kinetics will be applicable to all fields.

Tracer development is perhaps in its highest gear ever since the largest advances, particularly in molecular imaging, are likely to result from new, more specific, and more physiologically relevant tracers. A hot area is the design of tracers to follow genes of interest, reporter genes, and to track stem cells and disease processes (31).

While the ideal tracer may go in proportion to the site of interest and be stuck there so that static imaging provides full information, it is much more likely that tracers will continue to exhibit complex properties. Thus research into optimal acquisition and analysis of the temporal and spatial distributions of the tracer will remain an important field and will likely even grow in significance. Analysis of dynamic data can offer quantitative and possibly even absolute measures of many clinically important parameters. The use of tracer kinetic principles and image time series from modalities such as optical, MRI, ultrasound, PET, and SPECT will continue to have very important application to cardiac, brain, renal, liver, and other dynamic systems.

BIBLIOGRAPHY

- Zaret BL, Beller GA. *Clinical Nuclear Cardiology: State of the Art and Future Directions*. 3rd ed. New York: Mosby-Year Book; 2005.
- Rescigno A, Thakur AK, Brill AB, Mariani G. Tracer kinetics: A proposal for unified symbols and nomenclature. *Phys Med Biol* 1990;35:449–465.
- Kety SS, Schmidt CF. The nitrous oxide method for the quantitative determination of cerebral blood flow in man: Theory, procedure, and normal values. *J Clin Invest* 1948;27:476–483.
- Kety SS. The theory and applications of the exchange of inert gas at the lungs and tissues. *Pharmacol Rev* 1951;3:1–41.
- Donahue KM, Burstein D, Manning WJ, Gray ML. Studies of Gd-DTPA relaxivity and proton exchange rates in tissue. *Mag Res Med* 1994;32:66–76.
- Thomson LE, Kim RJ, Judd RM. Magnetic resonance imaging for the assessment of myocardial viability. *J Mag Res Imaging* 2004;19:771–788.
- Ostergaard L, Weisskoff RM, Chesler DA, Gyldensted C, Rosen BR. High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. Part I: Mathematical approach and statistical analysis. *Mag Res Med* 1996;36:715–725.
- Carson E, Cobelli C. *Modelling Methodology for Physiology and Medicine*. New York: Academic Press; 2001.
- Bertoldo A, Peltoniemi P, Oikonen V, Knuuti J, Nuutila P, Cobelli C. Kinetic modeling of [18F] FDG in skeletal muscle by PET: A four-compartment five-rate-constant model. *Am J Physiol Endocrinol Metab* 2001;281:E524–536.
- Akaike H. A new look at the statistical model identification. *IEEE Trans Auto Control* 1974;19:716–723.
- Cunningham V, Jones T. Spectral analysis of dynamic PET studies. *J Cereb Blood Flow Metab* 1993;13:15–23.
- Murase K. Spectral analysis: Principle and clinical applications. *Ann Nuclear Med* 2003;17:427–434.
- Bassingthwaight JB, Goresky CA. Modeling in the analysis of solute and water exchange in the microvasculature. In: Renkin EM, Michel CC, editors. *Handbook of Physiology—The Cardiovascular System*. Bethesda: American Physiology Society; 1984.
- Cherry SR, Sorenson JA, Phelps ME. *Physics in Nuclear Medicine*. 3rd ed. Saunders; 2003.
- Huesman RH, Mazoyer BM. Kinetic data analysis with a noisy input function. *Phys Med Biol* 1987;32:1569–79.
- Strang G. *Introduction to Applied Mathematics*. Wellesley: Wellesley-Cambridge Press; 1986.
- Mazoyer BM, Huesman RH, Budinger TF, Knittel BL. Dynamic PET data analysis. *J Comput Assist Tomogr* 1986;10:645–653.
- Riabkov DY, DiBella EVR. Estimation of kinetic parameters without input functions: Analysis of three methods for multi-channel blind identification. *IEEE Trans Biomed Eng* 2002;49:1318–1327.
- Kadmas DJ, Gullberg GT. 4D maximum a posteriori reconstruction in dynamic SPECT using a compartmental model-based prior. *Phys Med Biol* 2001;46:1553–1574.
- Meikle SR, Mattherw JC, Cunningham VJ, Bailey DL, Livieratos L, Jones T, Price P. Parametric image reconstruction using spectral analysis of PET projection data. *Phys Med Biol* 1998;43:651–666.
- Bertoldo A, Sparacino G, Cobelli C. Population approach improves parameter estimation of kinetic models from dynamic PET data. *IEEE Trans Med Imag* 2004;23:297–306.
- O'Sullivan F, Saha A. Use of ridge regression for improved estimation of kinetic constants from PET data. *IEEE Trans Med Imaging* 1999;18:115–125.
- DiBella EVR, Sitek A. Time curve analysis techniques for dynamic contrast MRI studies. *Information Processing in Medical Imaging* 2001;LNCS 2082:211–217.
- Barber DC. The use of principal components in the quantitative analysis of gamma camera dynamic studies. *Phys Med Biol* 1980;25:283–292.
- Samal M, Karny M, Surova H, Marikova E, Dienstbier Z. Rotation to simple structure in factor analysis of dynamic radionuclide studies. *Phys Med Biol* 1987;32:371–82.
- Buvat I, Benali H, DiPaola R. Statistical distribution of factors and factor images in factor analysis of medical image sequences. *Phys Med Biol* 1998;43:1695–1711.

27. Sitek A, DiBella EVR, Gullberg GT. Factor analysis with *a priori* knowledge—application in dynamic cardiac SPECT. *Phys Med Biol* 2000;45:2619–2638.
28. Tong L, Perreau S. Multichannel blind identification: from subspace to maximum likelihood methods. *Proc IEEE* 1998;86:1951–1968.
29. DiBella EVR, Clackdoyle R, Gullberg GT. Blind estimation of compartmental model parameters. *Phys Med Biol* 1999;44:765–780.
30. Riabkov DY, DiBella EVR. Blind identification of the kinetic parameters in three-compartment models. *Phys Med Biol* 2004;49:639–664.
31. Massoud TF, Gambhir SS. Molecular imaging in living subjects: seeing fundamental biological processes in a new light. *Genes Dev* 2003;17:545–580.

See also FLOWMETERS, ELECTROMAGNETIC; PHARMACOKINETICS AND PHARMACODYNAMICS; PHYSIOLOGICAL SYSTEMS MODELING; RADIONUCLIDE PRODUCTION AND RADIOACTIVE DECAY; STATISTICAL METHODS.

TRANSCUTANEOUS ELECTRICAL NERVE STIMULATION (TENS)

ANJU MADNANI
 SANJAY MADNANI
 RANDALL CARK
 JASON WELLS
 LSU Medical Centre
 Shreveport, Louisiana

INTRODUCTION

The field of pain management is a rapidly expanding one, and new treatment modalities are being discovered or rediscovered. Electroanalgesia has a long and sometimes dubious history, dating back to the ancient Egyptians. However, the publishing of “the gate control theory” of pain transmission in 1965 by Melzack and Wall transformed our understanding of pain, its transmission, and how it is modulated. With this discovery, electroanalgesia underwent a revolution, and transcutaneous electrical nerve stimulation (TENS) was developed and is continually being refined. Today, our understanding of the mechanism by which TENS produces analgesia continues to expand does its potential applications.

This article provides a review of pain, its definition, types, and physiology. It provides background information and theories surrounding the mechanism of analgesic action of TENS and the development of electroanalgesia. It discusses the usage, design, applications, and warnings surrounding TENS.

WHAT IS PAIN?

Pain is an unpleasant sensory and emotional experience associated with actual or potential damage, or described in terms of such damage. Pain serves as an essential defense mechanism to protect one’s body from potential damage. Indeed, the disastrous consequences of diminished or

absent pain signaling become readily apparent in diseases and conditions that result in partial or complete damage of the nerves that innervate the extremities (e.g., diabetic neuropathy, tabes dorsalis, tuberculoid leprosy, and many others). While serving an essential function, pain can often present for physiologically inappropriate reasons, continue far past the removal of noxious stimuli, remain long after wound healing, or even present for purely psychological reasons. This maladaptive and uncontrolled pain cycle afflicts an estimated 40 million Americans (1), and research into the causes and cures of pain is a rapidly expanding branch of medical science and forms the basis for a multibillion dollar a year, multidisciplinary industry.

Pain can be categorized either temporally as in acute or chronic pain or by the mechanism. Nocioceptive–inflammatory pain is produced after an appropriately perceived tissue injury. Neuropathic pain, however, is produced by nerve injury that is inappropriately perceived due to neuroplasticity. Often described as a burning or electric sensation, neuropathic pain can persist long after an injury or for completely idiopathic reasons. Even simple light touch or changes in temperature are enough to trigger severe bouts of extreme pain, lasting seconds to hours or longer (i.e., trigeminal neuralgia).

Phantom limb pain is another incompletely understood neuropathic phenomenon and occurs in 50–67% of postsurgical amputation patients (2). It is often described as a minor-to-severe cramping or, less commonly, as a burning sensation (3). While this commonly subsides with time, in ~10% of patients, this pain persists and is often refractory to NSAID or opiate therapy, traditional first and second line agents in the treatment of pain.

THE PHYSIOLOGY OF PAIN

The process of nociception is complicated, but can be divided into four distinct physiological processes transduction, transmission, modulation, and perception. Transduction, the translation of noxious stimuli into electrical activity at the sensory endings of nerves, occurs at unspecialized mechano-, thermo-, or polymodal (thermal and chemical) nociceptors, as well as at unspecialized nerve endings.

Polymodal nociceptors respond to a variety (i.e., chemical, mechanical, and temperature extremes) of intense noxious stimuli. Thermonociceptors are distinct from thermoreceptors that transmit non-noxious temperature information. This class of nociceptors functions from temperature ranges of roughly <5 to >45 °C. Mechanonociceptors are activated when intense pressure stimulates them; as with thermonociceptors, the mechanonociceptors are distinct from the receptors that transmit non-noxious light and strong touch, vibratory information, and so on. Additionally, visceral “silent” nociceptors exist in a default dormant state and are usually activated only in the presence of inflammatory mediators. These silent nociceptors likely are involved in hyperalgesia as discussed below.

In the peripheral nervous system, small unmyelinated C polymodal nociceptive fibers, as well as the larger,

lightly myelinated A δ mechano- and thermonociceptive fibers transmit noxious stimuli to the dorsal horn of the spinal column. The small C fibers are responsible for what is termed slow pain and transmit data at under 2.5 m·s⁻¹. These small fibers outnumber the larger, lightly myelinated A δ fibers, responsible for fast pain, which conduct at a rate of 4–30 m·s⁻¹, by a ratio of ~7:1 in the epithelium. The concept of slow and fast pain is easily conceptualized by a hypothetical injury of one stepping on a nail. The initial sharp sensation, or fast pain, is transmitted by the larger A δ fibers, while the nagging dull ache, or slow pain, is transmitted by the smaller, unmyelinated C fibers (Fig. 1).

The A δ fibers synapse with projecting neurons in lamina 1 of the dorsal horn of the spinal cord. In addition to this

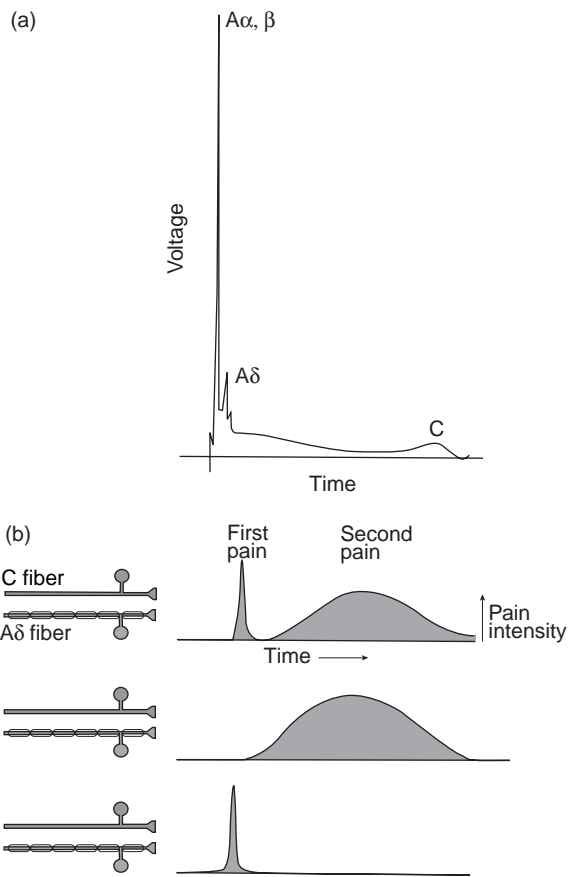


Figure 1. Propagation of action potentials in sensory fibers results in the perception of pain. (Modified from Ref. 4). (a) This electric recording from a whole nerve shows a compound action potential representing the summated action potentials of all the component axons in the nerve. Even though the nerve contains mostly nonmyelinated axons, the major voltage deflections are produced by the relatively small number of myelinated axons. This is because action potentials in the population of more slowly conducting axons are dispersed in time, and the extracellular current generated by an action potential in a nonmyelinated axon is smaller than the current generated in myelinated axons. (b) First and second pain are carried by two different primary afferent axons. First pain is abolished by selective blockade of A δ myelinated axons (middle) and second pain by blocking C fibers (bottom).

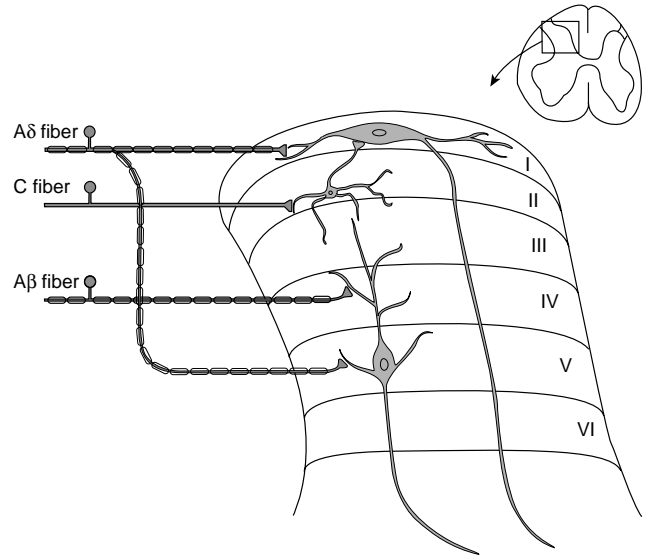


Figure 2. Nociceptive afferent fibers terminate on projection neurons in the dorsal horn of the spinal cord. Projection neurons in lamina I receive direct input from myelinated (A δ) nociceptive afferent fibers and indirect input from unmyelinated (C) nociceptive afferent fibers via stalk cell interneurons in lamina II. Lamina V neurons are predominately of the wide dynamic-range type. They receive low threshold input from the large-diameter myelinated fibers (A β) of mechanoreceptors, as well as both direct and indirect input from nociceptive afferent fibers (A δ and C). In this figure the lamina V neuron sends a dendrite up through lamina IV, where it is contacted by the terminal of an A β primary afferent. A dendrite in lamina III arising from a cell in lamina V is contacted by the axon terminal of a lamina II interneuron. (Adapted from Ref. 4.)

direct, afferent input, these projecting neurons receive indirect input from the stalk cell neurons in lamina II. These stalk cell interneurons of lamina II receive their afferent input from the C fibers that synapse with them. The projecting neurons of lamina V receive afferent input from the large myelinated A β , non-noxious, sensory fibers via dendritic synapse in lamina IV, from synapse with A δ fibers in lamina V, and project both to lamina III as well as higher cortical centers (5,6) (Fig. 2).

In the dorsal horn of the spinal cord at the synapse level, the afferent pain signal can be modulated to either lessen or amplify the body's response to the pain signal. Serotonin as well as norepinephrine act either directly presynaptically to inhibit the propagation of the pain signal or via activating inhibitory interneurons. The enkephalins, endogenous δ and μ opiate receptor agonists, function at this level to serve a similar inhibitory function. The neuromodulator peptide, substance P is released, along with glutamate from the C fibers, and both work allosterically to amplify the pain signal transmission to higher levels.

Once in the dorsal horn of the spinal cord and after synapse, the afferent pain signal is transmitted to higher cortical centers via either the spinothalamic, spinoreticular, spinomesencephalic, cervicothalamic, or spinohypothalamic pathways. Perception is the final process where all above processes as well as prior physical and psychological experiences interact and create the final subjective and emotional experience of pain. The opioids, both endo-

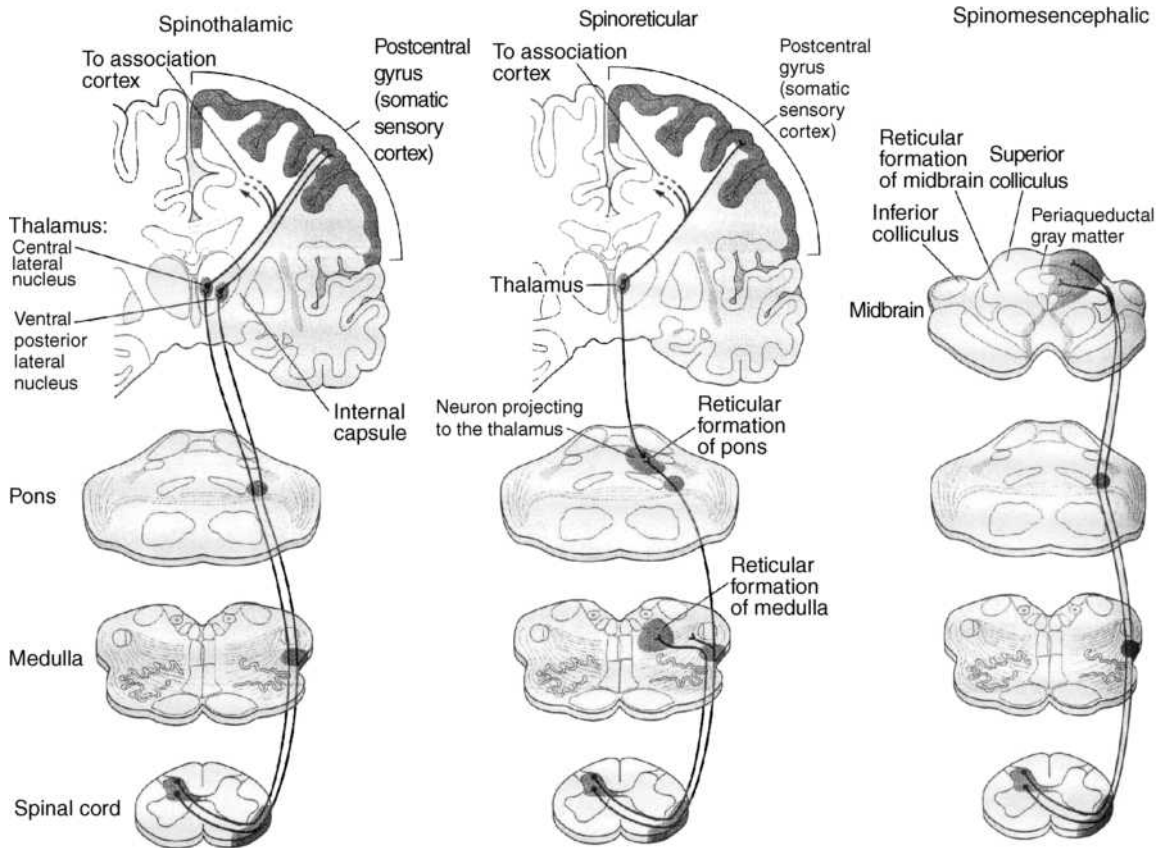


Figure 3. Three of the major ascending pathways that transmit nociceptive information from the spinal cord to higher centers. The spinothalamic tract is the most prominent ascending nociceptive pathway in the spinal cord. (Adapted from Ref. 7.)

ogenous and exogenous, function to alter perception of pain at the cortical level, as well as to activate inhibitory interneurons in the periaqueductal gray area (Fig. 3).

HYPERALGESIA AND SENSITIZATION

In certain situations, nociceptors can become exquisitely sensitive to stimulation or activated in greater numbers than usual. This results in hyperalgesia and is termed sensitization; this process occurs via distinct mechanisms both peripherally as well as centrally. While peripheral sensitization occurs in both acute and chronic phases of injury, central sensitization generally occurs in the chronic phase of insult, after repetitive noxious events.

Upon peripheral injury, for example, an epithelial incision, inflammation is affected via a large number of chemical mediators, such as prostaglandins, leukotrienes, bradykinin, serotonin, substance P, histamine, potassium, and others, released from both damaged, as well as surrounding tissues (5). These inflammatory mediators serve not only to result in inflammation, but also serve to decrease the threshold to stimulate surrounding nociceptors. This can be done by directly acting to affect sensitization or by working in tandem to sensitize nociceptors via another chemical mediator. For example, bradykinin is an important and extremely potent mediator of hyperalgesia.

It works not only to directly sensitize the nociceptive fibers (i.e., C and A δ fibers), but also serves to stimulate local tissue to produce prostaglandins, which themselves result in sensitization. In addition to bringing about sensitization of nociceptors, some chemical mediators directly activate nociceptors, for example, histamine activating polymodal nociceptors (Table 1).

With continued C fiber pain signal transmission due to persistent noxious insult, increased glutamate is released from their end plates in the dorsal horn. With this increased glutamate release, continued opening of postsynaptic calcium ion channels results. This is mediated by postsynaptic *N*-methyl-d-aspartate (NMDA)-type glutamate receptors. This process, termed "wind-up", results in a continual increase in dorsal horn neuron response to the pain signal. This is an example of pain signal modulation. In addition to this progressively increasing response to the pain signal, dorsal horn neurons can become more easily excitable to a lesser peripheral signal. This process, termed central sensitization, is also mediated by NMDA-type glutamate receptors. Additionally, there is an upregulation in production of a variety of neurotransmitters, neurohormones, and their receptors. Effectively, these changes of excitability and magnitude of C fiber response constitute a pain "memory" and also result in progressively larger areas of peripheral tissue coverage of the dorsal horn neuron. Central sensitization with resultant

Table 1. Naturally Occurring Agents that Activate or Sensitize Nociceptors^a

Substance	Source	Enzyme Involved in Synthesis	Effect on Primary Afferent
Potassium	Damaged cells		Activation
Serotonin	Platelets	Tryptophan hydroxylase	Activation
Bradykinin	Plasma kininogen	Kallikrein	Activation
Histamine	Mast cells		Activation
Prostaglandins	Arachidonic acid–damaged cells	Cyclooxygenase	Sensitization
Leukotrienes	Arachidonic acid–damaged cells	5-Lipoxygenase	Sensitization
Substance P	Primary afferents		Sensitization

^aModified from Ref. 4.

hyperexcitability helps explain allodynia, the perception of a non-noxious stimulus, such as light touch, as a painful stimulus. In light of these changes, it is obvious not only why chronic pain can be so difficult to treat, but also why it is important to break pain “cycles” before chronic changes begin to occur.

Allodynia is classically seen in several different chronic neuropathic pain syndromes for reasons that are not always completely understood, but likely stem from the chronic changes outlined above. Herpes zoster is perhaps the best known of these conditions with many sufferers reporting a severe dermatomal burning pain long after the peripheral nerve damage has healed. Allodynia is common following an attack, and severe bouts of pain can be precipitated from the friction between ones shirt and ones skin. Trigeminal neuralgia is another such chronic condition where allodynia is common. In this condition, lightly stroking one’s cheek or the process of eating can precipitate attacks of severe, stabbing transient pain, followed by longer periods of a moderate to severe burning sensation.

PSYCHOLOGICAL ASPECTS OF PAIN

As mentioned earlier, perception of pain is an individualized phenomenon. It is affected by culture, mood, and individual experiences (8,9). Chronic pain can be termed as pain that persists for a certain period, usually ~6 months, after an injury has healed or when the noxious source is idiopathic and central sensitization has occurred. The field of pain management employs a diverse, polymodal disciplinary strategy toward treating pain that extends far beyond simple pharmacotherapy. It includes interventional therapy, physical therapy, meditation, biofeedback, acupuncture, psychiatric therapy, electroanalgesia, and many other treatment modalities. There is a definite psychological component to chronic pain that can cause it to be perceived as much more severe than acute pain and make it refractory to traditional therapy, and chronic pain is frequently associated with depression.

THEORIES OF PAIN

Gate Control Theory

The gate control theory, published in 1965 by Ron Melzack and Pat Wall (10), was the theory from which modern electrotherapy has evolved and that has helped revolutionize

treatment and our understanding of pain. The theory states that pain perception depends on the balance of large, afferent sensory A β and small diameter afferent nociceptive A δ and C fiber activity. According to the theory, non-nociceptive sensory fibers can activate neurons in the substantia gelatinosa. These neurons can decrease or inhibit the pain signals of nociceptive neurons prior to higher level transmission. This theory explains the common practice of rubbing an acute wound to lessen pain. It is worth noting that the inhibitory effect of nonnociceptive neurons is a local one. No analgesic effect exists when rubbing one’s toes after an injury to one’s fingers (Fig. 4).

The theory’s emphasis on the modulation of inputs in the spinal dorsal horn and the dynamic role of the brain in pain processes had a clinical, as well as a scientific impact, and after this theory several methods were developed to modulate the input. Physical therapists and other health-care professionals began developing and refining different modulation techniques, such as implantable dorsal spinal electrostimulation, and later transcutaneous electrical nerve stimulation devices as well as rediscovering old ones, such as acupuncture. After this discovery, TENS became an important part in treating the acute and chronic pain. See below for a much more thorough discussion of the history of electroanalgesia.

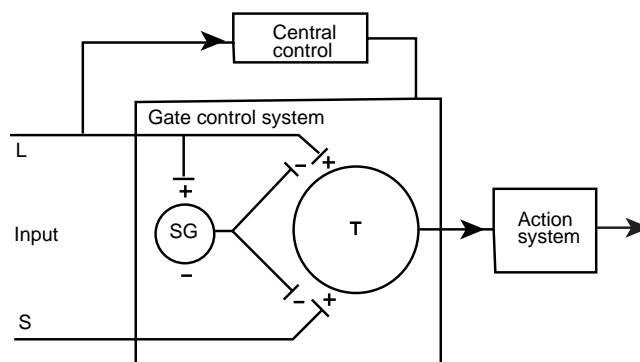


Figure 4. Schematic diagram of the Melzack–Wall gate control theory of pain mechanisms. Large- (L) and small-diameter (S) afferent fibers project to the substantia gelatinosa (SG) and first central transmission (T) cells. The inhibitory effect (-) of SG on the afferent terminals is increased (+) by activity in L fibers and decreased by activity in S fibers. A specialized system of L fibers (the central control trigger) activates certain cognitive processes that influence the modulating properties of the apinal gating mechanism via descending fibers. (From Ref. 10 R. Melzack and P. D. Wall, *Science*, 150:971–979, 1965, © 1965, AAAS.)

Other Theories Regarding TENS' Analgesic Effect

Other theories have been developed to explain the effectiveness of TENS, namely, the enkephalin and endorphin theories, and likely all three contribute to the analgesic effect. Multiple studies have demonstrated an increase in dorsal horn enkephalin production (11,12) as well as have demonstrated that blockade of opiate receptors lessens or even prevents analgesia from TENS (13–15). As briefly described earlier, enkephalins are μ and δ opiate receptor agonists and function as inhibitory neurotransmitters. Release of enkephalins from inhibitory interneurons decrease Ca^{2+} influx into the nociceptive neuron, thereby preventing, or lessening depolarization time, prevents or lessens excitatory neurotransmitters, such as glutamate, release, thereby negatively modulating the pain signal. Additionally, enkephalins function postsynaptically to activate K^+ conductance, thereby hyperpolarizing the dorsal horn projecting neuron and inhibiting pain transmission to further higher cortical centers (Fig. 5).

While enkephalins have a short half-life and function locally, recent studies (16–18) have demonstrated increased levels of circulating endorphins. In contrast to enkephalins, endorphins are circulating μ agonist neurohormones. As such, they act not only on the μ receptors in the dorsal horn of the spinal cord, but also function on central μ receptors to alter the perception of pain and negatively modulate the signal. The discovery that TENS increases endorphin levels is significant. The effect of increasing enkephalins produces a transient effect that lasts only as long as the electrical signal is applied, as is with direct nonnociceptive stimulation as described in the gate theory. However, use of TENS produces an increase in circulating endorphins that is proportional to usage. The net effect is an analgesic effect that persists after the TENS unit is removed and increases in potency and duration with repeated usage.

The Evolution of Electroanalgesia

The use of electroanalgesia is an ancient practice, and to thoroughly understand the theory and application of TENS, it is important to understand the evolution of electroanalgesia. The powers of certain fish, namely, the Nile Catfish (*Malopterus electricus*), Torpedo Fish (*Torpedo mamorata*), and the Electric Eel (*Gymnotus electricus*) to deliver electrical shocks resulting in paralysis and temporary sensory loss in affected limbs has long been known. The Nile Catfish appeared on walls of various Egyptian tombs, dating from ~2750 bc, and represents the earliest known documentation of this phenomenon of electrical discharge. While it is not known exactly when ancient man discovered the analgesic or anesthetic properties of such fish, it is quite likely that since their discovery by primitive man, these properties were readily apparent.

Exactly when the electrical properties were used for medicinal benefit is unclear, but the earliest known writings describing this benefit were made by made in ad 46 by Scribonus Largus, a Roman physician who described the usage of the torpedo fish as a treatment for intractable gout and headache pain (19). Quoting from his treatise *Compositiones Medicae*, Largus describes these remedies.

For any type of gout a live black torpedo should, when the pain begins, be placed under the feet. The patient must stand on a moist shore washed by the sea and he should stay like this until his whole foot and leg up to the knee is numb. This takes away present pain and prevents pain from coming on if it has not already arisen. In this way Anteros, a freedman of Tiberius, was cured (20).

“Headache even if it is chronic and unbearable is taken away and remedied forever by a live torpedo placed on the spot which is in pain, until the pain ceases. As soon as the numbness has been felt the remedy should be removed lest the ability to feel be taken from the part. Moreover several torpedoes of the same kind should be prepared because the cure, that is, the torpor which is a sign of betterment, is sometimes effective only after two or three” (21).

As time progressed, the usage of electroanalgesia spread as a treatment for varying medical conditions. Pedanius Dioscorides around 80 ad describes the usage of the torpedo fish for rectal prolapse. This represents likely the first description of electrical stimulation for intentional muscular contraction (19). Likewise, these treatments were used and espoused by Galen in the second century.

The knowledge of the usage of the electrical properties of such fish was not limited to Europe. Ibn-Sidah, an Islamic physician described placing an electric catfish on someone suffering a seizure sometime in the eleventh century (21). The use of the electric fish continued to grow and by the sixteenth century the number of remedies had expanded and included treatments for various arthralgias, myalgias, headaches, epilepsy, vertigo, and for inducing sleep both by European, Indian, and Middle Eastern physicians. By the seventeenth century the application of artificially generated electricity was made possible by the development of the electrostatic generator by Otto Von Guericke.

Major revolutions in electroanalgesia came in the mid-nineteenth century from Guillaume Benjamin Amand Duchenne. He introduced the usage of moistened electrodes, as opposed to the more painful dry electrode, as a means of delivering electroanalgesia for treatment of neuropathic pain. His focus on muscle contractions from electrotherapy and making strides toward to a somewhat standardized system of electrode placement were important advances as well.

Throughout the world, electrotherapy became increasingly popular toward the end of the nineteenth and beginning of the twentieth centuries. However, with this rise in popularity came a rise in dubious to downright fraudulent applications and practitioners treating all manners of maladies from skin ailments to weight loss. With an ever increasingly savvy public, the rise of fraudulent applications, the rise of modern pharmacotherapy, X rays, and the like, electrotherapy begin to fall out of favor, or at least popularity, in the early twentieth century (19).

However, technological advances in electrical storage and delivery along with new understandings of pain have produced a resurgence in application and research in electroanalgesia. Shortly after the publishing of the gate control theory, a flurry of exciting discoveries in this

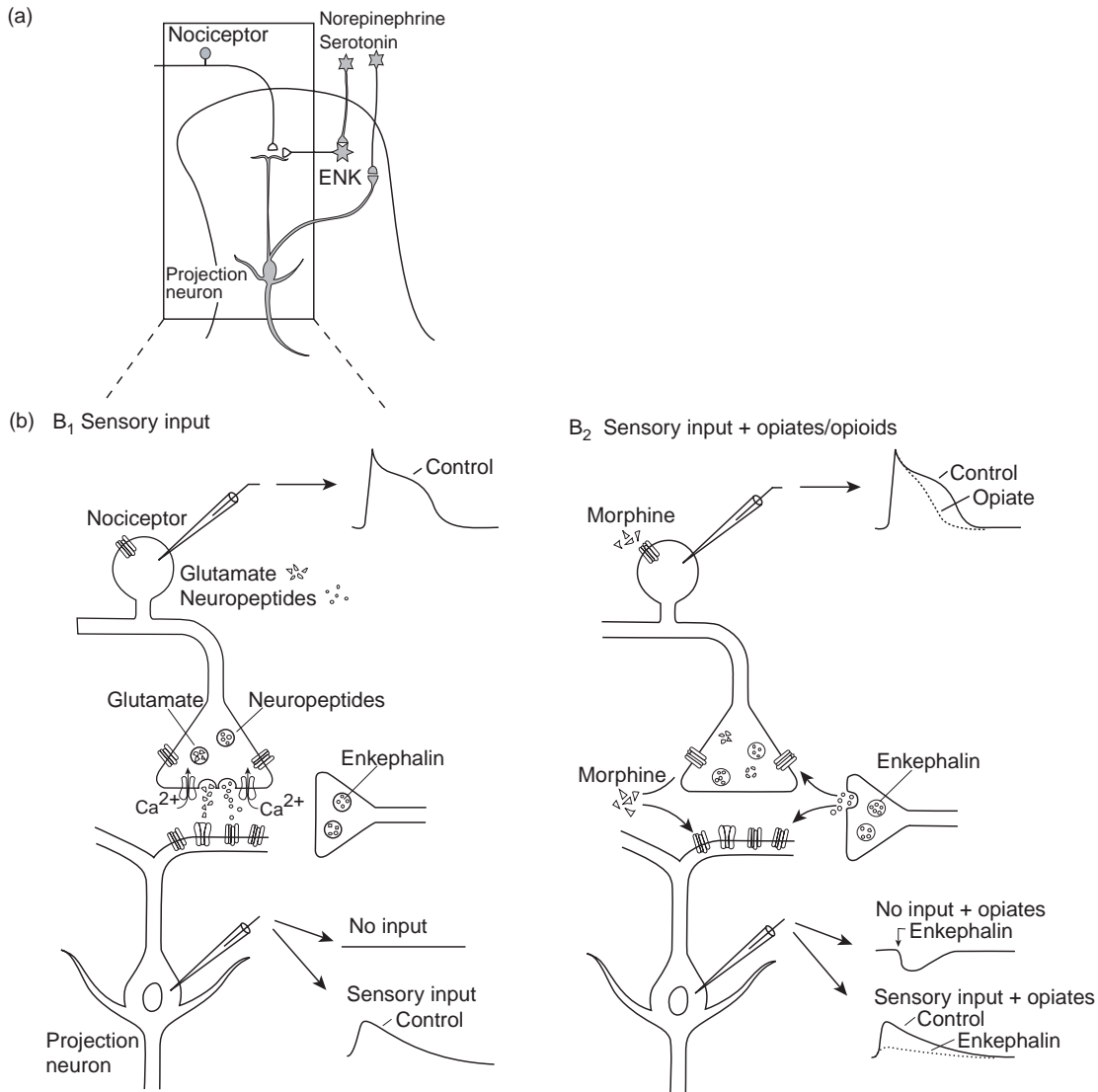


Figure 5. Local-circuit interneurons in the superficial dorsal horn of the spinal cord integrate descending and afferent pathways. (a) Possible interactions between nociceptor afferent fibers, local interneurons and descending fibers in the dorsal horn of the spinal cord. Nociceptive fibers terminate on second-order spinothalamic projection neurons. Local enkaphalin-containing interneurons (ENK) exert both presynaptic and postsynaptic inhibitory actions at these synapses. Serotonergic and noradrenergic neurons in the brain stem activate the local opioid interneurons and also suppress the activity of spinothaiamic projection neurons. (b) 1. Activation of nociceptors leads to the release of glutamate and neuropeptides from sensory terminals in the superficial dorsal horn, thus depolarizing and activating projection neurons. 2. Opiates decrease the duration of the nociceptor’s action potential, probably by decreased Ca²⁺ influx and thus decrease the release of transmitter from primary afferent terminals. In addition, opiates hyperpolarize the mambrane of the dorsal horn neurons by activating a K⁺ conductance. Stimulation of the nociceptor normally produces a fast excitatory postsynaptic potential in the dorsal horn neuron opiates decrease the amplitude of the postsynaptic potential.

realm took place starting with the 1967 demonstration by Sweet and Wall that *In vivo* electrostimulation of peripheral nerves produces analgesia as well as Shealy and Long’s work in the area of dorsal and anterior spinal cord surgically implanted stimulators. Shealy and Long discovered that peripheral nerve stimulation done in surgical candidates prior to an electrospinal implant placement produced nearly comparable analgesia to the actual dor-

sal horn implant (17)! This discovery laid the foundation for TENS development and widespread utilization today (22,23).

While somewhat beyond the scope of this article, it is worth noting that electroacupuncture experienced a widespread rediscovery in China in the 1950s. Though based on a different system of understanding of human physiology than traditional western medicine, this modality of



Figure 6. This represents one of the earliest families of stimulators, with the original model seen on the *left*. The first personal patient treatment model is depicted on the *right*, and a prototype for a miniaturized design is shown in the *center*. The original sponge electrodes are depicted in the *foreground*.

electroanalgesia is beginning to garner increasing interest interest in western medicine (19).

Transcutaneous Electrical Nerve Stimulation

Today electrotherapeutic treatment is one of the most important parts of multidisciplinary approach to treat acute and chronic pain. The TENS units themselves have undergone an evolution from large bulky units to the much smaller units of today. While there are numerous units available, each generally consists of one or more channels for electrodes, a display (either analog or digital), and various options to adjust the various parameters of the electrical current.

One of the First TENS Units Created (below)

The unit (*left*) is one of the first TENS unit available and is large and bulky with an all analog interface. Subsequent units (*center and right*) still remain analog but were more compact, though nowhere near the level of today's units (Fig. 6).

Several Modern TENS Units (below)

Pictured are just several of the numerous commercially available TENS units. Note the compact size of the models compared to older units as well as the digital TENS unit (*bottom*) (Fig. 7 and Table 2).

Electrode

The discovery that transcutaneous peripheral nerve stimulation provided nearly identical analgesic levels as dorsal root stimulation revolutionized electroanalgesia, and it almost goes without saying that the noninvasive, easy to employ nature of TENS is one of the modality's biggest assets. The electrode, the interfacing agent between the skin and machine, has undergone almost as much evolution as the TENS unit itself. The very nature of peripheral transcutaneous nerve stimulation is such that electrical currents must be applied for longer periods of



Figure 7. Several commercially available TENS units.

time in greater amounts. While the process of transferring an electrical current from machine to peripheral nerve may seem relatively simple on the outside, several notable problems present ranging from the actual transfer of the current to skin irritation to cost. Several distinct solutions currently are in use, and all present with a variety of tradeoffs (Table 3). Generally speaking, properties of a good electrode for TENS use include low cost/use ratio, good adhesion, comfortable, nonirritating to skin, good electrical conductivity, and easy to use.

Standard EKG or EEG electrodes were initially used for TENS with limited success, as these were designed for much lower amperage and much shorter usage than is needed for effective TENS. It quickly became apparent based on excessive skin irritation and poor adhesion and subsequent nonuniform current distribution that new electrode solutions were needed. One of the most popular current solutions involves silicone impregnated with carbon (Table 3). These carbon silicone electrodes provide the best cost/use ratio of the commercially available electrodes and can often last for months if properly cared for. However, a tradeoff exists in terms of convenience with these carbon silicone pads, as electroconductive gels, rich in suspended ions to facilitate the transfer of electric current from the TENS unit across the epidermis, must be applied

Table 2. Comparisons of Selected Modern TENS UNITS

Manufacturer/unit	Size, cm	Weight, g	Power source	Digital/Analog	Pulse Width, μ s	Stimulus Modes	Burst	Channels	Waveform	Output	
										Current, mA	Rate, Hz
Rehabicare/ ProMax	7.1 \times 10.1 \times 3.4	122.2	3 AAA Batteries	Digital	50–400	SD, Modulation, Normal, Burst	8 pulses/burst;	2	Asymmetric rectangular biphasic with zero net dc	0–100	2–160
Rehabicare/ Maxima3	8.4 \times 2.5 \times 6.3	121	9 V Battery	Analog	50–400	SD, Normal, Burst	8 pulses/burst; Burst at 2 Hz	2	Biphasic, asymmetrical with zero net dc	0–100	2–160
Rehabicare/ SMP-Plus	9.5 \times 6.4 \times 2.5	136.2	9 V Battery	Digital	40–300	SMP, Constant, Burst, Modulated Rate, Modulated Width, Multi- modulated	8 pulses/burst; Burst at 2 Hz	2	Biphasic, asymmetrical with zero net dc	0–60	2–125
Empi/Epix VT				Digital	0–400	ELF, Dual Pulse, High Frequency, Ramped Burst, Alternating Ramped Burst, Modulated Amp., Random Modulated, Cycle Burst, Rate Modulation, Multi modulation. Continuous, Burst, Modulated Rate, Multi-modulated	Varies	2	Balanced asymmetrical biphasic	0–60	2–150
Empi/Epix XL				Analog	0–400	Constant, Burst, Modulated Rate, Multi-modulated	Varies	2	Symmetrical biphasic square zero net charge,	0–60	2–150
BioMedical Life Systems/ BioMed 2000	9.9 \times 6.98 \times 2.54	132	9 V Battery	Analog	50–250	Constant, Burst, Width modulation,	8 pulses/burst; Burst at 2 Hz	2	Asymmetric rectangular biphasic	0–80	2–150
BioMedical Life Systems/ BioStim A6	9.9 \times 6.98 \times 2.54	132	9 V Battery	Analog	10–250	Constant, Burst, Width modulation, Rate Modulation, rate/width modulation, cycled burst	Cycled	2	Asymmetric rectangular biphasic	0–100	2–200
BioMedical Life Systems/ BioStim LX	9.5 \times 6.3 \times 3.2	226	4 AA Batteries	Digital	10–250	Constant, Burst, Burst Programmable, Width modulation, Cycled Burst	Cycled	2	Asymmetric rectangular biphasic	1–98	1–150
BioMedical Life Systems/ BioStim M7	8.2 \times 7.0 \times 4.5	266	4 AA Batteries	Digital	10–250	Constant, Burst, Burst Programmable, Width modulation, Rate Modulation, rate/ width modulation, cycled burst	Cycled	2	Asymmetric rectangular biphasic	0–98	1–200

Table 2. (Continued)

Manufacturer/unit	Size, cm	Weight, g	Power source	Digital/Analog	Pulse Width, μ s	Stimulus Modes	Burst	Channels	Waveform	Output Current, mA	Output Rate, Hz
BioMedical Life Systems/ BioStim Plus	9.9 \times 6.98 \times 2.54	132	2 AA Batteries	Digital	10–250	Constant, burst, cycled burst, Pulse rate modulation, pulse width modulation	Cycled/ 2 presets	2	Asymmetric biphasic, square wave	0–98	1–150
Vital/TENS Deluxe	9.1 \times 6.4 \times 2.3	130	9 V Battery	Analog	50–250	Burst, modulation, constant	7 pulses/burst; Burst at 2 Hz	2		variable	2–120
Kingly Star/ KS-168	40.6 \times 18.5 \times 6.8	1060.5	UM-1 \times 6 Battery or 9 V dc adapter	Analog		Constant, Burst, Modulation		4		0–100	2–200
Pain Management Technologies/ Medscope Pro	10.1 \times 5.9 \times 2.37	140	9 V Battery	Digital	50–260	Constant, Burst, pulse rate modulation, pulse width modulation	7 pulses/burst; Burst at 2 Hz	2	Biphasic square wave with zero net dc	0–80	2–150
Pain Management Technologies/ Bioscope	9.85 \times 6.05 \times 2.45	134	9 V Battery	Analog	60–250	Constant, Burst, Pulse rate and Pulse width modulation	\sim 7 pulses/burst; Burst at 2 Hz	2	Asymmetric biphasic square	0–15	2–150
ProMed Specialties/ ProM 100	9.1 \times 6.4 \times 2.4	130	9 V Battery	Analog	40–250	Constant		2	Asymmetric biphasic square with zero net current dc	0–80	2–150
ProMed Specialties/ ProM 200	9.1 \times 6.4 \times 2.3	130	9 V Battery	Analog	40–260	Burst, Constant, Modulation	9 pulses/burst; Burst at 2 Hz	2	Asymmetric biphasic square with zero net current dc	0–80	2–120
ProMed Specialties/ ProM 300	9.1 \times 6.4 \times 2.3	130	9 V Battery	Analog	40–260	Burst, Constant, Modulation	9 pulses/burst; Burst at 2 Hz	2	Asymmetric biphasic square with zero net current dc	0–80	2–50
Shining World Health Care Co./ SW 325	12.5 \times 6.6 \times 2.8	138	9 V Battery	Digital	200	Constant	n/a	2	Biphasic Spiked Wave	0–50	
Body Clock Healthcare/ Profile	10.5 \times 6.5 \times 2.75	100	2 AA Batteries	Digital	25–250	Constant, Burst, Width modulation, Rate modulation	Variable	2	Symmetric biphasic rectangular	0–100	1–200
Med-Dyne/TA3	9.1 \times 6.4 \times 2.3	130	9 V Battery	Analog	40–260	Constant, Burst, Modulation	9 pulses/burst; burst at 2 Hz	2	Asymmetric biphasic square with zero net current dc	0–80	2–160

Table 3. Basic Types of Electrodes

General electrode Type	Number of Uses	Typical Retail Cost Electrode (pair)	Typical features			
			Adhesion and Conduction	Composition Materials	Advantages	Disadvantage
Disposable	1 use	Under \$3.00	Pressure-sensitive tape surrounding conductive area (wet get in sponge)	Nonwoven Foam	Easiest to use Very good adhesion Comfort	Cost/use Skin irritation Poor electrical performance
Semireusable	3–10 uses	\$5.00–10.00	Conductive adhesive over entire surface	Foam Plastic film Carbon silicone	Ease of use Low skin irritation Comfort	Weak adhesion Care and storage Medium cost per use Good electrical performance
Reusable	>100 uses	\$4.00 ^a	Requires addition of conductive gel and tape adhesion	Carbon silicone	Very good electrical performance (if applied properly) Lowest cost per use	Most preparation to use Skin irritation Messy Requires gel and tape Skill required for optimal performance Poor adhesion Not as flexible in use

prior to each usage. The electrodes must then be affixed with tape to the skin. This process can be laborious if not impossible for the end user to do, depending on electrode location as well as physical disability. While numerous medical tapes exist, care must be taken in their selection. Some users display mild-to-severe allergic reactions to adhesives in various tape products. Likewise, certain adhesive tapes adhere too firmly to the skin and can result in injury with repeated application and removal.

In applications where cost is no object, sterility is needed, or convenience must be maximized, single use adhesive electrodes are used that consist of thin, porous material impregnated with adhesive electroconductive gel covered with cloth or plastic on one side to prevent adhesion to clothing or electrical current escape. These electrodes are used extensively in hospital or rehab facilities where numerous patients are seen and reusing electrodes is neither feasible nor sanitary as well as in individuals who desire or require maximum convenience.

A third option in electrode selection includes so-called “dry gel” electrodes manufactured from a conductive polysaccharide gum, called Karaya or Sterculia gum when taken from the Sterculia Urens tree native to India or a manufactured comparable material. These dry gel electrodes represent a good compromise between the disposable and reusable carbon silicone electrodes, as they are self adhesive, do not require electroconductive gel application, may be reused several times, and represent a significantly lower cost/use ratio than disposable pads.

Patients should be informed of the various pros and cons of the various electrodes as well as counseled regarding proper usage. According to Szeto, several factors should be considered when selecting the proper electrode:

1. How long will each application of the electrodes last, and therefore what is the level of adhesion needed?

2. Is the stimulation site readily accessible or will there be someone to assist? How simple must the application of the electrodes be?
3. What is the patient’s skin type (hairiness, oiliness, hyperallergic)? Will special pregelled electrodes be required?
4. Where is the painful are? This factor will help to determine the best electrode size, shape, cosmesis, and number.
5. Does the TENS user lead an active life? If so, a high performance electrode in terms of adhesiveness, pliability, and nonirritability would be needed.
6. Can the user take good care of the electrodes, and what are the financial arrangements? These issues will affect the cost-effectiveness of disposable or semireusable type of electrode (24).

Electrode size is another factor to consider in selection and depends on the location of pain, area required for stimulation, and personal preference. Numerous sizes and shapes of electrode pads are available and suitable for virtually any application.

Sample Electrodes

Electrode Placement

Electrode placement is crucial in maximizing positive outcome with TENS units. Most units employ two or more channels of current, which splits to two electrodes, and it is often advisable for multiple channels to be used to cover maximal areas, as many pain syndromes often do not exhibit pain localized to a specific point source. Numerous books on TENS or manufacturer information as well as various anatomical charts provide users with possible electrode placements. While it is impossible to describe

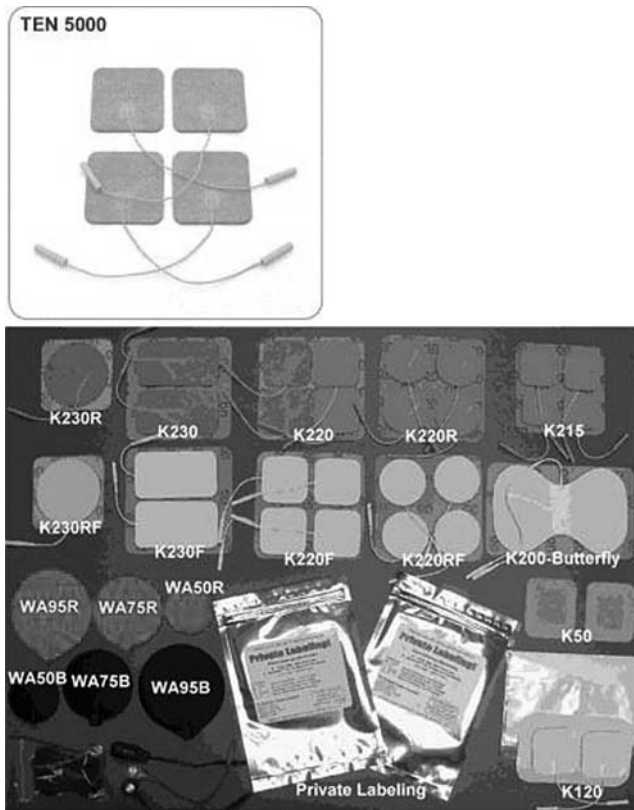


Figure 8. TENS electrodes.

proper electrode placement for every pain syndrome, certain electrode arrangements are frequently used.

For the purposes of this discussion, the channels will be referred to as I and II and the negative electrode as “a” and the positive as “b”. Parallel placement with one channel is utilized for relatively localized areas of pain, such as point pain or pain from a surgical incision. Electrode Ia is placed on one side of the incision, while Ib is on the other, producing a current that flows between the two with the area of pain in between the electrodes. Bilateral placement is similar, but generally defined as meaning Ia and Ib electrodes are placed on either side of the spine, symmetrically and close together, useful for localized, nonradiating neck or back pain. For radiating neck or back pain, a longitudinal arrangement is often utilized in which electrodes of one channel are on the same side of the spine and placed along the pain pathway. If the pain is bilateral, electrodes of another channel on the can be placed on the opposite side of the spine along the pain pathway (Fig. 9).

A crossed, or interferential, placement is useful for pain localized to large joints, such as shoulder, elbow, or knee. In this pattern, Ia and IIa are placed side by side with IIb below Ia and Ib below IIa, creating a square pattern with electrodes of the same channels diagonally opposite each other with the area of pain in the center of the square. Bracketed placement is generally reserved for treating the dermatomal neuralgia that frequently is associated with shingles, varicella zoster, out breaks. In this arrangement, electrodes Ia and Ib are placed along

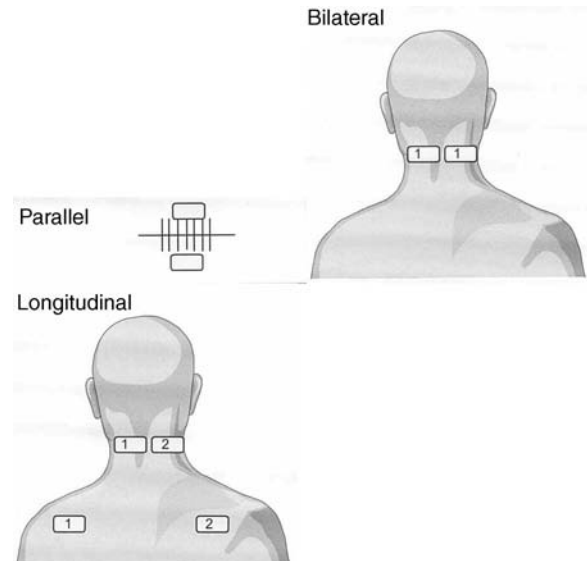


Figure 9. Bilateral pain electrode placement.

the dermatome cranial to the neuralgia and electrodes IIa and IIb placed along the dermatome caudal to the neuralgia (Fig. 10).

Occasionally, localized pain is so severe that the user cannot tolerate electrode placement over the affected site, and in this case a contralateral placement in the nonaffected hemisphere over the same anatomical area as on the affected side is utilized. This arrangement will sometimes permit sufficient pain relief for the user to eventually tolerate direct stimulation. While the exact mechanism of analgesia is not known, it is hypothesized the analgesic effect is the result of central inhibitory pathways (24). Certain syndromes, such as Reflex Sympathetic Dystrophy, lend themselves to this placement, and reflex vasodilatory effects may explain contralateral analgesia in these syndromes (25) (Fig. 11).

Certain pain syndromes, such as phantom limb pain, glove–stocking distribution peripheral neuropathy, or acute burns, fractures, lacerations, or other injuries of the hands or feet lend themselves to a placement of the electrodes proximal to the actual source of the pain. In this placement, the electrodes of one channel are simply placed along the dermatome of the pain source, but proximal to the pain (Fig. 12).

The final placement method to be covered is a linear, unilateral, overlapping pattern useful for pain along some, as in myofascial pain, or all, as in radicular pain, of an extremity, and follows a placement procedure outlined by Wolfe (25). After the dermatomal distribution of the pain is elucidated, electrode Ia is placed at the most proximal location where the user experiences pain. Distal to this electrode, the user identifies the site of maximal pain and places IIa here. At the most distal site of pain, electrode IIb is placed, and between IIa and IIb, electrode Ib is placed, being careful to keep all electrodes in the affected dermatome (Fig. 13).

It is important to note with the above placement, the electrical current covers the entire length of the pain the

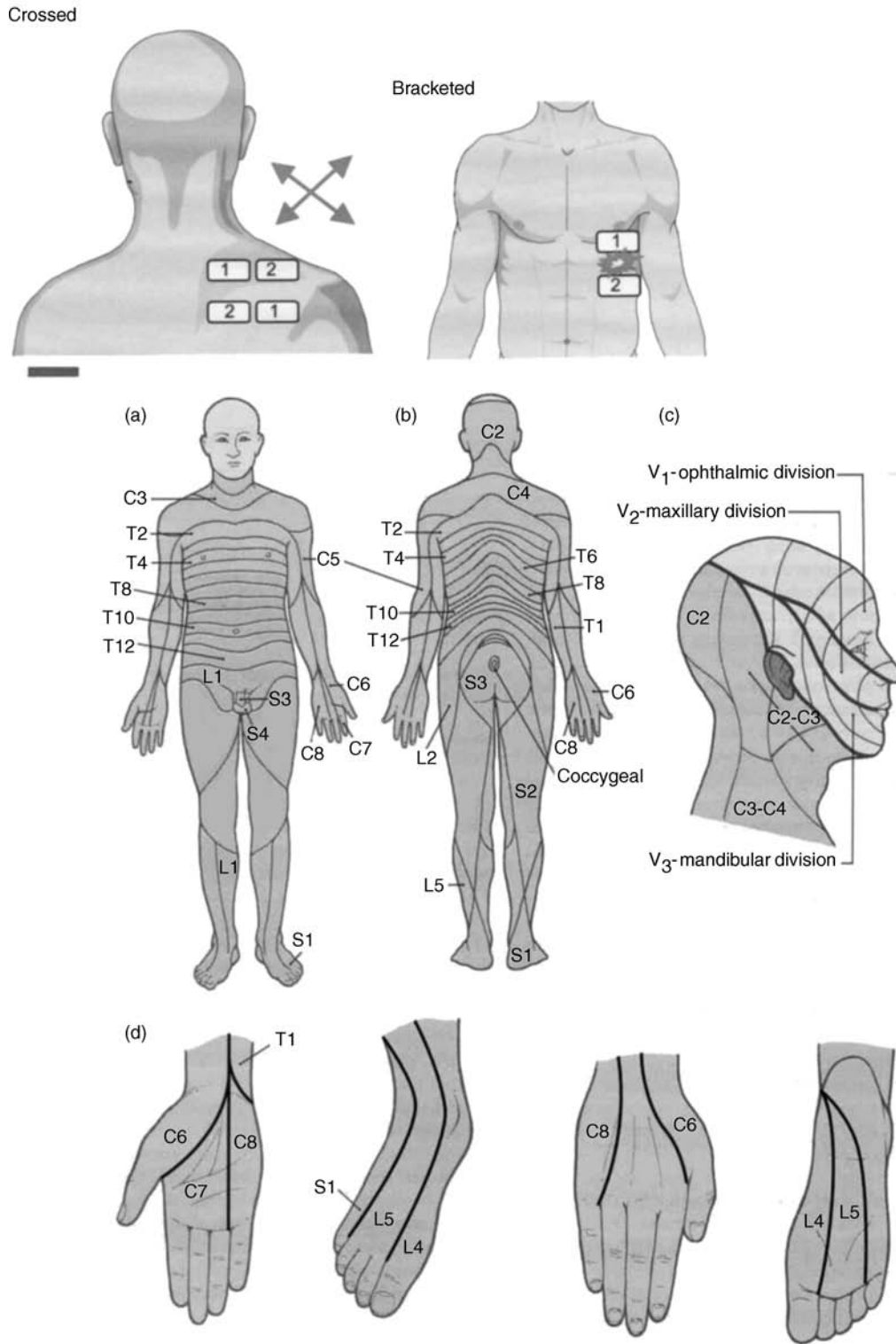


Figure 10. Dermatome maps of the peripheral distribution of spinal nerves (a and b) and trigeminal nerve (c) d. Details of dermatome maps on anterior and posterior surfaces of the hand and foot.

user experiences. If the user inadvertently places electrodes in a nonoverlapping pattern (i.e., Ia and Ib both proximal to IIa and IIb), the current will not cover the entire pain pathway; instead it will only travel between electro-

des of the same channel, leaving the area between Ib and IIa “uncovered”. The following should be generally avoided due to poor current coverage area: unilateral, linear (Fig. 14).

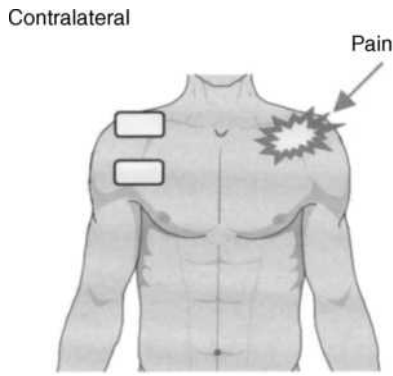


Figure 11. Contralateral electrode placement.

Electric Amplitude–Frequency Selection

Once the proper electrode type is chosen, as well as optimal electrode placement ascertained, the optimal electrical signal must be selected. Generally speaking, the most used currents include “classical” TENS with high intensity–low frequency currents for up to 12 h at a time, low intensity–high frequency currents for up to 45 min several times a day, and intermittent low frequency bursts are used. The varying current intensity–frequency produces analgesia via the different mechanisms as discussed above.

Classical TENS employs high frequency (60–200 pulses per second)/low intensity (15–60 mA) stimulation and produces a distinct “electrical tingling” sensation in the area of electrode pad placement that most users find pleasant. This current is not of significant intensity to produce significant muscular contraction. Pain relief from this form of stimulation is transient, occurring quickly once stimulus is applied and disappearing once current is removed, and the gate control theory likely explains the mechanism of analgesia. The high frequency pulses activate Aβ sensory afferent fibers and inhibit pain transmission in the dorsal column of the spinal cord.

Low frequency (1–5 pulses per s)/high intensity stimulation, producing sustained muscle contractions, results in slower onset pain relief that persists after the stimulus is removed. Numerous studies have demonstrated partial to near total inhibition of analgesic effect via administration of naloxone (13–15). The endorphin and enkephalin theories described previously likely largely account for the mechanism of analgesic activity, especially the endorphin

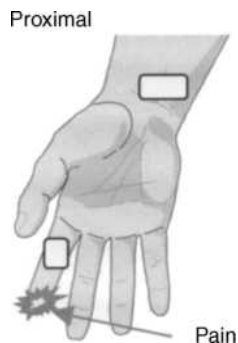


Figure 12. Proximal electrode placement.

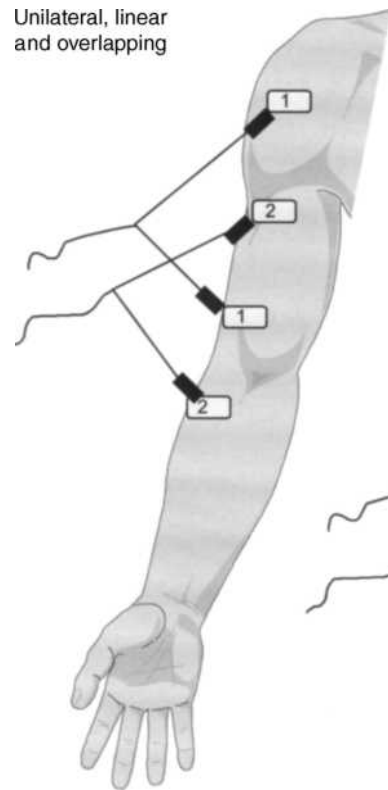


Figure 13. Unilateral, linear, and overlapping electrode placement.

theory and long-term analgesia. While effective at inducing long-term analgesic effects, the low frequency–high intensity method of stimulation is often perceived by many patients as less pleasant than high frequency stimulation.

As the long-term effects of low frequency intense stimulation are desirable, manufacturers have devised means of producing a more pleasurable sensation while at the same time stimulating muscle contraction enough for long-term analgesia via modulation of the current. To understand the modulation of current in TENS, a brief review of the current waveforms it employs is needed. Briefly, biphasic waveforms, consisting of both a positive and negative phase are used, and these may be either symmetric or

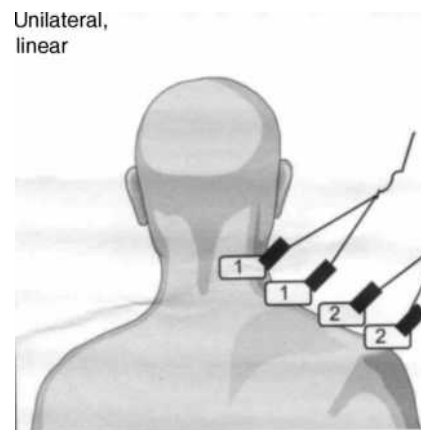


Figure 14. Unilateral, linear electrode placement.

asymmetric. If the current amplitude is equally positive and negative, the current is termed “electrically balanced,” also referred to as zero net charge (znc) or no net dc current. While both balanced and unbalanced electrical currents are employed, unbalanced current transmission can result in pH changes in the skin with long-term usage, do to ion exchange, which can result in skin irritation. Additionally, the current employed in TENS is a pulsatile current with interspaced periods of electrical activity and electrical silence. The periods of electrical silence may be either uniform or varying. The frequency of electrical pulses may be given in units of hertz (Hz), cycles per second (cps), or pulses per second (pps) (25). It is important to note that while frequency may be a constant 100 cps, the period between the pulses may be variable.

All of these variables in the current waveform may be adjusted to achieve a net effect that is both pleasant to the patient while sufficient to achieve muscular contraction. For example, the amplitude of the current may be varied over a constant time interval, the duration of the pulse may be varied, the time between pulses may be varied, or a combination of some or all of the previous may be used. As pain severity and character can frequently change, models that allow modulation of electrical current via one or more characteristic offer distinct advantages in patients’ individualizing their therapy as well as help prevent physiologic adaptation to the electrical stimulation. While numerous studies have delved into optimizing the electrical waveform (26–28), their conclusions have been varied, and it is likely, there is no optimal waveform. As such, TENS therapy is an individualized one, and patients should have frequent follow ups with their physician to ensure the patient is receiving optimal care for their specific complaint (Table 4).

TREATMENT PLANS

Treatment with TENS is an extremely variable and personalized process, and this cannot be underscored enough. It is vital for close healthcare supervision for a user to obtain the maximum therapeutic benefit from tens. TENS may only be used in an acute phase for a short period of time (e.g., incisional pain postsurgery) or for months or years (e.g., those suffering from chronic back pain). For chronic pain sufferers classical TENS may be used for several hours continuously daily. Modulated or low frequency/high intensity may be used for ~30 min three times a day for an indefinite period of time. When using TENS it is important to use as strong or nearly asstrong a current as the user can tolerate to achieve best results.

Fibromyalgia is a poorly understood chronic pain condition that presents unique management challenges not only because it is poorly understood, but also because it is often refractory to traditional treatment modalities. A recent double-blinded study by Cork et al. (29) explored cranial electrotherapy stimulation (CES) as a possible treatment for fibromyalgia. In this study, using electrodes clipped to the participants’ ear lobes, the Alpha-Stim CES device, delivered either modified square-wave stimulation at 100 μ A and a 50% duty cycle at 0.5 Hz for 1 h daily for 3 weeks or sham therapy (see Fig. 15). While there were no

Table 4. INDICATIONS for Use of TENS

<i>Systemic Pain</i>	
Bursitis	Phantom limb syndrome
Cancer	Raynaud’s syndrome
Causalgia	Rheumatoid arthritis
Multiple sclerosis	Synovitis
Neuralgia	Diabetic peripheral
Osteoarthritis	Neuropathy
Fibromyalgia	
<i>Head and Neck Pain</i>	
Cluster headaches	Suboccipital headaches
Dental disorders	TMJ Syndrome
Migraine headaches	Torticollis
Spondylosis	Trigeminal neuralgia
Sprains/strains	Whiplash
<i>Abdominal Pain</i>	
Diverticulosis	Labor
Dysmenorrhea	Postoperative pain
<i>Back Pain</i>	
Facet syndrome	
Intercostals neuralgia	Radiculitis
IVD Syndrome	Sprains/strains
Lumbago	Thoracodynia
Lumbosacral pain	Whole back pain
<i>Lower Extremity Pain</i>	
Ankle pain	Passive stretch pain
Foot pain	Sciatica
Fractures	sprains/strains
Ischialgia	tendonitis
Knee pain	Thrombophlebitis
<i>Upper Extremity Pain</i>	
Epicondylitis	
Frozen shoulder	Sprains/strains
Hand pain	Subdeltoid bursitis
Peripheral nerve	Wrist pain
Injury	



Figure 15. The Alpha-Stim CES device.



Figure 16. Electrode placement for CES.

differences in baseline pain scores of the participants in either group prior to beginning the study, after 3 weeks of CES therapy those in the treatment group displayed significantly lower Pain Intensity Scores, Tenderpoint Scores, and POMS Scores compared to the sham group. After 3 weeks the study was unblinded, and 23 of the 35 participants in the Sham group elected to switch over to active treatment for 3 weeks. Those switching from sham therapy to active CES therapy displayed significant reductions in the aforementioned pain scores as well (29) (Fig. 16).

WARNINGS AND CONTRAINDICATIONS

TENS is contraindicated in individuals with pacemakers, especially those with demand-type pacemakers as the electrical stimulation could cause misfiring or other malfunction of the pacemaker. Electrode placement in areas of sensory or circulatory deficits should be avoided due to the potential for burns or excessive muscular contraction. Electrodes should not be placed over the carotid sinuses to prevent a vasovagal reflex reaction with resultant hypotension. Electrodes should not be placed over the anterior neck due to potential to induce laryngospasm and subsequent asphyxiation. Electrodes should not be placed over the eyes. TENS should be avoided in pregnant women due to the potential to induce contractions and premature labor. Caution should be used in patients with implanted spinal stimulators as well as intrathecal opiate pumps. The unit should not be used with other electrical medical equipment, such as ECGs, EEGs, pulse oximeters, and electrocautery devices.

PRECAUTIONS

Tens has not been proven to have curative value and should be used only under the supervision of a physician. Patient selection is crucial, and not all patients will respond to TENS. The TENS has not been shown to exhibit curative value and should not be used for pain of unknown origin. The unit itself as well as wires and electrodes should be kept out of reach of children and water.

BIBLIOGRAPHY

1. Medical Data International, Market and Technology Reports, U.S. Markets For Pain Management Products, June 1999 Report RP-821922.
2. Bone M, Critchley P, Buggy DJ. Gabapentin in postamputation phantom limb pain: A randomized, double-blind, placebo-controlled, cross-over study. *Reg Anesth Pain Med* 2002 Sep-Oct; 27(5):481-486.
3. Zuurmond WW, van der Zande AH, de Lange JJ. Phantom pain following leg amputation: Retrospective study of incidence, therapy and the effect of preoperative analgesia. *Ned Tijdschr Geneesk* 1996 May 18; 140(20):1080-1083.
4. Fields, HL. *Pain*. New York: McGraw-Hill, 1987.
5. Jessell T, Kandel E, Schwartz J, editors. *Principles of Neural Science*. New York: McGraw-Hill; 2000 p 472-492.
6. Haines D, editor. *Fundamental Neuroscience*. Philadelphia: Churchill Livingstone; 2002. p 273-292.
7. Willis, WD Jr. *The Pain System: The Neural Basis of Nociceptive Transmission in the Mammalian Nervous System*. New York: Karger, 1985;
8. Melzack R. *The Puzzle of Pain*. New York: Basic Books; 1973. p 55-56.
9. Zborowski M. Cultural components in responses to pain. *J Soc Issues* 1952;8(4):16-30.
10. Melzack R, Wall PD. Pain mechanisms: A New theory *Science* 1965;150:971-979.
11. Luo F. A study on the cumulative effect of repeated electroacupuncture on chronic pain. *Sheng Li Ke Xue Jin Zhan* 1996 Jul; 27(3):241-244.
12. Han JS, et al. Effect of low-and high-frequency TENS on Met-enkephalin-Arg-Phe and dynorphin A immunoreactivity in human lumbar CSF. *Pain* 1991 Dec; 47(3):295-298.
13. Kalra A, Urban MO, Sluka KA. Blockade of opioid receptors in rostral ventral medulla prevents antihyperalgesia produced by transcutaneous electrical nerve stimulation (TENS). *J Pharmacol Exp Ther* 2001 Jul; 298(1):257-263.
14. Sluka KA, et al. Spinal blockade of opioid receptors prevents the analgesia produced by TENS in arthritic rats. *J Pharmacol Exp Ther* 1999 May; 289(2):840-846.
15. Han JS, Chen XH, Yuan Y, Yan SC. Transcutaneous electrical nerve stimulation for treatment of spinal spasticity. *Chin Med J (Engl)* 1994 Jan; 107(1):6-11.
16. Rodriguez E, et al. Effects of transcutaneous nerve stimulation on the plasma and CSF concentrations of beta-endorphin and the plasma concentrations of ACTH, cortisol and prolactin in hysterectomized women with postoperative pain. *Rev Esp Anestesiol Reanim* 1992 Jan-Feb; 39(1):6-9.
17. Liss S, Liss B. Physiological and therapeutic effects of high frequency electrical pulses. *Integr Physiol Behav Sci* 1996 Apr-June; 31(2):88-95.
18. Kho HG, Kloppenborg PW, van Egmond J. Effects of acupuncture and transcutaneous stimulation analgesia on plasma hormone levels during and after major abdominal surgery. *Eur J Anaesthesiol* 1993 May; 10(3):197-208.
19. Gordon G. *Electroanalgesia: Historical and Contemporary Developments*. 1998 Available at URL:<http://freespace.virgin.net/joseph.gadsby/index.htm>.
20. Schechter DS. Origins of electrotherapy. *N Y State J Med* 1971 May 1; 71(9):997-1008.
21. Kellaway P. The William Osler Medal Essay: The part played by electric fish in the early history of bioelectricity and electrotherapy. *Bull Hist Med* 1946;20:112-137.
22. Long DM. Fifteen Years of transcutaneous electrical stimulation for pain control. *Stereotact Funct Neurosurg* 1991;56(1):2-19.
23. Hymes A. Introduction: A review of the historical uses of electricity. In: Mannheim JS, Lampe G, editors. *Clinical*

- Transcutaneous Electrical Nerve Stimulation. Philadelphia: Davis Company; 1984 p 1–5.
24. Szeto A. Pain relief using transcutaneous nerve stimulation. In: Webster JG, editor, Encyclopedia of Medical Devices and Instrumentation. Vol 4 New York: John Wiley & Sons; 1988. p 2203–2220.
 25. Wolfe P. A Practical Approach to Transcutaneous Electrical Nerve Stimulation (TENS). New Brighton: Rehabicare, Inc.
 26. Barr JO, Nielsen DH, Soderberg GL. Transcutaneous electrical nerve stimulation characteristics for altering pain perception. *Phys Ther* 1986 Oct; 66(10):1515–1521.
 27. Katims JJ, Long DM, Ng LK. Transcutaneous nerve stimulation. Frequency and waveform specificity in humans. *Appl Neurophysiol* 1986;49(1–2):86–91.
 28. Repperger DW, et al. Microprocessor based spatial TENS (transcutaneous electric nerve stimulator) designed with waveform optimality for clinical evaluation in a pain study. *Comput Biol Med* 1997 Nov; 27(6):493–505.
 29. Cork R, et al. The Effect of Cranial Electrotherapy Stimulation (CES) on Pain Associated with Fibromyalgia. *Int J Anesthesiol* 2004;8(2). Available at <http://www.ispub.com/ostia/index.php?xmlFilePath=journals/ija/vol8n2/ces.xml>.

See also **BLADDER DYSFUNCTION, NEUROSTIMULATION OF; ELECTROANALGESIA, SYSTEMIC; ELECTROPHYSIOLOGY; FUNCTIONAL ELECTRICAL STIMULATION.**

TRANSPLANTATION, LIVER. See **LIVER TRANSPLANTATION.**

TRAUMA MANAGEMENT. See **CARDIOPULMONARY RESUSCITATION.**

TWEEZERS, OPTICAL. See **OPTICAL TWEEZERS.**

U

ULTRASONIC HYPERTHERMIA. See
HYPERTHERMIA, ULTRASONIC.

ULTRASONIC IMAGING

OLIVER KRIPFGANS
University of Michigan
Ann Arbor, Michigan

INTRODUCTION

Medical imaging has many modalities and most of them provide clinicians with unique features of a volume of interest (VOI) resulting from a chosen modality. Ultrasonic imaging is one technique for collecting anatomical and physiological information from within the human body. It can be used for diagnosis (imaging) and for image-guided therapy, where therapeutic intervention can be applied with direct image-based feedback. Other modalities include X ray (roentgen radiation), CT (computed tomography), MRI (magnetic resonance imaging), PET or PET/CT (positron emission tomography), and SPECT (single photon emission computed tomography). In contrast to most other imaging techniques, ultrasonic imaging is very attractive to professionals because it is cheap, real time (with >100 full frame images per second, >100 Hz), and it uses nonionizing radiation. Moreover, current clinical ultrasound machines can be integrated into laptop computers with very little external hardware for maximum portability and versatility. These combined features allow the use of ultrasonic imaging in a wide variety of settings, from private physician practices, to ambulances with on-site paramedics, to battle field situations, where very robust and lightweight equipment is required. Many other uses of ultrasonic imaging are found in science and industry these include, for example, ultrasonic microscopy, nondestructive testing and touch sensitive screens.

PHYSICAL PRINCIPLES

Ultrasonic imaging is based on ultrasound, which is sound produced at frequencies beyond those detectable in human hearing, that is, >20 kHz. In the same way that ultraviolet (UV) light is invisible to the human eye, ultrasound is inaudible to the human ear. Often objects that serve as carriers for ultrasound waves need to be treated as waveguides. Nonlinear effects become apparent for ultrasound propagation when leaving the range of elastic deformation during the propagation of waves through a medium. Physical material constants form ultrasound parameters, for example the speed of sound or the attenuation of sound. Very high frequency sound waves are treated by quantum acoustic laws. Historically, ultrasound was produced by oscillating platelets (1830), or pipes (1876).

Magnetostriction (1847) and the piezoelectric effect followed (1880 by Curie), and are still very much relevant mechanisms for medical and industrial ultrasound. In 1918, it was found that the use of oscillating crystals could be used to stabilize frequencies. The upper frequency for sound in a given solid material is determined by the separation of neighboring atoms in the host medium. This upper frequency limit is met when neighboring atoms, assuming the linear chain model, oscillate with a 180° phase shift, the so-called optical branch of oscillations in a solid (1–15).

Sound Waves

Sound waves are mechanical waves by nature and a medium is needed to carry them. These spatial-temporal oscillations occur nonsynchronously throughout a medium and cause density fluctuations, which in turn cause temperature oscillations if the rate of such fluctuations is larger than the time constant for thermal equalization. Typical properties to describe sound waves are

Displacement $\mathbf{s} = \mathbf{s}(t)$ of a particle due to a traversing wave

Sound particle velocity $\frac{d}{dt}\mathbf{s} = \frac{d}{dt}\mathbf{s}(t) = \mathbf{v}(t)$

Instantaneous mass density $\rho = \rho(t)$

Instantaneous pressure $p = p(t)$

where the later most is the deviation from the ambient static pressure. Mechanical properties, such as strain tensor σ_{ij} and stress tensor s_{ij} can be used to develop relationships between the above mentioned properties of sound waves (Fig. 1).

For mostly lossless media, such as water, one can use the laws of conservation of momentum and conservation of mass to derive the wave equation for sound waves,

$$\nabla p(r) = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} p(r) \quad \frac{B}{A} = \text{const} \left[\frac{\partial c}{\partial p} \right]_T + \text{const} \left[\frac{\partial c}{\partial T} \right]_p \quad (1)$$

where spatial variations in pressure ∇p are coupled to temporal variations $(\partial p)^2/(\partial t^2)$ via the speed of sound c . This simple relationship represents only a linear approximation, and is therefore valid only for waves of small displacements. The isentropic nonlinearity parameter B/A measures the amount of density change (ρ , $c = (\rho K)^{-1/2}$ (see Eq. 10) for a given pressure (p) and temperature (T) (6). Note that only second-order temporal derivatives in the pressure result in a spatial pressure change, that is, only accelerations can result in sound.

Planar and Spherical Waves. In general, mechanical waves can either travel as longitudinal waves by compressing and expanding the host medium itself or as transversal waves by exerting shear force on the host medium. Water is a very good host medium because it bears minimal

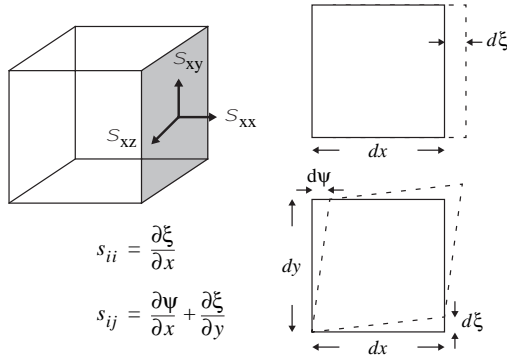


Figure 1. Ultrasonic waves travel through media by elastic deformation of matter. This deformation can be written in terms of strain σ_{ij} and stress s_{ij} tensors.

energy loss for traveling sound waves. The human body is 55–60% (8) composed of water, which ensures good acoustic accessibility. Exceptions are areas blocked by either bone or air, since bone and air provide poor ultrasound transmission characteristics. However, water mostly supports longitudinal waves. Transverse waves are attenuated at a high rate in water and can therefore be neglected. The three-dimensional (3D) wave equation in Eq. 1 reduces then to a one-dimensional (1D) equation with the general solution of an inward and outward propagating wave:

$$\left. \begin{aligned} p(x, t) &= F(x - ct) + G(x + ct) \\ p(r, t) &= \frac{1}{r}(F(r - ct) + G(r + ct)) \end{aligned} \right\} \begin{array}{l} \text{for Planar and} \\ \text{spherical waves} \end{array} \quad (2)$$

where x and r are the direction of propagation, t is time, and F and G are general, but continuously differentiable functions. The rationale for these arbitrary functions is their argument $x \pm ct$ and $r \pm ct$. This expression ensures the character of the wave as a traveling entity. Whatever the function F represents at time $t = 0$ at position $x = 0$, it will travel to position x/c after time t . In other words, by keeping the argument of the function $F = 0$, one can compute where and how fast the wave travels. Vice versa for the function G , except that it travels in the opposite direction (Fig. 2). For harmonic waves, these two functions are represented by harmonic functions, that is, sin, cos, or more general e^{ix} . Planar and spherical waves follow as:

$$\left. \begin{aligned} p(x, t) &= pe^{ik(x \pm ct)} && \text{planar waves} \\ p(r, t) &= \frac{p}{4\pi r} e^{ik(r \pm ct)} && \text{spherical waves} \end{aligned} \right\} \quad (3)$$

Here, k is the wave number, defined as $2\pi/\lambda$, which is the conversion between spatial coordinates of wavelength to

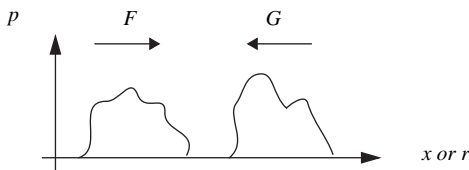


Figure 2. Illustration for the general solution of equation 1; two traveling waves F and G in opposing directions, $+r$ and $-r$, respectively.

radians in the complex exponential. Acoustic attenuation or absorption can be mathematically written as a imaginary valued wave number k_i , leading to the complex valued total wave number $k = k_r + k_i$.

Quantifying Sound

Sound intensity $I(\text{W} \cdot \text{cm}^{-2})$ and acoustical power P [W]

$$\mathbf{I} = \overline{p\mathbf{v}}, \quad P = \oint_S \overline{p\mathbf{v}_n} dS \quad (4)$$

are measures of the strength of the acoustic waves. In both equations, the temporal average of the product of pressure and particle velocity is computed. In the equation for intensity in equation 4, the temporal average value is a vector quantity, while in the equation for power it is a scalar because the velocity is taken as the normal component to the encapsulating surface element dS . In addition to acoustic intensity and power, very often one refers to a measure for the acoustic pressure. In SI units, pressure is given by Newton per square meter ($\text{N} \cdot \text{m}^{-2}$). However, sound pressures extend over a large range, and therefore a logarithmic scale, the dB scale, is commonly used to measure pressure.

$$\text{dB value} = \begin{cases} 20 \cdot \log p_{\text{rms}}/p_{\text{ref}} \\ 10 \cdot \log I_{\text{rms}}/I_{\text{ref}} \end{cases} \quad (5)$$

As can be seen from Eq. 5, a reference value must be used to compute the pressure level on a dB scale. Typically, this reference is chosen to be the peak output level of the system under test or it can be a fixed value such as when 1 mW into 50Ω is used on some oscilloscopes. Sound pressure level (SPL) and sound intensity level (SIL) reference values in underwater ultrasonics are $1 \mu\text{Pa}$ and $10^{-12} \text{W} \cdot \text{m}^{-2}$, respectively. In contrast, SPL for air bourne sound is $20 \mu\text{Pa}$, whereas SIL remains at $10^{-12} \text{W} \cdot \text{m}^{-2}$.

Acoustic Impedance

Impedance is a term commonly known from electromagnetism. However, it also applies to sound waves and relates sound pressure and particle velocity in a manner analogous to Ohm's law. One distinguishes at least four types of acoustic impedance: specific acoustic impedance (z) is used to compute the transmission of an acoustic wave from one medium into another; acoustic impedance (Z) is used to estimate the radiation of sound from vibrating surfaces; mechanical impedance (Z_m) is the ratio of a complex driving force and the resulting complex speed of the medium; and radiation impedance (Z_r), which is used to couple acoustic waves to a driving source or a load driven by a force.

$$\begin{aligned} \mathbf{z} &= \mathbf{p}/\mathbf{v} && \text{Specific acoustic impedance} \\ \mathbf{Z} &= \mathbf{p}/\mathbf{U} = \mathbf{z}/S && \text{Acoustic impedance} \\ Z_m &= \mathbf{f}/\mathbf{U} && \text{Mechanical impedance} \\ Z_r &= Z/S = \int d\mathbf{f}s/\mathbf{v} && \text{Radiation impedance} \end{aligned}$$

Here, \mathbf{p} is the acoustic pressure as a function of space and time, \mathbf{v} is particle displacement velocity as a function of

space and time, \mathbf{U} is a volume velocity, \mathbf{S} is the surface that emits the sound, \mathbf{f} is a complex driving force of the sound source, such as the force of a coil that makes the membrane of a loudspeaker move, and subsequently \mathbf{u} is the complex speed at which the forced medium is moving.

Moreover, a quantity called characteristic impedance is analogous to the wave impedance $\sqrt{\mu/\epsilon}$ of a dielectric medium. Its analytical form depends on the type of wave. Equation 7 shows the closed form expression for planar and spherical waves. It should be noted that the characteristic impedance for spherical waves can be complex valued and therefore pressure (p) and velocity (v) are not required to be in phase.

characteristic impedance:

$$z = \frac{p}{v} = \rho_0 \cdot c \quad \text{For planar waves}$$

$$z = \frac{p}{v} = \rho_0 c \cdot \frac{kr}{(1 + (kr)^2)^{1/2}} e^{j\theta} \quad \text{For spherical waves} \quad (7)$$

For these two special cases one can see that planar waves have pressure (p) and particle velocity (v) in phase. The ratio of pressure to velocity is a real number and is constant ($\rho_0 c$). Spherical waves, however, behave differently. Pressure and particle velocity are out of phase when measured close to the sound source. The ratio of pressure to velocity is a complex value (due to the $e^{j\theta}$ term and $\cot \theta = kr$) and it changes with distance (r). For large r , the spherical wave solution approaches the planar wave solution, that is, when $kr \gg 1$.

The parameter Z , the acoustic impedance, is often referred to as a frequency independent constant. The importance of this property lies in the nature of ultrasound. When imaging the human body, the sound waves travel through many layers of varying impedances, such as skin, fat, connective tissues, and organs. The crossing of each tissue boundary changes the sound waves in several ways. Typically, sound both transmits and reflects from tissue layers. While sound is mostly transmitted, small reflections are recorded and displayed as gray levels in a so-called B-mode image, where larger amplitudes of reflected waves is displayed as brighter pixels. More complicated scenarios include mode conversion between longitudinal and transverse waves. Reflection and transmission coefficients for pressure are directly related to the change in acoustic impedance as given in the following equation:

$$R = \frac{Z_2 - Z_1}{Z_2 + Z_1} \quad T = \frac{2Z_2}{Z_2 + Z_1} \quad 1 + R = T \quad (8)$$

Here Z_1 and Z_2 are the impedances of the proximal and distal side of the interfacial surface. Therefore, no reflection will be seen from sound entering a layer of equal impedance, but varying density and speed of sound compared to the current medium. Reflection and transmission coefficients for intensities are derived by squaring R and T in Eq. 8 (Fig. 3).

Attenuation

Acoustic attenuation manifests itself in many ways. Sound can be attenuated by mechanisms of reflection,

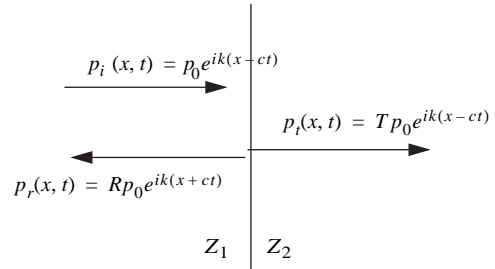


Figure 3. Acoustic propagation is altered when sound waves encounter an impedance change (Z_1 to Z_2), that is, a change in the product of local speed of sound and mass density.

absorption, or scattering. Reflection is caused by impedance changes, whereas absorption and scattering occur due to the internal structure of the medium. Viscous forces cause sound absorption following Lambert–Beer–Bouguer’s law

$$dI = -\beta I dx$$

$$I(x, \beta) = I_0 e^{-\beta x} \quad (9)$$

$$p(x, t) = p(x = 0, t) \cdot e^{ikx \pm ct + ik_i x}$$

where k_i is imaginary, and therefore $p(x, t)$ decays exponentially for $x > 0$. In general, acoustic waves in medical imaging are attenuated by traveling through layers of different tissues due to reflection and also due to attenuation inside tissue. Typical acoustic attenuation in biological tissues is of the order of $0.1\text{--}1.0 \text{ dB MHz}^{-1} \cdot \text{cm}^{-1}$, that is, a acoustic wave of 1MHz center frequency, which travels 0.5 cm deep into tissue (1 cm round trip), is diminished by 0.1–1.0 dB, or its amplitude is reduced by ~1–11%. However, a 2.25 MHz wave penetrating the abdomen to a depth of 15 cm at $0.7 \text{ dB MHz}^{-1} \cdot \text{cm}^{-1}$, will be amplitude attenuated by 47.25 dB or 99.6%. Good ultrasound systems have signal to noise and amplification capabilities up to 120 dB, and therefore allow penetration to a reasonable depth at megahertz frequencies. Typical frequency selections are 7.5–15 MHz for 1–3 cm depths and 2.25–3.5 MHz for 12–15 cm depths.

Pulse–Echo

Most medical imaging via ultrasound uses a pulse–echo method to obtain images. That is, sound waves are transmitted into the body and echoes from within the body are registered, and their origin is computed using complex algorithms. Pulse–echo is somewhat unique to ultrasonic imaging. Other modalities use transmission (X ray and CT) or register preexisting radiation from within the body (PET, SPECT).

Multiple transmissions at the same physical location can reveal the motion of targets. A fundamental assumption is the speed of sound in the investigated volume. Typically, water is assumed to be $1485 \text{ m} \cdot \text{s}^{-1}$, and human tissue varies between 1450 and $4080 \text{ m} \cdot \text{s}^{-1}$ (see Table 1), with an average soft tissue value of $1540 \text{ m} \cdot \text{s}^{-1}$ (6). In general, the speed of sound is inversely related to the compressibility K ($\text{m}^2 \cdot \text{N}^{-1}$) and mass density ρ ($\text{kg} \cdot \text{m}^3$)

Table 1. Speed of Sound in Various Human Tissues^a

Tissue	Mean Velocity, m · s ⁻¹	Tissue	Mean Velocity, m · s ⁻¹
Air	330	Brain	1541
Fat	1450	Blood	1570
Human tissue (mean)	1540	Skull bone	4080
		Water	1485

^aSee Ref. 6.

of the host medium:

$$c = \frac{1}{\sqrt{\rho K}} \quad (10)$$

Figure 4 shows the radio-frequency (rf) data for three firings at a set of moving targets. In this depiction, one can assume that the firings shown in a–c occur at a 1 ms interval. As time progresses, the scatterers move farther away from the transducer. At (a) the particles are $1/2 \cdot 26 \mu\text{s} \cdot 1485 \text{ m} \cdot \text{s}^{-1}$, that is, 19.3 mm, away from the transducer; in (b) the particles shifted to $1/2 \cdot 32.5 \mu\text{s} \cdot 1485 \text{ m} \cdot \text{s}^{-1}$, that is, 24.1 mm, away from the transducer. This shift of $6.5 \mu\text{s}$ or 4.8 mm is related to the interfiring time referred to by either pulse repetition interval or frequency (PRI or PRF). A PRI of 1 ms leads to the conclusion that the set of particles is moving at a speed of $4.8 \text{ m} \cdot \text{s}^{-1}$.

ULTRASOUND GENERATION

Physics

Sound is produced by anything that moves in an accelerated fashion. Nowadays, most practical materials are piezoelectric, such as quartz (SiO₂), polyvinylidene fluoride (PVDF), and lead zirconate titanate (PZT). Piezoelectricity is an effect that is associated with the crystalline structure of the materials. A piezoelectric crystal yields a voltage across its surface when under strain, and the reverse effect facilitates mechanical oscillations of the crystal in response to an alternating current (ac) electric field applied across its surface.

Transducer Construction

Ultrasound transducers are made from piezoelectric materials, as described above. Typically, a layer of material is

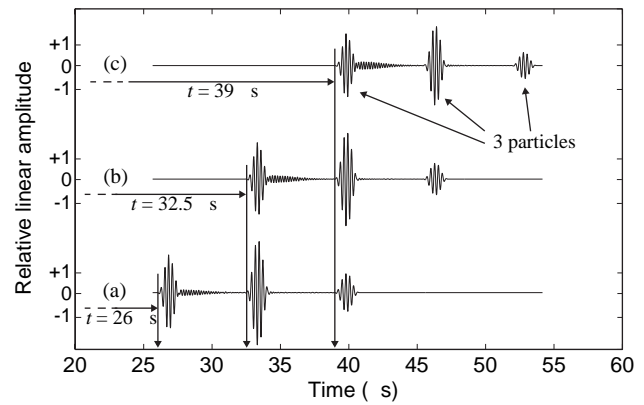


Figure 4. A set of three particles in water moves away from the ultrasound transducer during a set of three transmissions and receptions. (a) At time $t = 0 \text{ ms}$ the first backscatter signal of the set is received after $26 \mu\text{s}$. (b) For the second acquisition the first backscatter is received at 32.5 and $39 \mu\text{s}$ at (c). Signal travel time is directly related to travel distance by means of the speed of sound.

used to create a surface area for creation and transmission or reception of ultrasonic waves. The thickness of this layer is a function of the material properties and the desired acoustic frequency. As seen in Fig. 3, acoustic waves are reflected by impedance changes. An oscillating layer of piezoelectric material produces mechanical waves that propagate in the oscillation direction. These waves can be either compressional or shear waves. Here, the focus will be on compressional waves. Constructive interference of waves launched or reflected from the front surface and from the back surface of the crystal yield maximum pressure generation. High frequency transducers are made from very thin crystals due to their short wavelength and low frequency transducers are made from larger thickness crystals. For example, a 4 MHz transducer can be made from a 0.55 mm thick crystal. Table 2 lists the speed of sound in PZT_{5A} as $4400 \text{ m} \cdot \text{s}^{-1}$. The wavelength in PZT_{5A} at 4 MHz is 1.1 mm. Transducer crystals are typically machined to a thickness of $\lambda/2$, that is, 0.55 mm for 4 MHz. The rationale for this thickness is in the constructive interference of acoustic waves inside of the crystal. Figure 5 shows the bottom of the crystal moving up and down. Mechanical waves will launch from this surface and travel to either side of it. Assume that the

Table 2. Piezoelectric Properties of Typical Materials Used for Fabrication of Ultrasound Transducers

Property	Units	PVDF	PZT _{5A}	Quartz (x-cut)
Density	$\text{g} \cdot \text{m}^{-3}$	1.78	7.6	2.6
Relative permittivity	ϵ/ϵ_0	12	1700	4.52
Elastic modulus	$10^{10} \text{ N} \cdot \text{m}^{-1}$	0.3	4.9	
Piezoelectric constant	$10^{-13} \text{ C} \cdot \text{N}^{-1}$	$d_{31} = 20$ $d_{33} = 30$	$d_{31} = 180$ $d_{33} = 360$	
Coupling constant		0.11	$k_{31} = 0.35$ $k_{33} = 0.69$	
Speed of sound	$\text{m} \cdot \text{s}^{-1}$		4400	5740
Characteristic impedance	MRayl			15.2

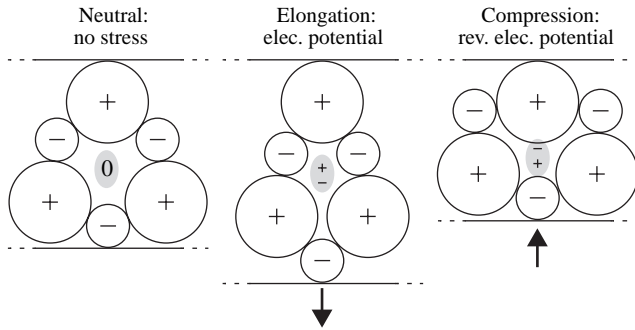


Figure 5. A piezoelectric crystal exhibits an electric charge on its surface when under mechanical stress (shown as the elementary cell response). The reverse effect is used to produce ultrasound; applying an alternating electrical potential across a piezoelectric crystal causes the crystal to vibrate along a given direction.

bottom side of the surface is facing air and that there is no sound transmitted into it. The sound wave traveling toward the top surface will be transmitted beyond that surface into the desired medium (e.g., tissue).

Reflected waves will travel downward and interfere with upward traveling waves. Moreover, reflected waves experience a phase shift of 180° . This is the reason that $\lambda/2$ is the required thickness and not λ .

Array Design. Modern clinical ultrasound imaging arrays are composed of hundreds of individual piezoelectric elements. Mostly these elements are arranged in a linear, 1D fashion, hence their name: linear arrays. The 3D in front of an imaging array are denoted axial, lateral, and elevational. Axial and lateral axes lie in the imaging plane, with the axial distance extending away from the transducer. The lateral axis is parallel to the transducer surface, inside the imaging plane, whereas the elevational axis extends perpendicular to the imaging plane. By convention, the origin is located at the transducer surface in the middle of the active aperture (see the section Acoustic Imaging). Each element may be rectangular with a fixed curvature for focusing in the elevational direction. Curvature as well as the elevational size of each array element determine the thickness of an image plane, which can be 1 mm. Typical element widths range from $\lambda_m/2$ to $3/2 \lambda_m$. The wavelength λ_m is the wavelength within the medium where the wave is launched. Arrays with element sizes of $\lambda/2$ or smaller are referred to as fully sampled. Element sizes are typically 0.5–1 cm in the elevational direction and hundreds of micrometers in the lateral direction. The size of the space between individual elements is called the kerf (Fig. 6) and it is meant to diminish acoustic crosstalk between adjacent elements. A major design criterion for arrays is the distance between the centers of elements, called pitch. This distance determines the location and amplitude of acoustic grating lobes. Variations in the pitch, either due to changes in kerf or element width cause the grating lobes to shift. Moreover, the total extent of the aperture is directly related to the full width at half the maximum (fwhm)

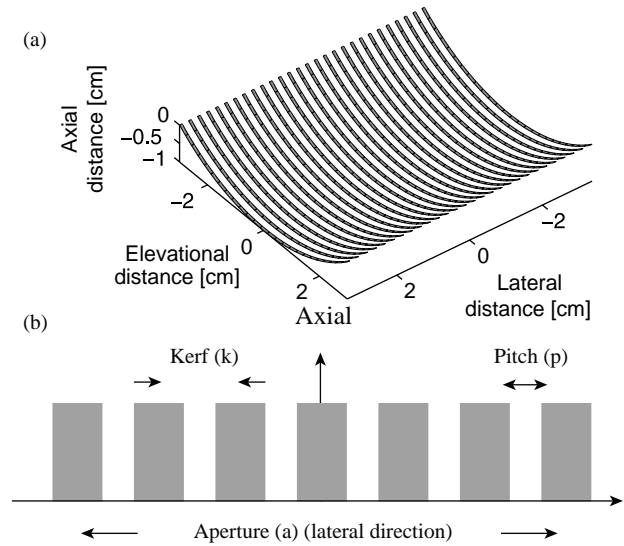


Figure 6. (a) Schematic of the piezoelectric elements of an array ultrasound transducer. Elevational curvature as well as spacing (white area, called kerf) between shaded elements is exaggerated. (b) Definition of geometric parameters: kerf, pitch, and aperture.

amplitude of the main lobe and the location of the side lobes. Equation 11 gives the fwhm of the main lobe and the angular position of the side lobes as well as grating lobes. Side lobes result from a transmit and receive aperture being small relative to the wavelength of the acoustic wave (λ/a), whereas grating lobes result from a steered and undersampled aperture, that is, an aperture with too few elements per wavelength.

$$\begin{aligned}
 \text{fwhm main lobe} \quad \Lambda &= \frac{aw}{p} \\
 \text{side lobes} \quad \theta_s &= \arcsin \frac{\lambda n}{a} \quad n \in N_0 \\
 \text{grating lobes} \quad \theta_g &= \arcsin \frac{\lambda n}{p}
 \end{aligned} \quad (11)$$

The consequence of improper selection of pitch and kerf for a given frequency is illustrated in Fig. 7a. At an imaging depth of 10 cm one can see a main lobe of almost 3 mm fwhm and strong side lobes and grating lobes. In this case the ratio of λ to element pitch is 0.4. For a ratio of λ to element pitch of 1.6, the width of the main lobe is 1.2 mm and the side lobes are mostly suppressed (Fig. 7b). Moreover, the grating lobes have disappeared when the ratio of λ to element pitch is increased.

High spatial resolution imaging is achieved by a proper selection of these geometry factors. Of additional importance to the acoustic field pattern are acoustic output, field of view, and practicality. The smaller the actual radiating area, the lower the acoustic pressure in the generated field. Moreover, a large pitch and/or element width will, for a given number of elements, cause the active aperture to increase in size, which limits the possible shift of the active aperture across the physical aperture of the array (see the section Acoustic Imaging). A

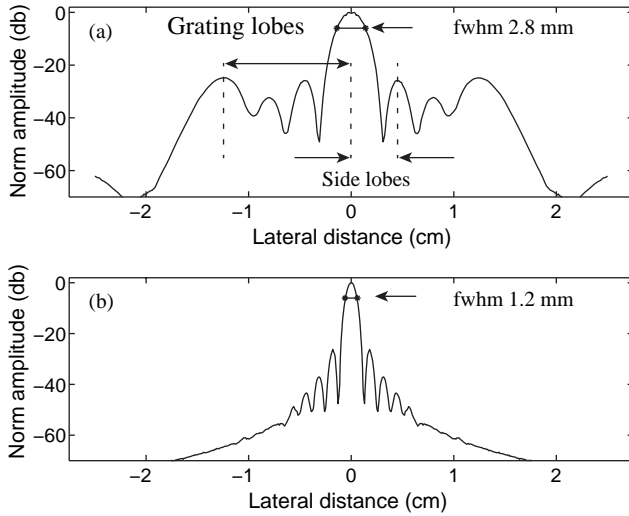


Figure 7. (a) Four elements ($200\ \mu\text{m}$ kerf, $1000\ \mu\text{m}$ pitch) using 4 mm aperture, yield a main sound lobe width of 2.8 mm fwhm. (b) Forty elements ($50\ \mu\text{m}$ kerf, $250\ \mu\text{m}$ pitch) using 1 cm aperture, cause the main lobe to narrow to 1.2 mm fwhm and the sidelobes and grating lobes to diminish. Half maximum corresponds to -6dB .

large number of elements is beneficial for focusing and field geometries, but it is physically difficult to electrically wire a substantially larger number of elements. Furthermore, it directly increases costs, since each element requires cabling and multiplexing electronics. Channels with transmit and receive electronics are in general multiplexed to the physical elements. The complications involved with cabling hundreds or thousands of elements as well as the challenge of real-time data acquisition of a large number of channels are directly related to the still limited usage of two-dimensional (2D) imaging arrays. A good introduction and very detailed overview can be found in Shung and Zipparo (16), as well as Angelsen et al. (1,17).

Acoustic Fields. Acoustic fields of transducers can be analytically derived from basic principles. Logically, this derivation originates at the sound source, a moving object or surface. Its motion and surface shape/orientation define the so-called source strength Q (Eq. 12). Using this source strength one can compute the actual pressure field at a distance r from the source. In order to do so, it is necessary to derive the fact that for all simple geometries, the ratio of source pressure P_1 to its source strength Q_1 and the ratio of a second source pressure P_2 to its source strength Q_2 are equal at the same distance (assuming the same frequency). However, the derivation of this equality is beyond the scope of this text. The pressure of a circular piston can now be written as a function of its source strength, as well as the pressure and source strength of a known source, typically a small sphere. Equation 13 gives the analytical expression for the pressure field of the piston transducer, where r is the distance of the observation point from the center of the aperture, θ is the angle between the axis extending perpendicularly

from the center of the transducer and the line from the center to the observation point, t is time, ρ is the mass density of the surrounding water, c is the speed of sound in water, U_0 is the sound particle velocity on the aperture, and ω and k are angular frequency and wave number, respectively. The integral is simplified for circular geometry and taken over the entire surface of the transducer's aperture. In *real world* simulations, one could take into account that the circumference of the aperture might be clamped or for other reasons not be able to oscillate with the same amplitude as the center of the aperture. To do so, one would keep U_0 inside the integral as a function of radius or even radius and in-plane angle of the aperture.

$$Q = \int_S (\mathbf{u} \cdot \mathbf{n}) dS \quad (12)$$

$$\frac{P_1}{Q_1} = \frac{P_2}{Q_2} \Rightarrow P_p = \frac{P_s}{Q_s} Q_p$$

$$P_p = \frac{\rho c}{-i2\lambda r} \int_S (\mathbf{u} \cdot \mathbf{n}) dS \quad (13)$$

$$P_p(r, \theta, t) = \frac{i\rho c U_0}{r} e^{i\omega t} \int_S \frac{e^{-ik\sqrt{r^2+s^2}}}{\sqrt{r^2+s^2}} 2\pi s ds$$

After deriving the general pressure field, one can compute special cases that are of particular interest, such as the central axis, as well as the far-field angular distribution of the radiation pattern. Further simplification of Eq. 13 yields approximate expressions for both cases and plots are shown in Fig. 8.

$$p(r, 0, t) = \rho c U_0 \left(e^{-ikx} - e^{-ik\sqrt{x^2+a^2}} \right) e^{i\omega t} \quad (\text{central axis})$$

$$p(r \gg a, \theta, t) = \frac{i\rho_0 c U_0}{2r} a(ka) e^{i(\omega t - kr)} \frac{2J_1(ka \sin \theta)}{ka \sin \theta} \quad (\text{far field}) \quad (14)$$

From the axial dependence one can see strong interference for locations close to the transducer surface. This region is called near the field or Fresnel region and it extends approximately $r = 4$ aperture diameters a away from the transducer. Beyond that point, the pressure falls off following a $1/r$ dependence, and this region is called the far field or Fraunhofer region. As a rule of thumb, the far field starts at $a^2(2\lambda)$. Imaging is impractical or impossible in the near field. However, one should keep in mind that this result is true only for a *single* element transducer, and subdividing the aperture into an array of small elements shifts the near-field–far-field transition toward zero based on the actual dimension of the array. Moreover, this transition point is also a function of the emitted frequency as represented by angular frequency and wave number in Eq. 13 and 14. For illustration purposes, a ka value of 8π was chosen for Fig. 8a. When plotting the angular field pattern in Fig. 8b. However, $ka = 4\pi$ was chosen to reduce the number of sidelobes and make the plot more readable. Similarly to Fig. 7, a strong main lobe and additional side lobes are evident in Fig. 8b, that interfere with the main lobe in the sense that appreciable sound will be detectable in nearly all directions. In fact, for this particular example echo amplitudes

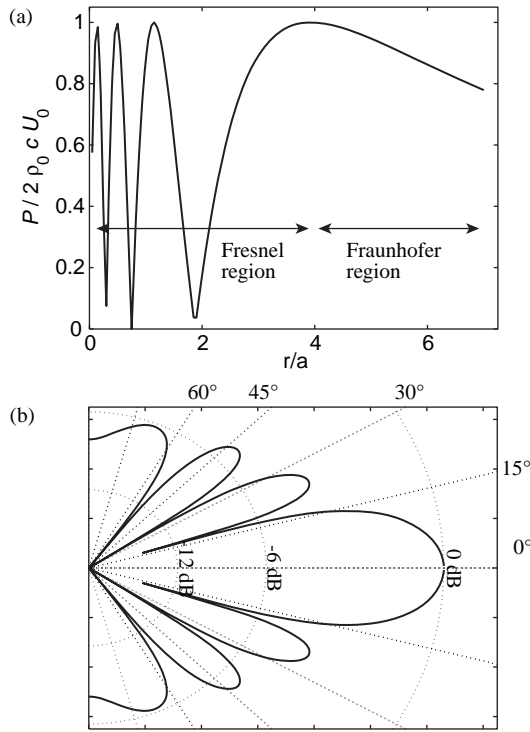


Figure 8. Acoustic field of a circular single element ultrasound source (transducer). (a) The axial transducer response can be divided into near field and far field, with an asymptotic $1/r$ dependence for large r . Phase interference is dominating in the near field. (b) Angular radiation patterns are described by the Fourier transform of the aperture function. A circular piston transducer yields a function that is defined as a Bessel function of the first type divided by its argument $[J_1(x)/x]$.

60–70° off the center axis will be only 50% lower than the main lobe. Note that this example is for educational purposes and is not at all a suitable design for imaging. Aperture sizes and frequencies as shown in Fig. 7, where a narrow and dominant main lobe can yield lateral and elevational spatial selectivity and time range gating, can yield depth information. In general, radiation patterns can be derived as Fourier transforms of the emitting aperture. Circular apertures are described by Bessel functions of the first type divided by their argument, that is, $J_1(x)/x$. For rectangular apertures the sine function $[\sin(x)/x]$ directly describes the field.

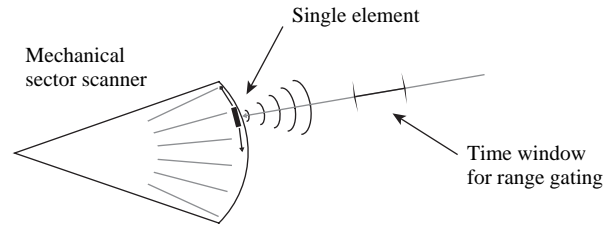
ACOUSTIC IMAGING

Ultrasonic imaging yields 2D images. Pixel columns represent the lateral extent and rows of pixels display reflections of the transmitted acoustic waves from progressively deeper tissues within the body.

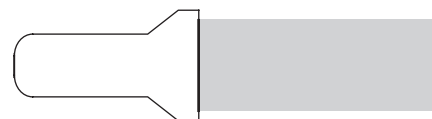
The most rudimentary way of obtaining lateral image data is done by using a single element transducer and wobbling it back and forth over a chosen sector. The most elegant way is to use an electronically controlled array of very small transducers and scanning and or steering an imaging beam across the region of interest. The former

method is rarely used anymore. The latter method is the modern standard for ultrasonic imaging of single image planes and a few clinical scanners are even already available for 3D image acquisition of steered elevational image planes.

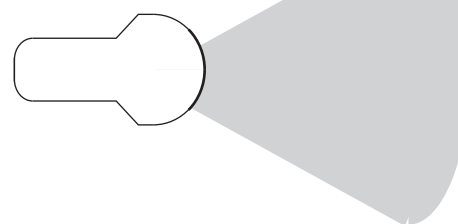
Depth information is encoded in the time that an acoustic wave takes to travel to a tissue site and back to the transducer. Transmitted sound can be received after a theoretically predictable wait time (see top section of Fig. 9). Such prediction requires the knowledge of the speed of sound along the traveled path. Unfortunately, an assumption of $1540 \text{ m} \cdot \text{s}^{-1}$ for soft tissue is not always precise. Various tissues in the human body differ from each other in terms of their specific speed of sound. When performing abdominal scans, aberration distortions can become significant due to the change in the speed of sound between connective tissues, fat layers, muscles, and abdominal organs. Other anatomical sites that provide difficulties for ultrasonic imaging include the human brain



Linear array: small parts, superficial vascular, obstetrics



Curved array: abdominal, obstetric, transabdominal, or for transvaginal or transrectal, or pediatric imaging



Phased array: heart, liver, spleen, fontanelle, temple

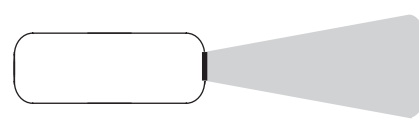


Figure 9. Transducer geometries include (curved) linear arrays and phased arrays. Linear arrays transmit and receive acoustic beams perpendicular to their surface area, whereas phased arrays steer the acoustic beam by a phasing scheme (Fig. 10).

and the heart. Fortunately, for newborns and infants ultrasound can be used to image the brain. Using a phased array (see bottom section of Fig. 9), pediatricians can image the brain directly through the fontanelle, which provides enough acoustical access for imaging. Premature newborns tend to have bleeding in the brain and develop larger ventricles, both of which can be imaged very easily through the fontanelle. However, when this acoustic access window closes, it is very difficult if not impossible to use ultrasound to image the brain. In adults, the temple fissure can be used to image (using a phased array Doppler system at 2.25 MHz) the germinal matrix near the foramen of Monroe. However, transcranial Doppler requires some guesswork on the orientation of vessels. Because of the limited access via the temple fissure, only small aperture and low frequency arrays with sub-optimal imaging capabilities can be employed.

The simplest imaging device is a linear array (range of transducer frequencies: 3–12 MHz). As its name suggests, this type of transducer has a linear arrangement of individual transducer elements. Images from a linear array are generally rectangular, and the image width corresponds to the width of the array. A set of adjacent elements (a subaperture) is used to fire a single image line or a portion thereof. Figure 10 shows how a subaperture can be used to steer and focus a beam. On the left side of each drawing, single sinusoids symbolize the electrical signals being used to excite the individual transducer elements (thick-lined vertical bars) of the array. After excitation an elementary spherically spreading wave launches from each element. Appropriate delay times applied to each electrical signal allow the ultrasound scanner to steer and focus the emitted wave front.

Additional wave conditioning includes amplitude shading of the subaperture. Typically, Gaussian-type functions or approximations thereof are used to weaken the transmit power for the outer most elements of a subaperture. The process is called apodization and yields lower side lobes since the side lobes are related to the Fourier transform of the aperture function. A gaussian apodization will result in a gaussian envelope for the side lobes, whereas no apodization (i.e., a constant amplitude excitation) would yield a sinc function $(\sin(x)/x)$ side lobe envelope.

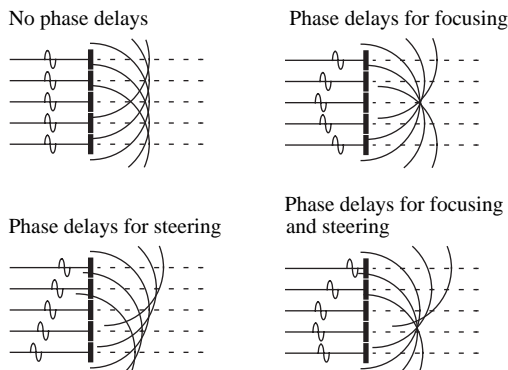


Figure 10. Delaying or advancing the phase of waves emitted from neighboring elements relative to the center element can achieve focusing and steering of the acoustic wave front.

To render an image, a subaperture is formed and linearly moved through the whole physical aperture. A major drawback of this imaging scheme is the dead time of the scanner, which occurs while waiting to receive the backscatter from the maximum depth. Fifteen centimeters of penetration require a wait time of at least (two fold for the round trip time):

$$t = \left[\frac{0.15 \text{ m}}{1540 \text{ m} \cdot \text{s}^{-1}} \right] \times 2 = 200 \mu\text{s}$$

In addition to this delay, some additional wait time may be required to suppress echoes from even larger depths. Image lines might be separated by 250 μm . For a physical aperture of 5 cm, for example, one has to transmit and receive 201 scan lines, which is equal to a time of $200 \mu\text{s} \times 201$ lines, that is, 40.2 ms per one full frame or a frame rate of 25 Hz. This frame rate seems reasonable, but one has to keep in mind that no extra wait time was added nor any other imaging overhead such as occurs during blood flow imaging using Doppler.

Nonetheless, if one takes into account the finite lateral width of a transmitted wave, one can subdivide the total image width into independent image segments in which beams can be fired simultaneously, or at least with minimal delay (see Fig. 11). Therefore an aperture of 5 cm could eventually be imaged with five simultaneous beams or a five-fold higher frame rate. Such high frame rates allow real-time ultrasonic imaging of the body and additional overhead, such as is mentioned above for Doppler or multi-zone focusing. This type of focusing is used when a large depth image would cause the acoustic beam to diverge too much before and after the focal point. By firing the same image line two, three, or four times, the same number of foci can be formed for tighter acoustic beams at larger depths or for shallow regions. This scheme will work to the limit that the beams are not overlapping, that is, there is no signal coming from adjacent image lines.

Other imaging modalities, such as X ray, do not suffer from slow acquisition time due to low wave speeds. Another method to overcome the speed of sound limitation is the use

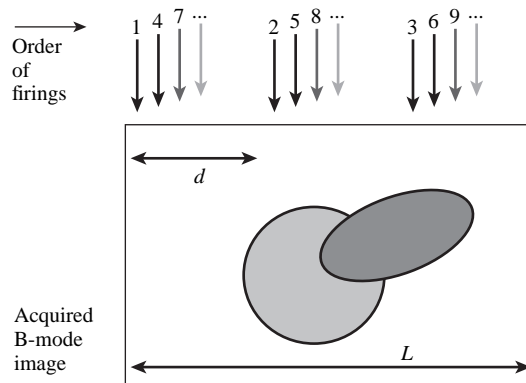


Figure 11. Interleaving scheme for image data acquisition. By minimizing the segment separation d , a maximum of L/d semisimultaneous transmit and receive zones can be achieved. The beam shape determines the smallest value of d , such that no overlap between adjacent firings will occur.

of *synthetic aperture* imaging. This method is already used in astronomy. Instead of 201 firings for 201 image lines, only one transmit is fired, and only one receive is recorded. Image reconstruction and especially spatial resolution will be computed/extracted from received data by extensive postprocessing. This scheme does not yield as much detailed backscatter data; however, it does yield very high frame rates. If only the speed of sound is limiting the data acquisition, then for the above example, the frame rate will increase by a factor of 201. Only the vast processing power of current computers makes synthetic aperture imaging possible. Typical applications for this type of image formation include full screen flow imaging, 3D, or 4D imaging (see respective subsections).

Other types of arrays include curved arrays and phased arrays which are popular for scans that require a larger image width then can be achieved by simply extending the physical aperture. Curved arrays, as seen in Fig. 9, form sector images. Because of the shape of the aperture, a relatively wide image can be achieved using a smaller footprint aperture. Scan lines are no longer parallel to each other but form a fan beam arrangement with field of view angles of up to 85° (150° for some endocavitary arrays). Typical bandwidths range from 2–8 MHz, which is a lower range than for linear arrays since this type of probe is intended for large penetration depths where frequency dependent attenuation prohibits very high frequencies. The use of nonionizing radiation to achieve real-time imaging with large fields of view makes these probes ideal for abdominal interoperative guidance of, for example, biopsy needles or radiofrequency (RF) ablation tumor treatments. These features are also ideally suited for obstetrics (see Fig. 12).

Phased arrays are also designed to form sector images. Contrary to curved arrays, where the natural shape of the



Figure 12. B-mode image of a fetus at the end of the second trimester (cross-sectional sagittal view). Low backscatter amniotic fluid is surrounding her head and upper body. The video reveals that the embryo is sucking her thumb. A curved array was used in this obstetrics exam.

physical aperture provides the basis for the sector shape, phased arrays steer the beam to form the image. As illustrated in Fig. 10, specific timing delays for the subaperture can not only focus to a specified depth, but also steer the beam in the lateral direction. Large fields of view can be achieved this way, but the development of increased side and grating lobes are a trade-off. Anatomical locations with small diameter access to larger distal regions can be imaged with this type of ultrasound array. For example, temple access can be used to image the frontal brain, and extension of the carotid artery above the jaw line is possible. Cardiac imaging typically relies on phased arrays due to acoustic shadowing from the rib cage, where one needs to image between narrowly spaced ribs in order to interrogate the much larger sized heart chambers.

B-Mode

B-mode is one of the most commonly used operation modes of a clinical ultrasound scanner. As explained earlier, ultrasound is a reflection or scattering based imaging modality, and the sophisticated generation of a sound wave allows the focusing of the sound to a specific location. Each transmission yields one scan line around the targeted focal point. If only one focal point is selected, one scan line extends over the total depth range, which is user defined in the current imaging settings. In order to record a full image frame using a linear array, the imaging software of the scanner electronically moves the active aperture of the array across the physical aperture to transmit and receive at a given line density. Typically, hundreds of scan lines are generated this way and displayed on a monitor. Figure 12 shows the cross-sectional sagittal (front to back, vertical slice) image of a fetus *in utero*. She is sucking on her thumb, as real-time video reveals. On the left of Fig. 12 one can see the head and the strong reflection of the skull bone. The very left side of the image is black, an artifact that could be due to maternal bowel gases that scatter the sound away from the transducer. The remainder of the skull is clearly visible from the forehead to the chin and from the back of the head to the neck area. Bones reflect sound waves well and result in a bright signal in the image. The black surroundings of the fetus are regions of amniotic fluid, which does not scatter sound due to the homogeneous nature of the fluid.

In front of the mouth, one can see the hand of the fetus. Once again the bones of the hand, namely, the knuckles, are pronounced since they scatter more ultrasound than the soft tissue of the hand. In the same fashion one can see the reflections of the spine.

M-Mode

M-mode (also called motion-mode imaging) does not yield full frame images per se, but rather one selected image line is rendered as a function of time. This is used for displaying motion of, for example, the periodic movement of heart valves. Any abnormalities or temporal variations can be directly seen as an image on the screen. The B-mode cross-section of a carotid artery is shown in Fig. 13a. Proximal and distal vessel wall delineates the dark vessel interior, as indicated by the arrows to the right.

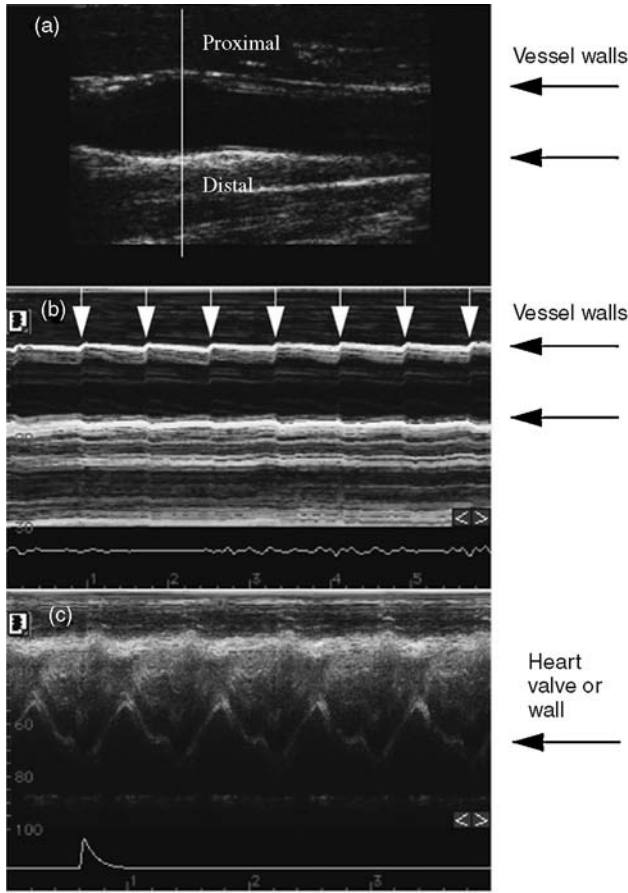


Figure 13. a. Longitudinal cross-section of carotid artery. The pixels along the vertical white line in (a) are plotted in (b) as a function of time. Black arrows indicate the proximal and distal vessel wall. Such walls are in motion as blood is pumped through the vessel in a pulsatile fashion. The repetitive pulsatile wall motion can be seen in the M-mode image in (b).

In a M-mode representation in Fig. 13b, pixels along the white vertical line in (a) are repeated parallel to each other over time. Figure 13b shows 5 s of repeated scans. For each heart beat a pulsatile wave travels through the arterial blood pool locally expanding the blood vessels. This expansion can be seen in B-mode as well as in M-mode representation. However, in B-mode it is an event in time occurring over several image frames, whereas in M-mode this event is plotted as the horizontal axis and therefore easy to detect. White arrows in Fig. 13b indicate the temporal expansion of the blood vessel. Figure 13c shows a much more pronounced motion. The transducer was pointed toward the heart and is therefore either imaging the heart wall or one of the heart valves, showing the typical cardiac pattern.

Doppler Imaging

Acoustic transmission of multiple beams along the same line can reveal temporal changes in the human body. As seen in Fig. 4, pulse echo fringes in a rapid fashion can track flow, as well as flow changes in time. A more formal

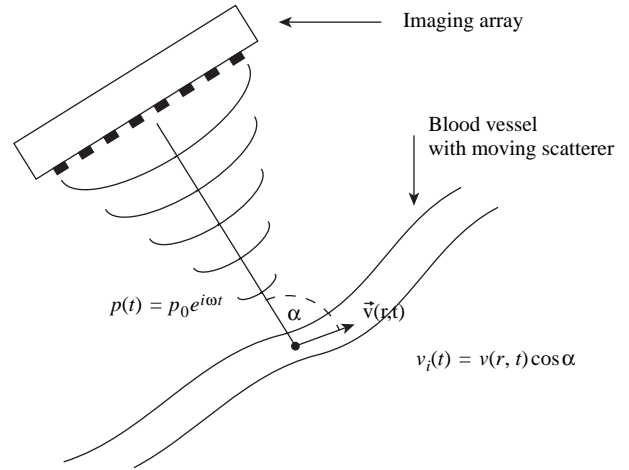


Figure 14. Illustration of Doppler imaging of a blood vessel.

derivation of the mathematical framework shall follow here.

Assume a vessel that is imaged at some angle α , which has acoustic scatterers such as red blood cells flowing at speed $v(r, t)$, as is depicted in Fig. 14. A recorded acoustic echo has an amplitude, frequency, and phase ($\alpha \cdot e^{i(\omega t + \phi)}$). The measured phase is the sum of temporal and spatial phase components. The first term $\omega \Delta t$ in Eq. 15 changes in time with ω and the second term changes in time with the velocity $v(t)$ of flowing red blood cells. Flow speed $v(t)$ and direction α ($\cos \alpha$) determine the magnitude of the measurable phase shift. Due to the $\cos \alpha$ term, any displacement that occurs parallel to the aperture will not be detected.

$$\Delta\phi(\Delta t) = \omega\Delta t - 2\pi \frac{v(t)\cos\alpha\Delta t}{\lambda} = \omega\Delta t \left(1 + 2\frac{v(t)}{c} \right) \quad (15)$$

It is assumed that the time between firings Δt is small enough that the scatterer does not move out of the main lobe of the beam pattern (Fig. 7). Moreover, it is assumed that $v(t)$ is constant during Δt . The absolute and relative received Doppler shift frequency can be directly derived from the change in phase, as the temporal derivative of the phase angle (Eq. 16).

$$\begin{aligned} f_{\text{receive}} &= \frac{1}{2\pi} \frac{\partial}{\partial t} \Delta\phi(t) = \frac{\omega}{2\pi} \left(1 + 2\frac{v_i(t)}{c} \right) \\ &= f_{\text{transmit}} \left(1 + 2\frac{v_i(t)}{c} \right) \\ \Rightarrow \frac{f_{\text{receive}}}{f_{\text{transmit}}} &= \left(1 + 2\frac{v_i(t)}{c} \right) = \left(1 + 2\cos\alpha \frac{v(t)}{c} \right) \end{aligned} \quad (16)$$

In the following example, a simulated scatterer is imaged and its traveling speed is measured. In the computation, the scatterer is travelling away from the imaging array from an axial distance of 2.50–2.52 cm, that is, 200 μm . On that path, the scatterer is imaged 32 times, once every 0.36 ms. Its speed is 2 $\text{cm} \cdot \text{s}^{-1}$. Figure 15a displays the backscatter of a scatterer as shown in Fig. 4, except that only one scatterer is imaged. Thirty-two

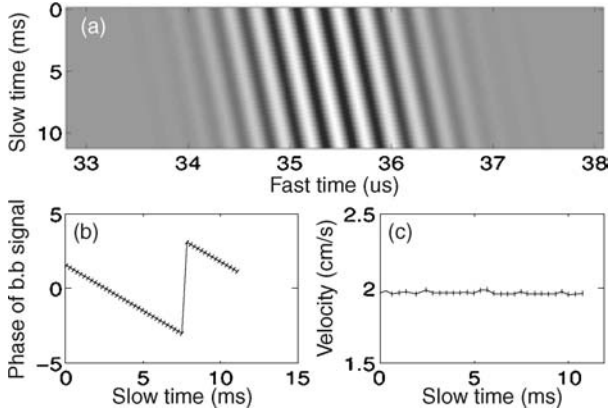


Figure 15. Doppler processing for velocity estimation. (a) Shows stacked (slow time) backscatter signals (fast time) of a moving scatterer. Complex analysis of such signals reveals the change in phase of the backscatter signal (b), and subsequently the scatterer velocity can be computed (c).

backscatter signals are stacked vertically, and the ordinate is labeled with the time at which the signal was measured (“slow” time in ms), whereas the abscissa shows the time frame of the measured radiofrequency signal (“fast” time in μ s) same as in Fig. 4. This arrangement of backscatter data is very similar to that used in M-mode imaging. Signal amplitude is displayed as gray scale with gray for zero, black for negative amplitudes, and white for positive amplitudes. Two major steps are performed in order to estimate the velocity of the scatterer from the backscatter signal. At first the signal $f(t)$ is transformed into a complex valued signal $f^*(t)$ by performing a Hilbert transform (Eq. 17). Measured signals are always real valued quantities. However, for computational purposes it is desirable to have complex valued data $f^*(t)$. This step allows us to directly measure the phase of the backscatter signal and yield the velocity of the scatterer after basebanding, which is the second step.

$$f^*(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{f(\tau)}{t - \tau} d\tau \quad (17)$$

$$= \frac{1}{\pi} \lim_{\epsilon \rightarrow 0^+} \left(\int_{-\infty}^{t-\epsilon} \frac{f(\tau)}{t - \tau} d\tau + \int_{t+\epsilon}^{\infty} \frac{f(\tau)}{t - \tau} d\tau \right)$$

Basebanding is a mathematical procedure used to remove the carrier frequency, the main transmit frequency, from a rf-signal. For a complex valued signal, $f^*(t)$, this is done by multiplying the signal with a complex harmonic of the same, but negative frequency ($-\omega_0$) as the carrier to obtain the complex valued envelope or amplitude modulation $a(t)$ and the phase ϕ (Eq. 18).

$$f^*(t)e^{-i(-\omega_0)t} = a(t)e^{i\phi(t)} \quad (18)$$

The phase is constant for each rf line in tissue, but varies between firings as targets, such as red blood cells, move. Figure 15b shows the phase of the basebanded signal and therefore the position of the scatterer. At ~ 7.5 ms, the phase wraps from $-\pi$ to $+\pi$ and continues

to decrease. This phase wrap was detected and unwrapped before computing the velocity as proportional to the derivative of the phase. This phase unwrapping is not performed in clinical ultrasound scanners. Rather one will see flow of the opposite direction being displayed on the screen as the Doppler processing unit concludes that the sudden increase in phase from $-\pi$ to $+\pi$ must be due to flow in the opposite direction. This artifact is called aliasing and is typically avoided by increasing the pulse repetition frequency (PRF), that is, the rate at which Doppler firings are repeated along the same scan line in order to track backscatter from blood cells. Inverting Eq. 16 for $v(t)$ and using data processed via Eq. 17 yields the flow velocity as given in Eq. 19. A comprehensive description of medical Doppler and Doppler physics can be found in and McDicken and Evans (18).

$$v(\tau) = \frac{c}{2\omega_0} \frac{d}{d\tau} \phi(\tau) \quad (19)$$

Pulse Wave Doppler. Pulse wave Doppler (PW Doppler) is used for measuring blood flow. The user can position a Doppler scan line and Doppler window to any location within the B-mode image, as seen in Fig. 16. The two short horizontal lines at a depth of 6.9 cm in the top B-mode image in Fig. 16 represent the sample volume. This is where the Doppler data is acquired. Typically, the beamformer of the scanner is set to the same sample volume location for transmit and receive. The axial size of this window can be adjusted and is displayed on the screen (here 2 mm, see Size in the right side data column of Fig. 16). Changes in the window’s size will affect the duration of the transmitted tone burst cycles. Commonly scanner software adjusts the duration of the transmit pulse to be twice as long as the chosen sample volume. Additionally, an angle (α) can be selected along which a blood vessel is oriented (here 0° , for example, along the Doppler scan line). As shown in Fig. 14 and Eq. 16, the measured flow

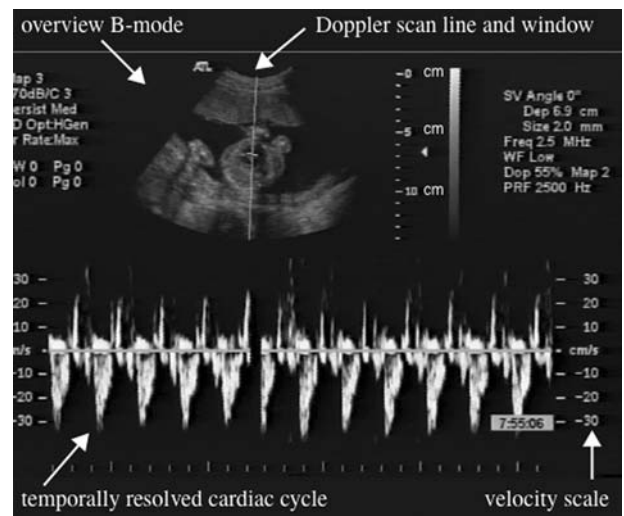


Figure 16. Pulse wave Doppler example. See text for more description.

velocity is only the projection of the actual flow vector onto the acoustic beam, which results in a $\cos \alpha$ term. By manually choosing the correct α , the actual flow velocity can be computed and displayed from the measured flow.

The bottom part of the screen image in Fig. 16 shows the temporally resolved blood velocity, where the abscissa represents time (here, a total of 5 s), and the ordinate represents velocity (here, from -30 to $+30 \text{ cm} \cdot \text{s}^{-1}$).

Traditional processing computes the power at each frequency and subsequently the associated velocity, as outlined in Eq. 20.

Here the phase ϕ of the basebanded backscatter signal $a(t)e^{i\phi(t)}$ is digitized along the slow time scale (ms). Fourier analysis is used to determine the frequency or rate of change (ω_n) of the phase $\phi_n = \phi(t_n)$. More precisely, the so-called spectral power $P(\omega_n)$ at each frequency ω_n is computed by Fourier analysis. This quantity yields how much contribution to the power there is for a given rate of change or speed v_n . These two quantities $P(\omega_n)$ and v_n are plotted in the velocity graph in Fig. 16 (lower plot). The gray level for a given point in the graph is determined by $P(\omega_n)$, whereas v_n or ω_n and the time t determine the location of the pixel. The indicated cardiac cycle in Fig. 16 shows contributions from high velocities that yield high Doppler frequencies. At the end of this cycle the blood flow slows down and contributions to high frequencies diminish and formerly white pixels are now plotted in black. Other operations such as windowing of the phase signal ϕ_n are neglected here for simplicity.

$$\begin{aligned} f^*(t)e^{-i(-\omega_0)t} &= a(t)e^{i\phi(t)} \\ P(\omega_n) &= \left| \sum_n \phi_n T_n e^{i\omega_n n} \right|^2 \\ v_n(t) &= \left(\frac{\omega_n(t)}{\omega_0 T_{rep}} \right) \left(\frac{c}{2} \right) \cos \alpha \end{aligned} \quad (20)$$

Color Flow Doppler. Color flow Doppler allows the user to see a 2D map of flow in the current B-mode image. Instead of measuring flow only along a single scan line as in Pulse wave Doppler (PW Doppler), all lines inside a chosen region of interest (ROI) are fired repetitively (4–16 times) and analyzed for flow. Velocity resolution and frame rate are limited due to the large number of acoustic transmissions and the computational burden of analyzing the resulting received waveform data. In the same fashion as in B-mode, interleaved imaging can be used to counterbalance the reduction in frame rate caused by the necessary increase in (Doppler) scan lines (see Fig. 11).

In Fig. 17, one can see a regular B-mode image of a carotid artery, extending from the left to the right side of the image, parallel to the transducer face. The gray scale B-mode image is overlaid with Doppler information inside a user-defined color flow box, which is either a rectangle or a parallelogram slanted either to the left or right. For this specific example, a rhomboid with 20° rightward steering was used in order to measure the velocity of blood flow. One should remember that vessels parallel to the transducer do not yield a Doppler or phase shift and therefore can not be identified with Doppler methods. However, a 20° rightward steering provides enough angle deviation to measure flow

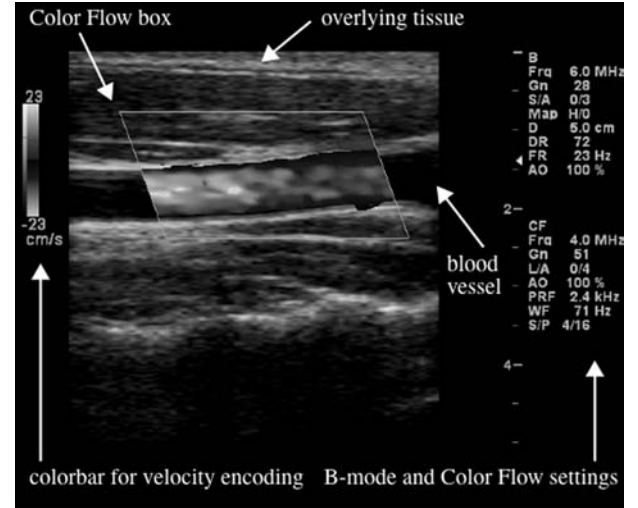


Figure 17. Color flow example showing blood flow in the carotid artery. The color data (here in gray scale) encodes magnitude and direction of flow. A color bar allows the quantitative conversion to actual speed, here a maximum of $\pm 23 \text{ cm} \cdot \text{s}^{-1}$.

and to display velocities everywhere in the chosen color flowbox. Ultrasonic waves are transmitted parallel to the slanted rhomboid.

Typically, blood flow is encoded from dark red to red to yellow when it is approaching the transducer, indicating low, moderate, or high velocities, respectively. Alternatively, it is colored in shades of blue to cyan when it is moving away at low to high speed. The maximum detectable velocity magnitude is directly related to the PRF used. A PRF of 2.4 kHz was chosen in the example shown, resulting in a maximum detectable velocity of $\pm 23 \text{ cm} \cdot \text{s}^{-1}$. Note that the PW Doppler example in Fig. 16 used a 4% greater PRF but allows a 30% greater velocity range. This is related to the greater burst length used in PW Doppler (15–20 cycles) compared to CF Doppler (2–6 cycles) and the subsequent data processing. The PW Doppler is designed for high velocity resolution and simultaneous temporal resolution, whereas CF Doppler is designed for great spatial resolution. Moreover, one should notice the very low backscatter level of blood relative to the surrounding tissue, which can be as much as 40 dB below that of soft tissue. The blood vessel in Fig. 17 appears black relative to the tissue on the proximal and distal side of it.

As mentioned above, CF Doppler demands not only more acoustic transmissions, but also more computing power to estimate the actual flow velocities from the measured acoustic backscatter. This imaging modality became practical when Kasai et al. (12) succeeded in designing a method (see Eq. 21) by which the mean Doppler frequency, that is, the mean velocity in every pixel, could be computed in real time by cross-correlations of the quadrature components, I and Q , of the analytic (complex) backscatter signal. Quadrature components I and Q are obtained by mixing the rf-signal at the hardware level with $\sin(\omega_0 t)$ and $\cos(\omega_0 t)$. This corresponds to the complex base banding given in Eq. 18, since $e^{-i\omega t} = \cos \omega t - i \sin \omega t$. It follows that I and Q are the real- and complex- valued parts of the

basebanded signal, $a(t)e^{i\phi(t)} = I + iQ$.

$$\phi = \frac{\sum_{n=1}^{N-1} Q[n]I[N+n] - I[n]Q[N+n]}{i\sum_{n=1}^{N-1} Q^2[n] + I^2[n]} \quad (21)$$

$$\bar{v} = \frac{c}{2\omega_0 T_{rep}} \arctan \phi$$

Power Doppler

The previous two methods of flow quantification suffer from a lack of good flow detection. In perfusion studies, it is often necessary to detect very small amounts of flow volume travelling through capillaries at velocities of the order of $1 \text{ mm} \cdot \text{s}^{-1}$, which is $\sim 0.1\%$ of the speed observed in the carotid artery. In order to overcome the poor sensitivity of PW and CF Doppler, power Doppler displays the integral of the power spectrum $P(w)$, shown in Eq. 20. Typically, the integration value is also averaged over a very long period of time (several heart beats). Averaging at least one cardiac cycle results in a nearly constant value for flow. Physically, this value represents the amount of blood flowing, but not the velocity, since it is the integral of all detected velocities.

First, imaging capillary flow yet remains difficult, even in power Doppler mode, partially because blood cells do not scatter much of the transmitted acoustic signal. Small blood vessels, such as the capillaries, provide small fractional blood volume, which further decreases the total backscattered signal. Second, at $1 \text{ mm} \cdot \text{s}^{-1}$, flow velocities in capillary beds are difficult to differentiate from static soft tissue background at zero frequency shift, without being suppressed by the wall filter. This filter is used to prevent tissue motion from being incorrectly ascribed as real flow. In Fig. 17 flow that causes $< 71 \text{ Hz}$ frequency shift per Doppler firing is filtered out of the Doppler data to eliminate flow speeds of $< x \text{ cm} \cdot \text{s}^{-1}$.

The ultrasound machine transmits Doppler pulses every 0.42 ms (reciprocal of 2.4 kHz). Backscatter will contain phase shifts between $-\pi$ and $+\pi$, which corresponds to $+23$ and $-23 \text{ cm} \cdot \text{s}^{-1}$ flow speed.

The acoustic wavelength of the color Doppler (CF panel on the right side in Fig. 17) transmits equals 0.385 mm in tissue ($c = 1540 \text{ m} \cdot \text{s}^{-1}$, $Freq = 4 \text{ MHz}$, $\lambda = c/Freq$) and each firing is separated 0.42 ms (reciprocal of $PRF = 2.4 \text{ kHz}$). The Doppler electronics measures phase shift, therefore it can not measure more than a shift of 2π or $\pm\pi$. Two-pi corresponds to λ or $\pm\pi$ to $\pm\lambda/2$, hence the maximum detectable speed of $v = s/t = 45.8 \text{ cm} \cdot \text{s}^{-1}$, with $s = \lambda/2$ and $t = 1/PRF$. However this value does not match the displayed $23 \text{ cm} \cdot \text{s}^{-1}$. Doppler pulses work in pulse-echo mode, in which any displacement dx results in a time shift of 2-times dx/c . Finally the maximum detectable flow speed is given by equation 22. For a given wallfilter (WF) the minimal detectable flow is given by the ratio of the PRF to wall filter times the maximum flow, that is, $(2.4 \text{ kHz}/71 \text{ Hz}) * 23 \text{ cm} \cdot \text{s}^{-1} = 0.68 \text{ cm} \cdot \text{s}^{-1}$.

$$v = \frac{\delta s}{\delta t} = \frac{0.5 \cdot c / (2f)}{1/PRF} = \frac{0.5 \cdot 1540 / (2 \times 3.75 \times 10^6)}{1/2.4 \times 10^3}$$

$$= 23 \text{ cm} \cdot \text{s}^{-1} \quad (22)$$

3D Imaging

Current ultrasound images are naturally 2D because ultrasound imaging arrays are only 1D. One-dimensional arrays are still predominant in the market. Even so, great efforts in the ultrasound community are pushing ultrasonic imaging toward 2D arrays. The transition from 1D to 2D is especially apparent in the naming scheme of current arrays:

- 1D: Is the classic linear or focused array, which has one row of elements that allows focusing and steering in the lateral imaging plane. The elevational focus is constant due to a fixed elevational curvature of each element.
- 1.25D: Extra rows of elements on either side of the main row allow changes in the elevational aperture, but there is no electronic elevational focusing, nor steering.
- 1.5D: This class of arrays has a 2D set of elements, where the elevational elements are connected symmetrically to the center row. This array can focus in the elevational direction but not steer.
- 1.75D: A 2D set of individually driven elements is available for this type of array, but the number of elements in the elevational direction is much less than in the lateral direction. Elevational focusing is possible, but only limited elevational steering is available.
- 2D: Elevational and lateral directions should be equivalent and indistinguishable for a true 2D array. Full apodization, steering, and focusing is possible in 3D. Currently there are some commercial systems that use 2D arrays particularly in cardiac imaging.

Hardware and software implementations allow the 3D reconstruction of a scanned volume even when using 1D arrays. Sophisticated 3D hardware position sensors allow ultrasound scanners to register the position and orientation of a 1D array in 3D space. Therefore, any acquired image in a set of many can be aligned with others in the set to render a 3D volume (see Fig. 18). However, these hardware additions are costly and can be inconvenient. Moreover, they might show limitations due to interference with electromagnetic fields or nearby metallic objects. Software solutions use correlations between adjacent image frames to determine the transducer translation or rotation. Figure 18 rudimentarily illustrates how individual frames taken in freehand fashion are “stitched” together to form a 3D volume, which can be rendered in various ways. Figure 19 shows an anatomical example of the bifurcation of the ascending carotid aorta rendered as a 3D volume. Some implementations on clinical scanners, however, already use the 4D nomenclature by adding time as the fourth dimension.

CONTRAST IMAGING

As in every other clinical imaging modality, agent based imaging enhancements are available for ultrasonic

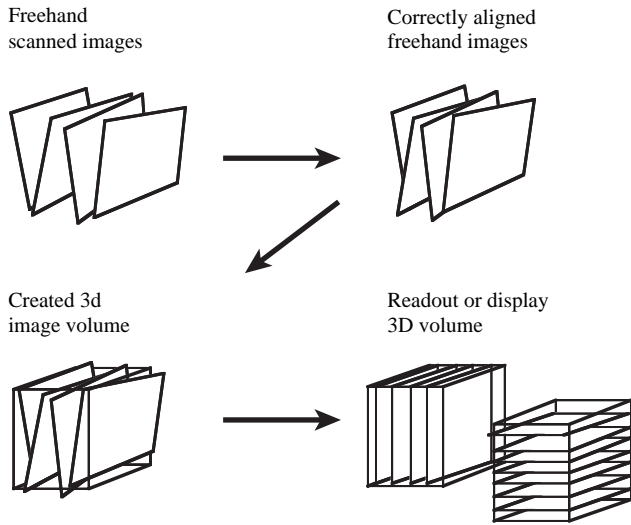


Figure 18. Illustration of 3D image reconstruction. Sets of spatial misaligned images are stacked according to their spatial position and orientation and used to fill a 3D image volume. Afterward this volume can be read out in any slice plane direction or even rendered as a 3D volume, such as shown in Fig. 19.

imaging. However, a limited number of clinical applications is approved by the Food and Drug Administration (FDA). As of 2004, the only FDA approved application for ultrasound contrast agents is for cardiac procedures, and more precisely, for outlining the border of the heart chamber. Other countries or regions such as Europe and Japan have a variety of agents approved. However, considerable research has been performed on ultrasound contrast agents, and it is likely that more FDA approvals will follow in the future. Contrast agents are used to enhance ultrasound image quality; therefore, imaging techniques implemented in current ultrasound scanners will be described.

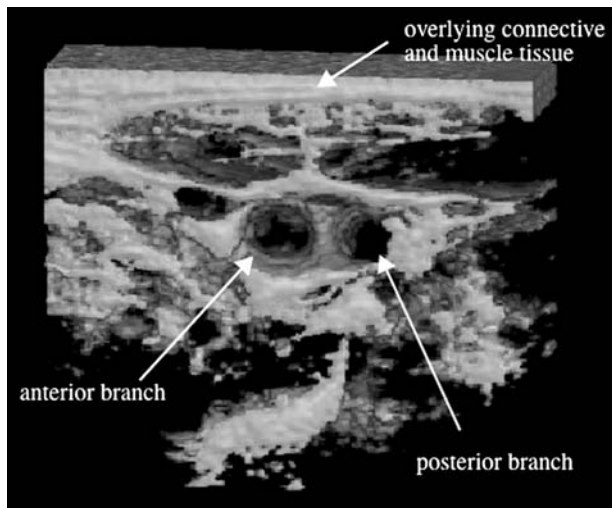


Figure 19. Three-dimensional reconstruction of the ascending carotid artery. This imaging mode uses 3D correlations and compounding to align individual images as the imaging array is swept across the vessel.

Moreover, the physics of contrast agents as well as their optimal clinical use will be discussed.

Clinical Background. Every year 135 million (Source: Amersham Health Inc. owned by GE Healthcare.) ultrasound scans are performed in American hospitals. Only 0.5% of these procedures actually use contrast-enhancing products. For example, better diagnosis of myocardial perfusion and better visualization of fine capillary-level vasculature will be possible when ultrasound contrast agents are certified by the FDA. Ultrasound is a relatively inexpensive imaging technique, and better diagnostic information can be obtained.

New contrast-agent-based ultrasound imaging modes include: Harmonic imaging, Pulse inversion (with harmonic or power mode), Microvascular imaging, Flash contrast imaging, as well as Agent detection imaging.

Enhancing Contrast. A major duty of ultrasound contrast agents is the improvement of ultrasound based image acquisition. The definition of contrast is given in Eq. 23, where I_1 and I_2 are the echo intensities before and after contrast administration, respectively. Even though this is a very simple formula, the mechanism for contrast improvement can be rather complicated.

$$\Lambda = \frac{I_2 - I_1}{I_1} \tag{23}$$

The key for contrast improvement for ultrasonic contrast agents lies in the physical principles of sound transmission, reception, and the nonlinear characteristics of bubbles themselves. For example, an increase in backscatter *amplitude* in the presence of contrast agents relative to the average human tissue backscatter level would improve the overall image. Furthermore, the creation of acoustic *frequencies* that occur in the backscatter signal of bubbles but not in the transmit signal nor in the backscatter of tissue, would provide a mechanism by which bubbles can improve the overall image. An important fact to keep in mind is that ultrasound contrast agents do not enhance the visibility of human tissue nor of blood, but the bubbles themselves can be visualized better than tissue or blood. Nevertheless, imaging perfusion of tissue or measurement of the amount of blood flowing through a vessel can be greatly improved by the usage of ultrasound contrast agents.

Modern Agents. Modern agents are not just gas bubbles. A sophisticated shell coating is used to prevent coalescence and reduce diffusion of the interior gas into the surrounding medium. This shell can be made of serum albumin. Lipids are also used as stabilizing agents. The gases filling the interior of the shell are chosen for low diffusion rates from the bubble into the blood stream as well as because of their low solubility in blood. Table 3 lists commercially available contrast agents. Currently FDA approved contrast agents include Imavist by Alliance Pharma/ Photogen, Definity by Bristol-Myers Squibb Medical Imaging Inc., Albutex by Molecular Biosystems, and Optison by GE/Amersham.

Table 3. Modern Ultrasound Contrast Agents^a

Manufacturer	Agent Name	Interior Gas	Shell Material
Acusphere	Al-700	Perfluorocarbon	Copolymers
Alliance Pharma. / Photogen	Imavist	Perfluorohexane, air	Surfactant
Bracco	SonoVue	Sulfur hexafluoride	Phospholipid
Bristol-Myers Squibb Medical Imaging, Inc.	Definity	Perfluoropropane	Lipid bilayer
	MRX-815-stroke	Perfluoropropane	Lipid bilayer
Molecular Biosys.	Albunex	Air	Albumin
	Oralex	Air	Dextrose
GE/ Amersham	Optison	Perflutren	Albumin
Nycomed Imaging AS	Sonazoid	Perfluorobutane	Lipid
Schering AG	Echovist	Air	Galactose
	Levovist	Air	Lipid layer
Sonus Pharma.	EchoGen	Dodecafluoropentane	Albumin
	SonoGen		Charged surfactant

^aSee Refs. 3,7, and 9.

Acoustic Bubble Response. Ultrasound contrast agents can be viewed as systems known as harmonic oscillators, with a given amplitude, phase, and frequency. Min-naert has derived Eq. 24, which gives the resonance frequency of a gas bubble as a function of its size. For example, a 3 μm radius (R_0) air bubble (adiabatic coefficient κ) in water (mass density ρ_L) under atmospheric pressure P_0 has a resonance frequency f of 1.1 MHz (Table 3). This is a very fortunate relationship since capillaries of the human circulatory system are as small as 8 μm in diameter and typical clinical frequencies used are 1–10 MHz.

$$f(R_0) = \frac{1}{2\pi} \frac{1}{R_0} \sqrt{\frac{3\kappa P_0}{\rho_L}} \quad (24)$$

The amount of acoustic scattering of the bubble surface (scattering cross-section σ_S) is described by the Rayleigh equation. This equation is used for scatterers that are small (μm) relative to the acoustical wavelength used (mm). For gas bubbles in water, the Rayleigh equation can be written with a series of mathematical terms for corresponding physical oscillation modes.

$$\sigma_S = 4\pi a^2 (ka)^4 \left[\left(\frac{\kappa - \kappa_0}{3\kappa} \right)^2 + \frac{1}{3} \left(\frac{\rho - \rho_0}{2\rho + \rho_0} \right)^2 \right] \quad (25)$$

The first term represents a monopole type bubble oscillation, whereas the second term describes a dipole term. One can see that the monopole term dominates the scattering due to the large compressibility (κ) difference between water and air–gas. Density differences (ρ) between water and air/gas are large too, however, the monopole term dominates the acoustic scattering (see Table 4).

Mathematical and Physical Modeling. The equation of motion of a bubble can be readily derived from an energy balance of kinetic (T) and potential energies (U) using the Lagrange formalism ($L = T - U$). It should be mentioned that the momentary inertial mass of the bubble as an oscillator does change with time. This is a major reason why ultrasound contrast agents are nonlinear systems, as

will be shown shortly. One of the first equations describing the motion of a gas bubble excited by ultrasound was derived by Rayleigh–Plesset and is given in Eq. 26. This formula is derived under the assumption that the interior gas follows the ideal gas law, and other forces acting on the bubbles are comprised of the internal vapor pressure p_d , the external Laplace pressure caused by the surface tension σ , the viscosity η_L of the surrounding host medium (water), the mass density ρ_L of the water, as well as the static p_0 and acoustic $p_\infty(t)$ pressures.

$$R(t) = \frac{1}{R(t)} \left(-\frac{3}{2} \left(\frac{\partial}{\partial t} R(t) \right)^2 + \frac{1}{\rho_L} \left(\left(p_0 + \frac{2\sigma}{R_0} - p_d \right) \left(\frac{R_0}{R(t)} \right)^{3\kappa} + \dots + p_d - \frac{2\sigma}{R(t)} - \frac{4\eta_L \frac{\partial}{\partial t} R(t)}{R(t)} - p_0 - p_\infty(t) \right) \right) \quad (26)$$

Figure 20 shows a simulation of the bubble response to a short tone burst excitation. Transmitted acoustic pressures of 1 and 50 kPa were simulated. Graph (a) shows a 1.1 MHz and 50 kPa pressure waveform as transmitted by a simulated ultrasound transducer. Graph (c) shows the subsequent radial oscillations of the simulated bubble (resting radius of 3 μm). Graph (b) shows the spectral response of these oscillations for a sound pressure of 1 kPa. The bubble oscillates mostly at the driving frequency of 1.1 MHz. An increase in sound pressure amplitude (i.e., 50 kPa) reveals the nonlinear nature of gas bubbles. In panel (d), in addition to 1.1 MHz one can also see higher harmonics of 2.2 MHz, 3.3 MHz, and so on.

Table 4. Scattering Coefficients for Monopole and Dipole Terms in Eq. 25 of a Water or Air Filled Sphere Under Water

Material	Bulk		Monopole Magnitude	Dipole Magnitude
	Modulus κ , MPa	Density ρ , $\text{kg} \cdot \text{m}^{-3}$		
Water	2250	1000	0	0
Air	0.14	1.14	2.9×10^7	0.33

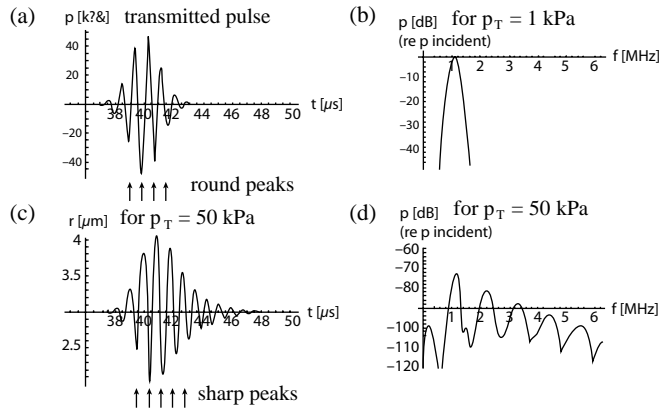


Figure 20. For higher transmit sound pressures (p_T) gas bubbles exhibit nonlinearities at similar magnitudes as the fundamental frequency [(b) vs. (d)]. Higher harmonics increase in their relative amplitude with sound pressure. Radial oscillations of the gas bubble contrast agent shows the nonlinear response during the compressional phase (arrows at small radii). Harmonic imaging takes advantage of the strong nonlinear character of contrast agent, which can be stronger than tissue depending on the number density of the bubbles.

Modern ultrasound contrast agents cannot be modeled using the free gas bubble Rayleigh–Plesset model. Elastic layer-based models presented by de Jong (5), Church (4), and Hoff (10) contain additional parameters, such as the mass density of shell material, a second surface tension term, a second viscosity term, or the elastic modulus of the shell material.

Imaging Modes

The following sections will cover imaging modes that rely on nonlinear backscatter, either originating from body tissue or due to nonlinear reflections from contrast agents. Each mode will be theoretically described and illustrated with examples.

Harmonic Imaging

Nonlinear tissue and contrast agent backscatter is the basis for harmonic imaging. Human tissue, as well as ultrasound contrast agent, can be driven in a nonlinear fashion such that the backscattered signal contains not only the original frequency f_0 , but also $2x f_0$, $3x f_0$, and so on (see Fig. 20d). Higher harmonics increase in their relative amplitude with sound pressure. Harmonic contrast imaging takes advantage of the strong nonlinear character of contrast agent, which is stronger than tissue. As a result, nonlinear backscatter from tissue will be smaller in amplitude than that of contrast agent and the vascular system will be visible over tissue. Subharmonic emissions are also characteristic for bubbles ($1/2x f_0$, $1/3x f_0$, etc.) and can be used to distinguish bubbles from tissue.

Figure 21 is comprised of a direct gas bubble simulation and an illustration of harmonic bubble behavior. As can be seen from the simulation results, harmonic scattering

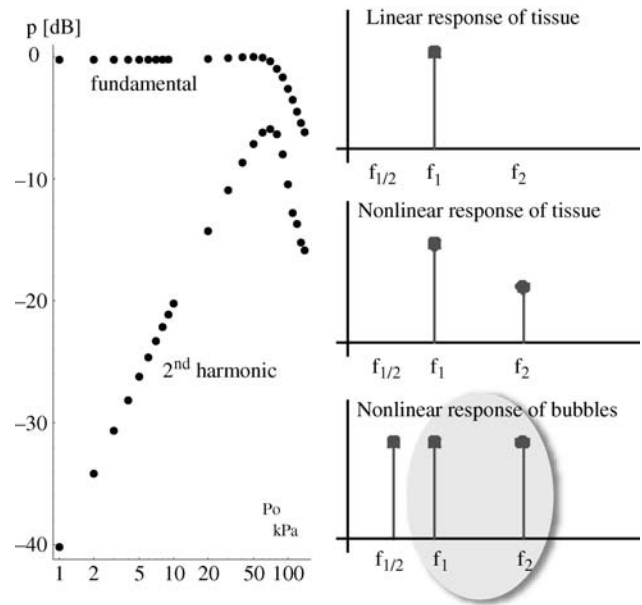


Figure 21. Higher harmonics increase in their relative amplitude with sound pressure (left), P_{harmonic}/P_0 . Harmonic imaging takes advantage of the strong nonlinear character of contrast agent, stronger than tissue, depending on the number density of the bubbles. Subharmonic emissions $f_{1/2}$ are also characteristic for bubbles.

of gas bubbles increases with increasing incident sound pressure. Specifically, the normalized amplitudes of the fundamental and the second harmonic frequency are shown here ($p_{\text{fundamental}}/p_0$ and p_{harmonic}/p_0). Increasing the incident sound pressure causes a proportional increase in the fundamental response, while the contribution of the harmonics increases more strongly. At 50–100 kPa, the backscatter amplitude of the second harmonics peaks and for higher pressures more and more energy is distributed over a wide range of frequencies, from sub-harmonics f_0/n to higher harmonics $n \cdot f_0$. Because the model as described in Eq. 26 does not take into account any losses, such as those induced by radiation or viscous damping, backscatter predictions for large excitation pressures will not be accurate. However, at modest amplitudes tissue scatters in a linear fashion, gas bubbles contribute harmonic signals, and the amplitude ration between fundamental and harmonic components can approach 1 for bubbles driven at sufficiently large amplitudes (illustration in lower right panel of Fig. 21).

Figure 20 shows the oscillation of a gas bubble in a large amplitude acoustic field. Panel (a) shows the excitation sound pressure waveform of 50 kPa amplitude. Radial excursions as well as the sound pressure spectrum at a distance of 5 cm from the bubble are plotted on the bottom panels (c) and (d), respectively. Most notably are the sharp peaks for bubble radii smaller than the initial bubble radius (see arrows in Fig. 20c). At positive incident sound pressures, the bubble is compressed and exhibits a large internal pressure. At the same time, the bubble’s resonance frequency changes. This change in resonance is the reason

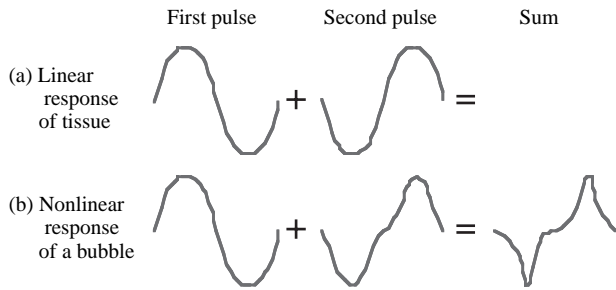


Figure 22. Tissue and contrast agent can yield different backscatter when exposed to an acoustic field of moderate intensity (typically of a mechanical index (MI) of less than 0.2). (a) Tissue responds in a linear fashion, that is, it yields a 0 and a 180° phase-shifted signal. The sum of both waves is zero. (b) For ultrasound contrast agent, the backscatter will include higher harmonics. A 180° phase shift in the transmit frequency f_0 will result in a 360° phase shift in the second harmonic $2f_0$. However, a 360° phase shift will result in the original signal. Therefore, the sum of a regular pulse and an inverted pulse will cancel for f_0 frequency components but not for $2f_0$ components.

for their highly nonlinear character. Panel (b) shows the scattered sound pressure spectrum for a low incident pressure of 1 kPa. No harmonic contributions can be seen within 40 dB of the pressure amplitude at the fundamental frequency.

Pulse Inversion Mode

As its name suggests, pulse inversion uses inverted pulses to gain contrast in the image. Figure 22 illustrates how this imaging mode works. As opposed to regular B-mode imaging, this mode requires *two* transmissions per image line. The first transmission does not differ from regular B-mode. However, the second transmission is 180° phase shifted, that is, inverted with respect to the first pulse transmitted. In case of tissue the backscatter for both pulses will mostly remain the same in magnitude. However, contrast agents respond with harmonic contributions which differ for 0 and 180° pulses. For very low acoustic pressures, this difference might not be distinguishable from the linear backscatter, but a mechanical index of 0.2 or less is sufficient to perform pulse inversion imaging. The mechanical index is a measure for the sound pressure amplitude and will be defined in the bioeffects section.

High sound pressure amplitude pulses can have two effects. First, they cause nonlinear tissue backscatter, thus reducing the contrast between tissue and contrast agent in the vascularity. Second, they destroy the contrast agent. One has to keep in mind that contrast agents are comprised of encapsulated gas bubbles. Therefore the bubble can shatter and the contained gas can subsequently dissolve into the blood pool. This shattering is possibly a source for bioeffects of ultrasonic imaging and will be discussed in a later section.

Theoretical modelling of the oscillatory behavior of gas bubbles using the above-introduced model directly shows the nonlinear response of gas bubbles. Figure 23a shows the two-pulse sequence required for pulse inversion ima-

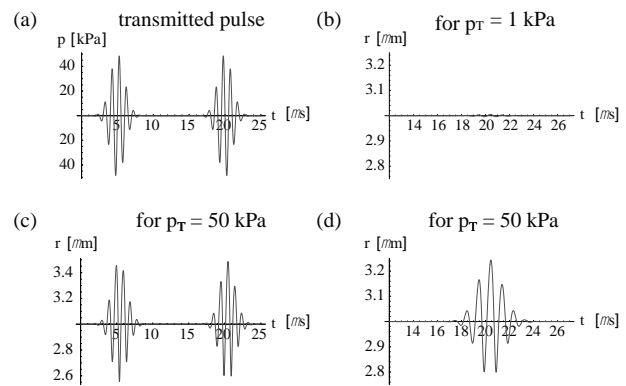


Figure 23. Two transmitted pressure waveform pulses with 180° phase shift with respect to each other, are displayed at time equal 5 and $20 \mu\text{s}$ in panel (a). This concept of inverted pulses is illustrated in Fig. 22(a). For 1 kPa sound pressure amplitude the sum of the oscillations of these two pulses is near zero (b). However, for a sound pressure of 50 kPa, the radial excursions of the contrast agent bubbles is nonlinear (as shown in Fig. 20) and the sum of the 0 and 180° signal is almost as great ($3.2 \mu\text{m}$ in d) as the oscillation itself ($3.4 \mu\text{m}$ in c). This technique is used in pulse inversion contrast agent detection scheme.

ging. For transmit pressures of 1 kPa, the difference in bubble response to 0 and 180° is very small. However, on the same scale, the response to a 50 kPa excitation yields two signals whose fundamentals f_0 cancel, but whose second harmonic contributions $2f_0$ add in phase (see Figs. 22 and 23 d).

A major downside of pulse inversion imaging is its sensitivity to motion. Spatial shifts cannot be distinguished from changes in back-scatter due to changes in transmit phase. A clever work-around to this problem is illustrated in Fig. 24, where a three-pulse transmit of varying phase is shown. Situations in which motion is anticipated or intestinal peristalsis or chest wall excursions during breathing. The latter is on the order of $2 \text{ cm} \cdot \text{s}^{-1}$, or for a 1 kHz PRF, $20 \mu\text{m}$ per firing. A signal of 1 MHz center frequency will experience a phase shift of 4.8° when spatially translated $20 \mu\text{m}$. It is assumed that the time duration in which the three pulses are fired is small compared to that of the body motion. If so, then the motion can be approximated as being linear on the time scale of the firings. The first firing is transmitted with zero phase, the second one with $5 + 180^\circ$, the third one with 10° phase shift. Linear response of tissue will result in backscatter signals of 0° , $5^\circ + 180^\circ$, and 10° phase shifted signals. Averaging the first and third backscatter signal will yield an average of 5° phase shift. Adding this average signal to the second backscatter signal will yield a value of zero for linear tissue, even while it is in motion. Bubbles, however, will yield a similar, nonzero, signal whether in motion or not.

Coded Excitation and Coded Harmonic Excitation. *Coded excitation* is a way to overcome poor signal to noise ratios (SNR) in ultrasonic imaging. It was already used in radar imaging for the same purpose, before its introduction to medical ultrasound. The most simple solution for poor

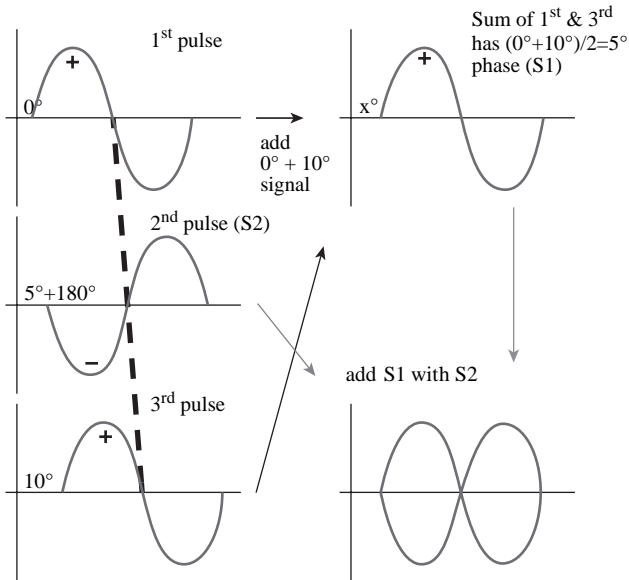


Figure 24. A three pulse sequence with a linear shift in transmit phase (0°, 5°+180°, and 10°) is used to suppress tissue signal, even when the tissue is in motion. The 0 and 10° backscatter signals are averaged to yield a 5° signal from tissue (S1). The 5° phase-shifted and inverted transmit signal (S2) is then added to the averaged signal S1. For moving tissue the result will be zero. If the tissue motion is linear during the time span of the three transmit pulses (~1 ms), then pulse will be separated by an offset of x°. This offset is in addition to the initial 5° offset between first & second and second and third pulse. Therefore tissue will yield zero for S1+S2. Ultrasound contrast agents, however will yield S1≠S2 as shown in Figs. 22 and 23.

signal/noise ratios is to increase the amplitude of transmitted sound. However, the FDA regulates sound pressure amplitudes because of the likelihood of acoustic cavitation in the presence of large sound pressure amplitudes. Moreover, the number of acoustic transmissions per unit time (pulse repetition interval, PRI) as well as the length of individual pulses (burst length) are regulated by means of the overall deposited energy that will eventually result in tissue heating. Both effects will be discussed in the Bioeffects section below. In addition to regulatory and safety concerns, increasing the number of cycles in a traditional transmit pulse also reduces the axial resolution.

However, improvements in signal to noise ratios without sacrificing axial resolution can be achieved despite regulations on pressure amplitudes and burst lengths. Transmission of specially designed and unique signals can significantly improve their detectability. This concept is called Coded harmonic excitation.

Various code types shall be discussed here to illustrate how Coded excitation works. Figure 25 illustrates two codes. On the left side is the most simple code, a so-called pulse train (or uncoded tone burst), that is, a series of pulses or sinusoids. When transmitting four cycles at a certain frequency one can use a frequency filter that is sensitive at the transmit frequency, but only for signals that are four cycles long. A more complicated type of coded excitation is the code shown on the right side of Fig. 25.

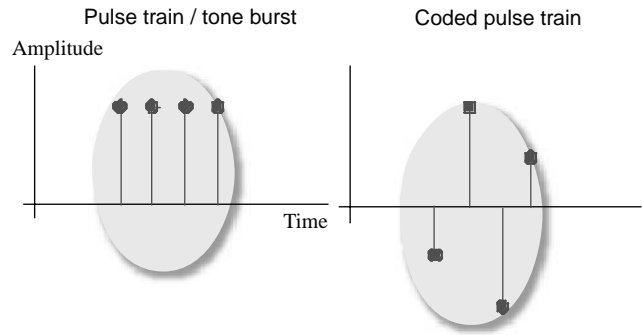


Figure 25. Some signal to noise problems can be overcome by the use of coded excitation. Instead of a single pulse, a pulse train (commonly also known as a tone burst) or coded pulse train is transmitted for better backscatter detection. See Fig. 26 for temporal /spatial resolution.

This signal also transmits four cycles, but now the four cycles all differ in sign as well as in amplitude. Both features make this code more unique and therefore more detectable when ambient noise lowers SNR.

Above mentioned frequency filters are explained in Figs. 26 and 27. The first column in Fig. 26 illustrates the transmit pressure waveforms and the second column shows the receive filter used. A mathematical technique termed convolution is used to match the transmitted waveform with the anticipated receive signal by means of an appropriately designed filter. The average reader might not have adequate signal processing background to be familiar with this concept. Therefore, Fig. 27 will be used to explain Coded Excitation for the example of Golay code

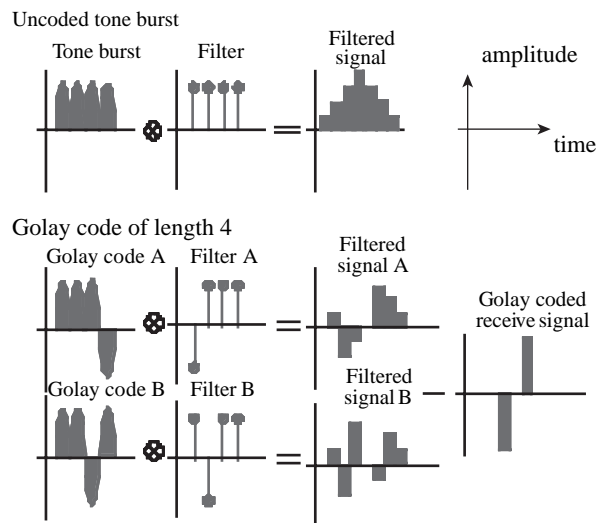


Figure 26. Coded excitation is implemented on modern ultrasound scanners to overcome signal loss and low SNR. A simple solution is the increase in burst length. However, this simultaneously results in a decrease in axial/temporal resolution. Specially coded waveforms are designed for an increase in SNR by reasonably maintaining spatial resolution. Golay codes, for example, can be designed for a 10-fold increase in SNR by losing only a factor of two in temporal resolution.

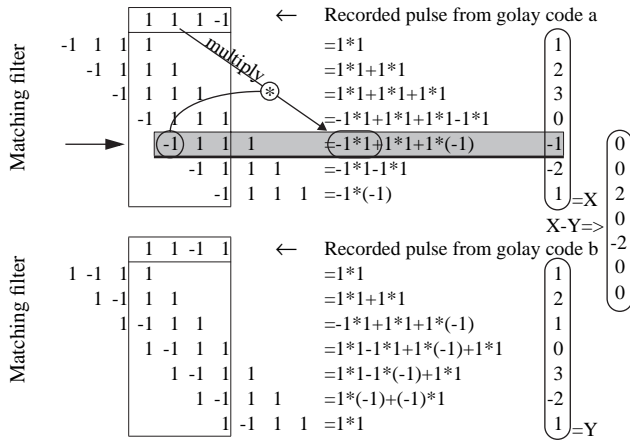


Figure 27. The mathematical process of convolution is described in this figure. For this purpose the reader should refer to the Golay code of length 4 given in Fig. 26 and to the text section on Coded excitation for a detailed description.

of length 4 (see Fig. 26). This code consists of two transmit pulse packages. The first and second packages are four pulse sequences of $[+1, +1, +1, -1]$ and $[+1, +1, -1, +1]$, respectively. Such pulses are shown on the top row of each box in Fig. 27. The pulses are transmitted into the body and the resulting echoes, which should show the same pattern, are recorded. Matching filters are shown below the pulse packages and have been replicated at a total number of 7 positions. Each filter replication has an overlap with the transmit pulse ranging from 1 to 4 digits. For a given position, numbers in the same column are multiplied and then the resulting values of the overlapping columns are added.

For clarification the highlighted case in the top system for position 5 (arrow) is explained in detail. The filter $[-1, 1, 1, 1]$ and the recorded pulse $[1, 1, 1, -1]$ overlap for three digits and the computational method of convolution is equivalent to computing the inner product of two vectors: $[1, 1, -1] \cdot [-1, 1, 1] = 1 \cdot (-1) + 1 \cdot 1 + (-1) \cdot 1$, which sum to -1 . This value is displayed in the column on the right. This convolution procedure is done for both transmits. Subtracting the two resulting 7 digit vectors $X = [1, 2, 3, 0, -1, -2, 1]$ and $Y = [1, 2, 1, 0, 3, -2, 1]$ will yield $[0, 0, 2, 0, -2, 0, 0]$, a receive vector (called Golay-coded receive signal in Fig. 26) with good temporal/spatial resolution. In the given case it is essentially three time steps long: $[-2, 0, 2]$. The uncoded tone burst is 7 time steps long (see top of Fig. 26).

Back to Fig. 26, the result of the convolution is shown in the third column. For the uncoded tone burst one can see a triangular shaped signal that peaks where the receive filter perfectly matches the incoming pressure wave. The temporal duration of the convolution increases as the duration of the tone burst increases. This unwanted effect of loss of spatial resolution is overcome by the use of more sophisticated codes. Golay transmit codes and receive filters, for example, can mostly maintain spatial resolution while improving SNR. Two differing transmit codes are used and the resulting convolutions are subtracted from each other, similar to pulse inversion imaging.

Coded excitation in the form of a Chirp-type pulse train yields a signal to noise gain of $\sqrt{\Delta F T}$; for example, a signal with a 1 MHz frequency sweep and $T = 16 \mu\text{s}$ pulse duration will yield a fourfold increase in SNR, which corresponds to 12 dB. Chirps are tone burst of continuously increasing transmit frequencies within the pulse train. Continuously in frequency decreasing pulse trains are called chirps, which almost spells chirp backward. General Barker and Golay codes perform well too. For a standard Barker sequence the side lobe amplitudes of the transmit/receive signal are -22 dB down relative to the amplitude main lobe. A Golay code with 16 pulses per burst yields a ten-fold increase in SNR. Here, only a length-4 sequence is shown for clarity (see Fig. 26).

More detailed mathematical background, as well as clinical feasibility can be found in publications by Nowicki (19) and Misaridis (20), for example.

BIOEFFECTS

Ultrasound related bioeffects are centered on two main categories: cavitation and heating. Acoustic output (and therefore to some degree acoustic cavitation) is regulated by the FDA through the implementation of the mechanical index (MI), as well as the temperature index (TI) in the so-called Output Display Standard (Standard for Real Time Display of Thermal and Mechanical Acoustic Output Indices on Diagnostic Ultrasound Equipment (1992), published by the National Electrical Manufacturers Association as UD-3). The output display standard is used to obtain approval by the FDA for medical use of ultrasound scanners.

Cavitation is a well-studied effect and manifests itself, for example, by the erosion of ship propellers and for therapeutic purposes in lithotripsy. Hydrodynamically generated microscopic gas bubbles imploding in close proximity to the metallic surface of the propellers punch small holes into the metal. This occurs due to the existence of water jets generated by asymmetrically imploding gas bubbles.

To date, there are only very limited clinical reports on the occurrence of bioeffects in diagnostic ultrasound. Nevertheless, *in vitro* cell cultures and animal models are used to study bioeffects at acoustic parameters beyond the limits for diagnostic imaging. No limits for single parameters (e.g., pressure amplitude, wave frequency, or pulse length) were identified *per se* due to the complicated biological end points (21). However, two fundamental quantities, MI and TI, were introduced as parameters by which to judge the probability of bioeffects (11,21). Both quantities gain in presence, especially with the MI being used by the FDA to limit the acoustic output of scanners.

$$\text{MI} = \frac{p[\text{MPa}]}{\sqrt{f[\text{MHz}]}} \quad \text{TI} = \frac{Wp[\text{Nm} \cdot \text{s}^{-1}]}{W_1^2[\text{Nm} \cdot \text{s}^{-1}]} \quad (27)$$

Pressure p and frequency f are quantities that can be directly measured experimentally. Power W_p , however, is a derived quantity and is related to the pressure waveform as

shown in Eq. 28 (22).

$$PII = \frac{\int_{t_1}^{t_2} v_h(V(t))^2 dt}{10^4 \rho c M_L (f_c)^2} [J \cdot cm^{-2}] \tag{28}$$

$$I_{SPTA} = PII(j, v, z_{m,jPII}) \times PRF [W \cdot cm^{-2}]$$

The pulse intensity integral *PII* is computed over a truncated duration of the acoustic wave, namely, from *t*₁ to *t*₂. These times correspond to the 10 and 90% indexes of the accumulated total energy in the tone burst and can in most circumstances be regarded as the burst length. When calibrating an acoustic transmitter, one often measures an electric waveform produced by a underwater microphone, called a hydrophone. A conversion function *v_h*, is then used to determine the actual pressure of the measured electric waveform *V(t)*. *M_L* is the hydrophone’s frequency dependence. Together, *v_h*, and *M_L* allow the complete conversion of the electrical hydrophone signal to the pressure of the measured wave. Mass density and speed of sound of the medium in which the wave travels are labeled as *ρ* and *c*. Even when the burst length is taken into account, one does not know the total transmitted power without the PRF. Spatial peak and temporal average intensities (*I_{SPTA}*) are computed with knowledge of the PRF, yielding units of watts per centimeter squared (*W · cm⁻²*). For computing the TI, however, one needs to multiply that value with the geometric cross-section of the acoustic beam. In reality, the computation of the TI is more complicated and takes into account more parameters such as the chosen exposure conditions, as well as nominal perfusion parameters. In clinical use, perfusion parameters are estimated from user selected target tissue types.

An additional step toward reducing the possibility of bioeffects is the usage of the principle of ALARA (as low as reasonably achievable), that is, the reduction of acoustic output and reduction of exposure time to the lowest reasonable minimum. Typical diagnostic acoustic procedures operate at or below the FDA limit for *I_{SPTA,3}* of 720 mW · cm⁻² (see Table 5 and (23)). Currently, the FDA regulates output limits via Tracks 1 and 3 of 510(k). Track 1 uses the *I_{SPTA}* levels shown in Table 5, whereas Track 3 allows device manufacturers to increase their output to a general 720 mW · cm⁻² if their device provides user feedback via MI and TI display standards (24).

Biological effects of ultrasound can be investigated after the acoustic output of a sound source is quantified and

qualified as described in the previous paragraphs. *In vitro* cell studies are typically the first method for research on the biological effects of ultrasound. Such studies can be performed in a highly controlled environment and are therefore repeatable. This is not necessarily the case for patient studies where many factors cause unavoidable variabilities.

It was found in a study on anesthetized rats that the interaction of the incident sound field with ultrasound contrast agent can cause pete-chial hemorrhages (punctuate sites of bleeding from blood vessels) in heart tissue. In a study by Li et al. (13,14) ultrasound contrast agents were injected in a similar fashion as is used in clinical contrast echocardiography. Heart tissue was observed in real time using a phased array ultrasound scan head (1.7MHz transmit frequency) operating in a harmonic mode native to the clinical scanner used. Postmortem heart tissue analysis showed that bleedings scaled monotonically, proportionally to the square of the peak rarefactional pressure (*P_{rare}*) amplitude of the sound field. Pressure amplitudes of 0.6–1.8 MPa *P_{rare}* were used in that study. This pressure range corresponds to MIs between 0.5 and 1.5. Real-time ultrasonic imaging showed that premature ventricular contractions (PVC) were triggered by ultrasound in the presence of contrast agents. However, no significant PVC were observed for MIs of 0.5 and 1.0, but up to 40 PVCs were observed for a 3 min exposure at a MI of 1.9, the maximum allowed by the FDA.

Currently, there are very limited clinical reports on unanticipated bioeffects of ultrasound. Intended effects exist since ultrasound can be used to treat kidney stones and during such procedures it is likely that tissue bleeding occurs. There have been reports of effects associated with diagnostics, but the review of these reports has yielded no establishment of a causal relationship with ultrasound exposure. However, a study by van der Wouw in healthy male volunteers concludes: "Imaging of contrast agents with high acoustic pressures can cause PVCs if end-systolic triggering is used. This effect is related to both the dose of contrast agent and acoustic pressure. It does not occur during end-diastolic triggered imaging. Precautionary measures would include using lower MIs or end-diastolic triggering" (25).

It has been seen that the level of bioeffects varied with the ultrasound contrast agent used. This is probably due to different shell materials and internal gases, which cause the bubbles to oscillate at different amplitudes. If agent rupture occurs, cavitation type damage in produced in the surrounding tissue (13,14). Thermal and nonthermal effects exist and they are under investigation. According to Natori (26) temperature increases of no more than 1.5°C above normal are considered clinically acceptable and nonthermal, that is, cavitation based effects can only be found where gas bodies are present, such as postnatal lung and intestines, or via ultrasound contrast agents. Therefore regular B-mode imaging is considered by many unaffected (26).

Finally, it should be mentioned that there are no bioeffects in the absence of contrast agent and also no bioeffects in the presence of contrast agents but absence of an ultrasonic field. Moreover, low doses (10–50 μL · kg⁻¹)

Table 5. Acoustical Output Limits for Clinical Ultrasound Scanners per FDA Regulation

Tissue	<i>I_{SPPA,3}</i> , mW · cm ⁻²	<i>I_{SPTA,3}</i> , W · cm ⁻²	MI
Peripheral vasculature	720	190	1.9
Cardiac	430	190	1.9
Fetal and other	94	190	1.9
Ophthalmic	17	28	0.23

^aIntensity values are derated for tissue attenuation with an acoustic attenuation of 0.3 dB · cm⁻¹ · MHz⁻¹.

^b*I_{SPTA}* and MI are defined in the text, *I_{SPPA}* is defined as the intensity of the spatial peak pulse averaged waveform.

of ultrasound contrast agent yielded little if any bioeffects. In general, radiological contrast agents, or the imaging procedure itself, may bear the risk of bioeffects. That risk has to be balanced with the medical need for the procedure.

ACKNOWLEDGMENTS

I wish to thank Katy, Kim, Mario, John, and Jessi for their invaluable efforts for proof reading and their technical comments. Also, I would like to thank the reviewers for their contributions and very helpful commentary. Most of all, I wish to thank my wife Naki for her patience during many lost hours of family time.

BIBLIOGRAPHY

- Angelsen BAJ. *Ultrasound Imaging Waves, Signals, and Signal Processing*; 2000.
- Averkiou M, et al. Ultrasound contrast imaging research. *Ultrasound Q* 2003;19(1):27–37.
- Becher H, Burns PN. *Handbook of Contrast Echocardiography—LV Function and Myocardial Perfusion*. Berlin: Springer-Verlag; 2000.
- Church CC. The effects of an elastic solid surface layer on the radial pulsations of gas bubbles. *J Acoust Soc Am* 1995; 97(3):1510–1521.
- deJong N, et al. Absorption and scatter of encapsulated gas-filled microspheres: Theoretical considerations and some measurements. *Ultrasonics* 1992;30:95–105.
- Duck FA. *Physical Properties of Tissue*. San Diego: Academic Press; 1990.
- Feinstein SB. The powerful microbubble: From bench to bedside, from intravascular indicator to therapeutic delivery system, and beyond. *Am J Physiol Heart Circ Physiol* 2004;287: H450–H457.
- Gray H, Williams PL, Bannister LH. *Gray's anatomy: The anatomical basis of medicine and surgery*. New York: Churchill Livingstone; 1995.
- Grayburn PA. Current and future contrast agents. *Echocardiography* 2002;19(3):259–265.
- Hoff L. Acoustic characterization of contrast agents for medical ultrasound imaging. Ph.D. dissertation at the Norwegian University of Science and Technology in Trondheim, Norway; 2000.
- Holland CK, Apfel RE. Thresholds for transient cavitation produced by pulsed ultrasound in a controlled nuclei environment. *J Acoust Soc Am* 1990;88(5):2059–2069.
- Kasai C, Namekawa K, Koyano A, Omoto R. Real-time two-dimensional blood flow imaging using an autocorrelation technique. *IEEE Trans, Ultrasonics, Ferroelectrics, Frequency Control* 1985;SU-32(3):458–464.
- Li P, Cao et al. Impact of myocardial contrast echocardiography on vascular permeability: An in vivo dose response study of delivery mode, pressure amplitude and contrast dose. *Ultrasound Med Biol* 2003;29(9):1341–1349.
- Li P, Armstrong WF, Miller DL. Impact of myocardial contrast echocardiography on vascular permeability: Comparison of three different contrast agents. *Ultrasound Med Biol* 2004; 30(1):83–91.
- Lindsay RB. The story of acoustics. *J Acoust Soc Am* 1965; 39(4):629–644.
- Shung KK, Zipparo M. Ultrasonic transducers and arrays. *IEEE Eng Med Biol* 1996; 20–30.
- Angelsen BAJ, et al. Which transducer array is best. *Eur J Ultrasound* 1995;2:151–164.
- McDicken WN, Evans DH. *Doppler Ultrasound: Physics, Instrumentation and Signal Processing*. New York: John Wiley & Sons, Inc.; 2000.
- Nowicka A, Litniewski J, Secomski W, Lewin PA. Estimation of ultrasonic attenuation in a bone using coded excitation. *Ultrasonics* 2003;41:615–621.
- Misaridis TX, et al. Potential of coded excitation in medical ultrasound imaging. *Ultrasonics* 2000;38:183–189.
- Abbott JG. Rationale and derivation of Mi and Ti—A review. *Ultrasound Med Biol* 1999;25(3):431–441.
- American Institute of Ultrasound in Medicine (AIUM). *Standard Specification of Echoscope Sensitivity and Noise Level Including Recommended Practice for Such Measurements*; 1978.
- O'Brien WD, et al. Acoustic output upper limits proposition. *J Ultrasound Med* 2002;21:1335–1341.
- Nyborg WL. History of the American Institute of Ultrasound in Medicine's efforts to keep ultrasound safe. *J Ultrasound Med* 2003;22:1293–1300.
- van derWouw PA, et al. Premature ventricular contractions during triggered imaging with ultrasound contrast. *J Am Soc Echocardiogr* 2000;13(4):288–294.
- Natori M. Ultrasound safety: Overview and what we do need in daily clinics for a safe use of diagnostic ultrasound. *Inter Congress Ser* 2004;1274:125–128.

See also COMPUTED TOMOGRAPHY; ECHOCARDIOGRAPHY AND DOPPLER ECHOCARDIOGRAPHY; IMAGING DEVICES; MAGNETIC RESONANCE IMAGING.

ULTRAVIOLET RADIATION IN MEDICINE

J. J. LLOYD
Regional Medical Physics
Department
Royal Victoria Infirmary
Newcastle-upon-Tyne,
United Kingdom

INTRODUCTION

It is to the philosophers and physicians of the ancient civilizations that we should attribute the earliest history of ultraviolet radiation (UVR) in medicine (1). For example, the Greek sun god, Apollo, was also the spiritual god of healing, providing the first documentation of an association between sunlight and health. In 525 BC, Herodotus observed that the strength of a human's skull was related to sunlight exposure, >2000 years before the formal discovery of the role of sunlight in vitamin D metabolism. At about the same time the Egyptians were using psoralens from plant extracts and sun exposure in the treatment of vitiligo. However, it was not until Jonathan Ritter in 1801 discovered the UV region in the solar spectrum that the science of photobiology could really begin.

The Danish physician Niels Finsen (1860–1904) is regarded by many as the father of modern UV therapy. In a series of articles published between 1893 and 1896,



Figure 1. Patients receiving sun heliotherapy for the treatment of tuberculosis. (Picture courtesy of the University of Denver.)

Finsen stressed that it was the UVR in the solar spectrum that was responsible for sunburn and not the radiant heat, as the name implies. In parallel with his scientific investigations, Finsen was also an active clinician. He is best remembered for his successful treatment of *lupus vulgaris* (tuberculosis of the skin, mainly on the face) and in 1903 was awarded the Nobel Prize for medicine in recognition of this work. The photograph shown in Fig. 1 shows patients with *lupus vulgaris* being treated with sunlight (heliotherapy). Treatment of disease using artificial sources of UVR is known as actinotherapy or phototherapy.

Following the pioneering work of Finsen, the early part of the twentieth century saw the rapid expansion of heliotherapy and actinotherapy throughout Europe and the United States. The practice of actinotherapy continued to expand through the middle part of the twentieth century and was accompanied by an enormous literature on the subject during the 1920s and 1930s. This rapid growth is reflected by the many revisions of the handbook *Actinotherapy Technique* first published by the Sollux Publishing Company in 1933 and reprinted for the ninth time (7th ed.) in 1949 (2). (A copy of this can be seen on-line at <http://www.meridianinstitute.com/eaem/hanovia/hanocont.html>) Most of the irradiation protocols for the countless number of diseases, such as diabetes and angina, described in this book are now of historical interest only. The advent of effective antibiotics and the realization that the successes claimed in many of these diseases were little more than anecdotal have resulted in a more limited role of UVR in clinical medicine. However, today there is considerable interest in photobiology, both in treatment of diseases such as psoriasis, and in research into the photobiological basis of skin aging, carcinogenesis, and photodermatoses. Greater leisure time and ease of travel has led to potential for greater exposure to solar UV and greater potential for short- and long-term deleterious effects. This article presents a review of photobiology with particular emphasis on sources and measurement of UV relevant to medical diagnosis and treatment, but also encompassing

biological effects of UV, natural UV exposure, medical applications, and hazard assessment.

THE ULTRAVIOLET SPECTRUM

Ultraviolet radiation is part of the electromagnetic spectrum and lies between the visible and the X-ray regions. Different wavelengths in the UV spectrum show enormous variations in ability to cause biological damage, and for this reason the UV spectrum is divided into three spectral regions: UVA, UVB, and UVC. The notion to divide the UV spectrum into different spectral regions was first put forward at the Copenhagen meeting of the Second International Congress on Light held during August 1932. It was recommended that three spectral regions be defined as follows:

UVA 400–315 nm

UVB 315–280 nm

UVC 280–100 nm

The subdivisions are arbitrary and differ somewhat depending on the disciplines involved (3). The boundary between UVA and UVB is sometimes set at 320 nm and that between UVB and UVC is sometimes regarded as 290 nm. The short wavelength limit for UV is sometimes quoted as 100 nm and sometimes as 200 nm. The UVA region has recently been divided into UVAI (340–400 nm) and UVAII (320–340 nm).

Due to potential confusion with this terminology and because of rapidly changing biological effects as a function of wavelength, it is recommended that, in publications in photobiology, bandwidths are quoted explicitly and ideally the full spectrum of UV sources is described.

BIOLOGICAL EFFECTS OF ULTRAVIOLET RADIATION

An understanding of biological effects of UV is vital in appreciating the requirements for sources and detectors in medical applications. As UV does not penetrate tissue readily, it is the eyes and skin that are organs of particular interest. Comprehensive reviews of the health effects of UV are given in a recent book from the National Radiation Protection Board in the United Kingdom (4). Briefer reviews can be found elsewhere (5,6). In this section, the effects of UV in normal subjects is briefly described; effects in disease are discussed in later sections.

In order to cause a biological effect, UVR must be absorbed and initiate a photochemical process. The biological molecule that absorbs the radiation is known as a chromophore. A plot of the effectiveness of a chromophore at absorbing radiation as a function of wavelength is the absorption spectrum. A plot of the effectiveness of the UVR of different wavelengths in causing a given biological effect is called an action spectrum. The shape of the action spectrum will depend both on the absorption spectrum for the chromophores initiating the effect and also the optical properties of the skin that influences the radiation reaching the chromophores. The transmission of UVR

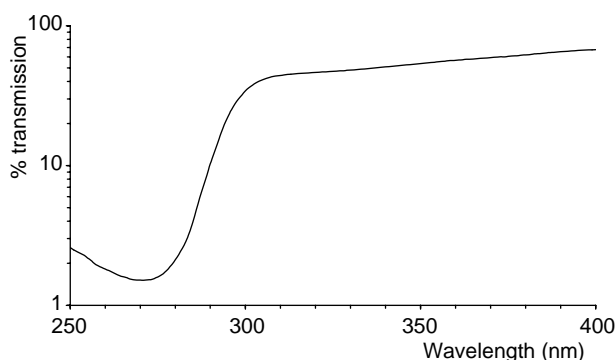


Figure 2. Fractional UVR penetration through the epidermis as a function of wavelength (B.6 L. Diffey, personal communication).

through the epidermis is shown in Fig. 2; UVA has much higher transmission than UVB or UVC, and therefore penetrates deeper into the skin.

Effect of UV Exposure on the Eyes

Both UVC and UVB radiation is predominately absorbed by the conjunctiva and cornea. The lens absorbs radiation in the range 300–370 nm, whereas radiation in the range 400–1400 [visible blue light to near-infrared (IR) radiation] is absorbed on the retina. Acute effects of UVR overexposure are photokeratitis (corneal injury) and photoconjunctivitis. Photokeratitis may occur in mountaineers as a result of high solar UV at altitude and high UV reflection from snow; this is referred to as “snow-blindness”. Many epidemiological studies suggest a link between chronic ocular UV exposure and ocular disorders such as cataracts and the appearance of yellow/brown deposits in the cornea. Absorption in the UVC range is of little interest in epidemiological studies since there is no natural exposure at these wavelengths. However, corneal absorption is very high between 190 and 220 nm and radiation in this range is used therapeutically in laser photorefractive surgery. Also certain artificial sources, such as germicidal lamps, emit significant UVC, and therefore care is required in their use.

Effect of UV Exposure on the Skin

The acute effects of UVR on the skin include erythema (sunburn), skin thickening, tanning, and vitamin D production. Long-term effects include induction of skin cancer and premature skin aging. Individual response to UVR varies greatly. Fair skinned people burn more easily in the sun, find it difficult to obtain a tan, and have a higher risk of skin cancer compared to darker skinned individuals. In photobiology it is usual to categorize skin into six types, as shown in Table 1.

Acute Effects. Erythema is skin redness caused by inflammation and dilation of small blood vessels. Despite being extensively studied the underlying cause of UV induced erythema is poorly understood (7). The erythema response of the skin in an individual can be defined by the minimum dose required to produce a just perceptible redness. This is referred to as the Minimal Erythema Dose

Table 1. Characteristics of Different Skin Phototypes^a

Skin Type	Skin Color	Sensitivity to Sunburn	Ability to Tan	Skin Cancer Risk
I	White	Very high	Virtually nil	High
II	White	High	Poor	High
III	White	Medium	Good	Medium
IV	Olive	Low	Very good	Low
V	Brown	Very low	Very good	Very low
VI	Black	Very low	Very good	Very low

^aFrom Ref. 65.

(MED). A standard erythema action spectrum has been defined taking into account many published studies (8) (see Fig. 3). Skin sensitivity is maximum to radiation in the UVC and UVB range up to 298 nm. At longer wavelengths, the sensitivity drops rapidly reducing to be about 10,000 times less at 400 nm. The MED is a threshold measurement made by visual assessment. For quantitative studies a device measuring reflectance (11) can be employed to quantify erythema. This device works by measuring the decreased green reflectance, relative to red, from haemoglobin in the dermal blood vessels.

Within an individual the sensitivity to erythema can vary considerable from site to site (12,13). Minimal erythema dose measured on the forearm may be twice that on the back, and the buttock skin is more sensitive than the back.

Chronic Effects. There are three types of skin cancer associated with UVR: squamous cell carcinoma (SCC), basal cell carcinoma (BCC), and malignant melanoma (MM). The SCCs appear as persistent red crusted lesions on exposed sites and have an incidence about a quarter of BCC. The BCCs appear as raised translucent nodules, normally on the face. Basal cell carcinoma appears to be related to cumulative UVR exposure, whereas SCC and MM may be related to intermittent exposure to UVR. In support of this idea is that fact that BCC and SCC tend to occur on habitually sun-exposed sites whereas MM occurs more commonly on intermittently exposed sites. Although MM has a much lower incidence than non-melanoma skin cancer, it accounts for ~80% of all skin cancer deaths.

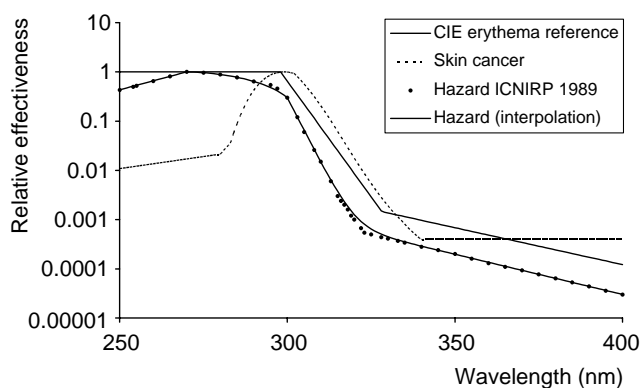


Figure 3. Standard action spectra for erythema (9), non-melanoma skin cancer (10), and UV hazard (6).

Melanoma incidence is increasing rapidly, probably due to increased recreational sun exposure, although other non-UV factors also play a part. Since the 1970s melanoma has seen the largest increase in incidence compared to all other major cancers in the United Kingdom and now stands at 8/100,000 person-years. It is predicted that the incidence will continue to rise, perhaps doubling over the next 30 years in the United Kingdom (14).

Photoageing is the term used to describe features of chronically sun exposed skin. These features include: wrinkles, “age spots”, and thick leathery skin.

Effects on the Immune System

There is some evidence that exposure to UVR may have a systemic immunosuppressive effect. It is possible that UV exposure may increase the incidence of infectious diseases and play a role in the promotion as well as initiation of skin cancers.

ULTRAVIOLET SOURCES

Ultraviolet radiation is emitted during transition of a molecular electron from an excited, high energy, to a less energetic state. As the possible energy levels are fixed for a given molecule, radiation is emitted with distinct photon energies. When excitation is by heating, the release of radiation is termed incandescence. Alternatively, excitation can be generated by passage of an electrical discharge through a gas. This is the basis for most artificial UV sources used in medicine, for example, mercury arc lamps and fluorescent UV lamps. The UV emitting LEDs (light emitting diodes) are available, although these tend to have low output and be restricted to long wavelength UV. A further recent development has been the production of UV emitting lasers.

Solar Ultraviolet Radiation

Although direct use of solar radiation as a medical therapy (heliotherapy) is no longer employed, an understanding of solar UV can be helpful clinically, particularly in the investigation of abnormal reactions to sunlight. In clinical phototesting solar simulating sources may be used. A brief summary is therefore given here, for more information the reader is directed to various recent reviews (15–18).

The spectrum of extraterrestrial solar radiation approximates that of a black body at ~5800 K. At the earth's surface, this spectrum is modified by atmospheric attenuation. The stratospheric ozone layer prevents almost all radiation with wavelengths <290 nm and a substantial proportion of UVB (70–90%) reaching the earth. When the sun is lower in the sky, the path length through the atmosphere is greater. The UV intensity is therefore reduced at all wavelengths, but more so for UVB than UVA. The relative intensity of UVA compared to UVB is therefore greater in the winter than summer in the United Kingdom. At mid-day in the summer in the United Kingdom the ambient total UV is ~4 mW/cm², of which UVB contributes ~5%. However, due to its greater deleterious effect, UVB contributes to ~80% of the harmful effects of solar UV.

Incandescent Sources

Incandescent sources emit a smooth broad spectrum of radiation with the peak wavelength inversely related to the absolute temperature. Conventional tungsten bulbs used for domestic lighting have peak emission in the infrared (IR) region and emit very little UVR. Tungsten halogen bulbs operate at higher temperatures and may produce rather more UVR. These sources are not used for medical applications due to the low UV output, although, they may be used as reference sources for UV meter calibration.

Mercury and Metal Halide Arc Lamps

The radiation emitted from a mercury-vapor arc lamp arises from two mechanisms. Line or characteristic radiation is produced as a result of excitation of the constituent atoms, together with a spectral continuum that is chiefly due to ion and electron recombination. Lamps can be produced that operate at different pressures. A low pressure mercury arc lamp consists of a fused silica tube filled with argon at ~1 Pa and containing a drop of mercury. A discharge occurs between the electrodes sealed into the ends of the tube. More than 90% of the radiant energy produced by the discharge is at 253.7 nm. Various other characteristic mercury spectral lines occur, for example, at 313 and 365 nm, and these sources are useful for wavelength calibration of spectroradiometers.

If the pressure in the lamp tube is increased, then the spectral lines broaden and also more radiation occurs in the continuum (see Fig. 4). The “alpine sunlamp” is a type of medium pressure arc lamp that was widely used for phototherapy, but has since been superseded by fluorescent lamps. The addition of metal halides to a high pressure mercury discharge lamp greatly enhances the UV output. These lamps were common in phototherapy departments at one time, but have now been mostly replaced by fluorescent UVB lamps. These lamps are still occasionally used for cosmetic tanning. By incorporating optical filters between the lamps and the irradiated subject, absorption of the UVC or UVC and UVB components can be achieved (see Fig. 5).

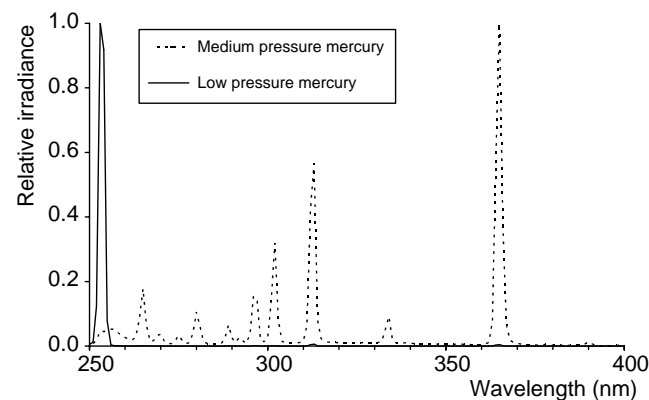


Figure 4. Ultraviolet emission spectra from low and medium pressure mercury lamps.

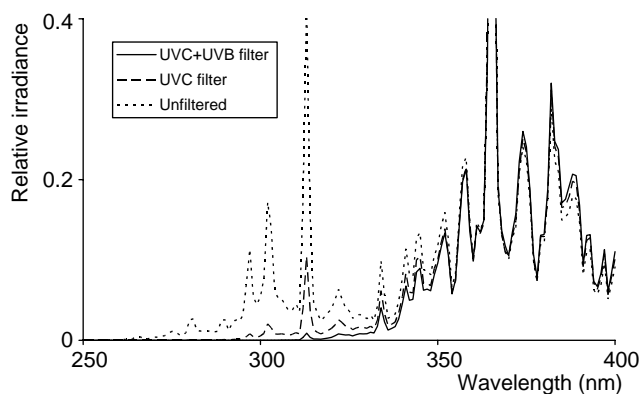


Figure 5. Ultraviolet emission spectra from metal halide lamp with and without filters.

The medium pressure mercury lamp is also the source in the “Wood’s lamp”, although in this case optical filtration is incorporated into the lamp housing to limit the short-wavelength UV and visible light emissions. This results in an approximate monochromatic source of 365 nm. The Wood’s lamp is used in diagnostic fluorescence techniques in dermatology.

Fluorescent Lamps

A fluorescent lamp is a low pressure mercury discharge lamp that has a phosphor coating applied to the inside of the envelope. Fluorescence radiation is produced by the excitation of the phosphor by the 253.7 nm radiation. The spectral power distribution of the fluorescence radiation is a property of the chemical nature of the phosphor material. In addition to a continuum due to the phosphor, the mercury characteristic lines are superimposed. These lines are present in all mercury fluorescent lamps irrespective of the phosphor material. However, it is possible to suppress the emission of certain lines either by using a lamp envelope material that absorbs unwanted short-wavelength radiation or by incorporating a suitable filter in the lamp housing. Because fluorescent lamps are relatively cheap, stable, and efficient and can produce a large high intensity irradiation field, they are now very widely used in phototherapy and cosmetic tanning.

The spectral power distribution of UVA lamps used in cosmetic tanning and PUVA phototherapy are shown in Fig. 6. Broad-band UVB fluorescent lamps (e.g., TL12, Phillips, Eindhoven, the Netherlands) incorporate a phosphor that results in a continuous spectrum from 270 to 380 nm, with a peak at 313 nm. This lamp is used in phototherapy, however, it is being replaced by the narrow-band lamp (Phillips TL01), which has ~80% of its output within 2 nm of the peak at 311 nm (Fig. 7).

The black light fluorescent lamp emits a similar spectral power distribution to the UVA fluorescent lamp in the region 315–400 nm, but with suppression of the mercury lines in the visible spectrum. This results from using a visible-absorbing, UVA transmitting, glass envelope. When switched off, the lamp envelope appears almost black. A purplish light is perceived when the lamp is operating.

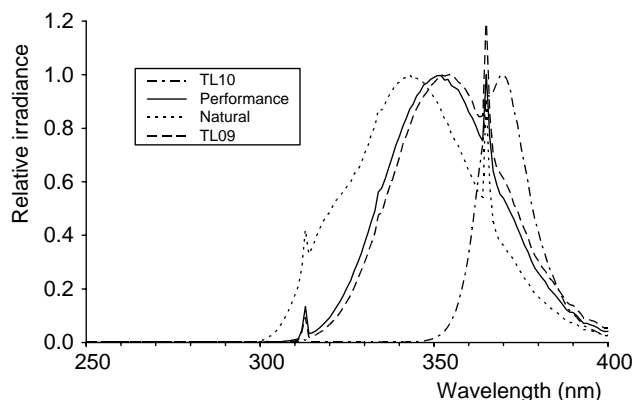


Figure 6. Ultraviolet emission spectra from fluorescent UVA lamps used in PUVA therapy and commercial tanning studios (Cleo Performance and TL09). Also shown are the Phillips TL10 fluorescent lamps emitting longer wave UVA, that were popular for tanning at one time and the Phillips natural fluorescent lamp that have been promoted by the manufacturers as “a more natural way to tan”.

Xenon Arc Lamps

In the xenon arc the radiation is emitted primarily as a continuum, unlike the mercury arc, which essentially emits a line spectrum. The production of the continuum is optimum under conditions of high specific power, high current density, and high internal pressure, leading to compact, bright sources. Because of high operating temperatures the lamp envelope is normally constructed from fused silica. Unlike arc lamps containing mercury that has to vaporize, xenon lamps contain a permanent gas filling and the full radiation output is available immediately after switching on so that no run-up period is necessary.

Solar Simulators

Because of the similarity of the spectrum to that of the solar spectrum, the xenon lamp has been employed as a laboratory source of sunlight, the so-called solar simulator. In order to improve the match to sunlight, a WG320 filter

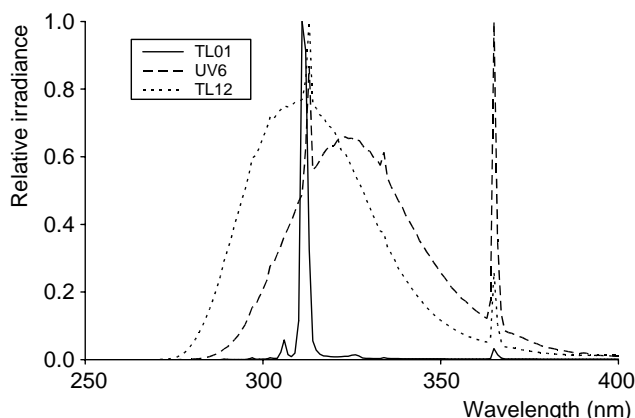


Figure 7. Ultraviolet emission spectra from fluorescent UVB lamps.

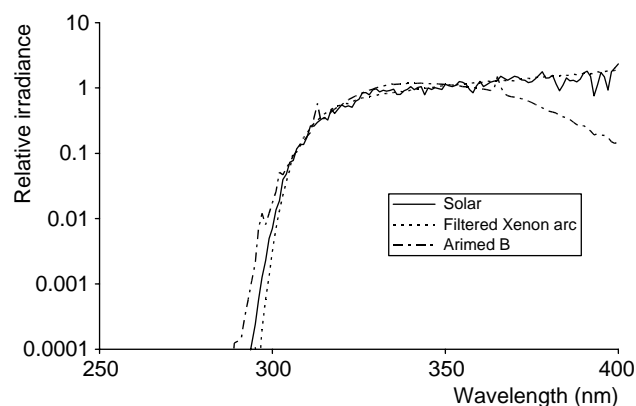


Figure 8. Ultraviolet emission spectra from solar simulator sources together with the solar spectrum obtained at mid-day on a sunny summer's day in Newcastle-upon-Tyne (UK). The relative irradiance has been displayed on a logarithmic scale to allow comparison of small differences in the UVB range.

(Schott AG, Mainz, Germany) should be used and often a visible light absorbing filter (e.g., Schott UG5 or UG11) is also employed. A drawback of the arc lamp solar simulator is that it is difficult to irradiate a large area. Fluorescent lamps can be used for this purpose and the best match to sunlight is probably given by the Arimed B lamp. The suitability of a lamp as a solar simulator can be assessed by comparing the percentage relative cumulative erythral effectiveness (%RCEE) for a number of wavebands and compare this with the %RCEE values for a "standard sun" (17). Various solar simulator sources are shown in Fig. 8.

Monochromatic Radiation

A radiation monochromator consists of a xenon arc lamp together with a double diffraction grating monochromator. By adjusting the angle of the diffraction grating, and by adjusting the input and output slit widths a narrow range of UV wavelengths can be obtained. These instruments are widely used in photobiology studies, both *in vivo* and *in vitro*. They are also routinely used for testing the erythral sensitivity of patients in phototesting clinics. Although the radiation is referred to as monochromatic, in fact bandwidths of between 30 and 5 nm are typically employed. Narrowing the bandwidth <5 nm may lead to unacceptably low irradiance.

Lasers

Lasers produce truly monochromatic radiation. There are a number of lasers that produce UV radiation with emissions possible in the UVA, UVB, and UVC range. Argon Fluoride lasers produce 193-nm radiation (UVC) and are used for corneal refractive surgery (19). Dye lasers have been used to investigate the action spectrum for erythema in human skin (20).

THE MEASUREMENT OF ULTRAVIOLET RADIATION

Techniques for the measurement of UVR may be divided into three classes; biological, chemical, and physical. In

Table 2. Radiometric Terms

Term	Unit
Wavelength	Nanometer, nm
Radiant energy	Joule, J
Radiant flux	Watt, W
Radiant intensity	Watt per steradian, W/sr
Radiance	Watt per square meter per steradian, W/m ² /sr
Irradiance	Watt per square meter, W/m ²
Radiant exposure (often referred to as "dose")	Joule per square meter, J/m ²

general, physical devices measure power, while chemical and biological systems measure energy.

Quantities and Units

In clinical and photobiological UVR dosimetry, it is customary to use the terminology of radiometry rather than that of photometry. Photometry is based on visible light measurements weighted by the human eye's response curve, and therefore not relevant as the eye does not respond to radiation at wavelengths <380 nm. The common radiometric terminology is listed in Table 2. These radiometric quantities can also be expressed in terms of wavelength by adding the prefix "spectral". In clinical photobiology, the derived unit of milliwatt per square centimeter is commonly used and radiant exposure tends to be referred to as dose. Note that dose in this context differs from the term used in radiobiology where dose indicates energy absorbed per unit mass of tissue.

The most frequent radiometric calculation is to determine the time for which a patient who is prescribed a certain dose (in J/cm²) should be exposed when a radiometer indicates the irradiance in mW/cm². The relationship between these quantities is

$$\text{exposure time (min)} = \frac{1000 \times \text{dose (J/cm}^2\text{)}}{60 \times \text{irradiance (mW/cm}^2\text{)}}$$

Weighted Irradiance

It is useful to derive quantities by weighting the spectral irradiance by an appropriate action spectrum. The erythemally effective irradiance E_{eff} is defined as

$$E_{\text{eff}} = \sum s(\lambda)E(\lambda)\Delta\lambda$$

Where $s(\lambda)$ is the CIE erythral action spectrum (8), $E(\lambda)$ is the measured spectral irradiance, and $\Delta\lambda$ is the bandwidth of measurement. The "Global UV index" (21) can be used as a standard way of expressing the erythemally weighted solar irradiance. A measure of the integrated effective irradiance received is given by the Standard Erythral Dose (SED) (9). One SED is equal to an effective dose of 100 J/m². In the past, some have used the MED as a standard measure of erythral dose. However, this is not correct, MED should be reserved to describe an individual's erythral response and not a standard measure of dose received. The number of SED required to cause just

perceptible erythema ranges from ~ 1.5 to 6 for skin types I to IV. It can be demonstrated that erythematous responses to UV sources with different spectra are similar if the doses expressed are SED (7).

Physical UV Detectors

Physical UV detectors utilize either thermal or photon mechanisms (22).

In thermal detectors, the absorption of radiation increases the temperature in the detector element and this rise in temperature is measured by some means. Thermopile UV detectors are the simplest and commonest thermal device used to measure UV irradiance. A multijunction thermopile is formed from a number of thermocouples in series, which generate a voltage that is proportional to incident energy in the form of heat. Thermopiles intended for use with UV radiation are fitted with quartz windows, which transmit well in the UV range. The advantage of thermal detectors is that they have a relatively flat spectral response over a wide wavelength range.

Photon detectors operate by absorbing discrete quanta of photon energy, and therefore have a threshold wavelength above which no radiation is detected. The lower wavelength limit is related to optical properties of the detector or associated filters. The response of photon devices are therefore inherently wavelength dependent and thus have to be calibrated for each source of interest.

Photoemissive detectors have a photocathode from which electrons are ejected when photons are absorbed. These electrons are then collected by an anode and a current is produced. The simplest of this type of detector is the vacuum phototube consisting of an evacuated tube with a potential difference applied between the cathode and the anode. This device has a gain of unity and a low responsivity (~ 0.05 A/W). A gas-filled phototube has a gain of ~ 10 due to secondary ionization of the gas in the tube that has the effect of producing a greater anode current. Photomultiplier tubes have a series of electrodes (called dynodes) having successively greater potential differences applied between them. As electrons hit each dynode they release further electrons that in turn release more electrons at the next dynode, and so on, leading to a high overall gain (typically 10^6). The responsivity of photomultiplier detectors tends to be high ($\sim 5 \times 10^4$ A/W).

Photodiodes, or junction photodetectors, have a depletion region formed by the junction of n and p doped semiconductor material. On absorption of a photon in this region, electron hole pairs are formed that are then swept out of the region and cause a current to flow in an external circuit. These devices can either be operated in a zero bias or reverse bias mode. For UV detection, GaAsP, GaP, or Si photodiodes are used. These photodiodes are small, cheap, and rugged with good responsivity (~ 0.1 A/W), and are therefore ideal UV detectors.

UV Radiometers

A radiometer is a complete UV measurement device, consisting of a detector and a meter to amplify and display the detector output. Narrow-band UV radiometers are used to measure the irradiance of a source in the different UV

bands. The detector consists of a diffuser to collect the UVR, a filter, and the sensor (e.g., a photodiode) itself. Lambert's law states that the irradiance falling on a surface varies with the cosine of the incident angle. A good diffuser should possess an angular response close to the ideal cosine response—of particular importance in phototherapy, where arrays of tubes are used leading to a large source area irradiation.

A good example of a hand-held radiometer for phototherapy is the IL1400A UV radiometer system (International Light, Newburyport, MA) used with the following probes (see Fig. 9):

- SEL 240 solar blind vacuum phototube, fitted with a SCS280 filter.
- SEL 033 Silicon photodiode detector, fitted with a UVA filter.

Both of these detectors are fitted with domed Teflon diffusers that have a good angular response. The SEL 240 responds over the 185–320 nm range, and the SCS280 filter is a sharp-cut filter, removing all wavelengths < 280 nm. The combination of the two gives a system that responds to wavelengths in the 265–332 nm range, which is well matched to the UVB band. The SEL 033 detector responds

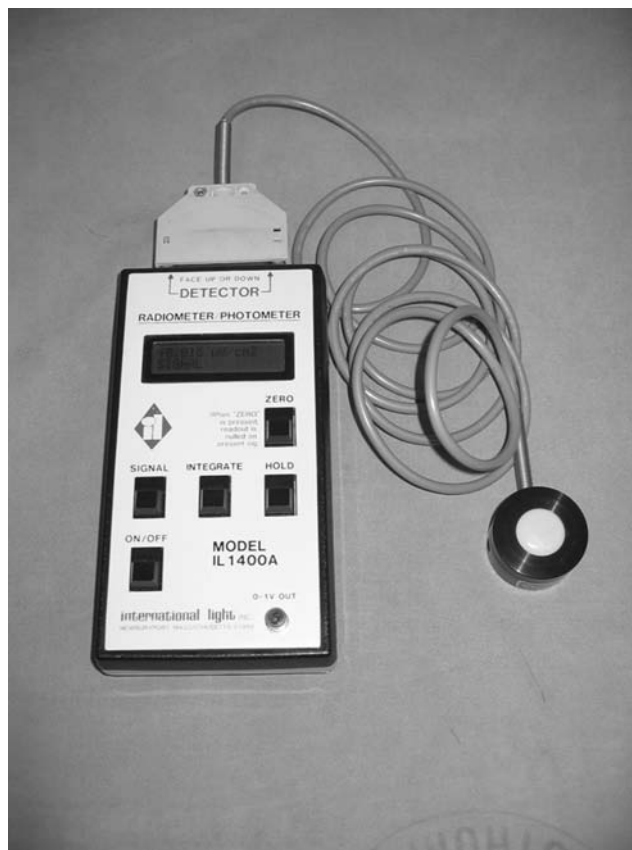


Figure 9. A hand-held UV radiometer suitable for routine monitoring of phototherapy equipment. Model shown in an International Light IL1400A radiometer with UVA detector (see text for further details).

over the 200–1100 nm range, therefore the UVA filter used is a wide-band filter that restricts this response to the UVA range (325–388 nm). It is important to note that narrow-band radiometers have a response that varies with wavelength, and therefore they must be calibrated for each type of source they will be used to measure.

A broad-band radiometer possesses no filter, and therefore measurement of the total irradiance from a source is obtained. Typically these incorporate a thermopile detector connected to an amplifier and display electronics.

Spectroradiometry

Spectroradiometry is concerned with the measurement of the spectrum of a source of optical radiation. In many cases, spectral measurements are not required as ends in themselves, but for application to the calculation of biologically weighted radiometric quantities.

The three basic requirements of a spectroradiometer system are (1) the input optics, designed to conduct the radiation from the source into (2), the monochromator, which usually incorporates one or two diffraction gratings as the wavelength dispersion elements, and (3) an optical radiation detector, either a photomultiplier tube or a solid-state photodiode. The UVR is collected by a diffuser, which should have good angular response, and transmitted to the monochromator by a light guide. A quartz optical fiber is required as quartz has a very high UV transmission, unlike ordinary glass. The monochromator allows the separation of polychromatic radiation into very narrow bandwidths (typically 1 nm) thereby allowing the spectral irradiance of a source to be measured. This is achieved by using a diffraction grating, which disperses the incident radiation. Scanning spectroradiometer systems operate by measuring at a given wavelength, then changing the angle of the diffraction grating, making another measurement, and so on. For photobiology applications it is recommended that a double diffraction grating system is used. The second diffraction grating achieves a reduction in the radiation transmitted outside the waveband of measurement (so-called "stray radiation"). Very small amounts of stray radiation may have large implications in the measurement of erythemally weighted irradiance if it occurs at biologically highly effective wavelengths. The spectroradiometer system requires a high responsivity detector as narrow bandwidths are selected. All of the spectral examples in this article were obtained with this type of spectroradiometer system (the majority using a model DMc150 monochromator from Bentham Instruments, Reading, U.K.). Spectroradiometers need to be calibrated with reference to a standard lamp, which in turn has a output calibrated by a national standards laboratory. The disadvantage of double-diffraction grating scanning spectroradiometers systems are the high cost, large size, and relatively long time required to obtain a complete spectrum.

Recently small, rugged, and relatively cheap spectroradiometers systems have become available (e.g., Sola-Scope 2000, 4D controls Ltd, Redruth, U.K.). These devices consist of an input optic, single diffraction grating, and solid-state multidetector array. In this device, there are no moving parts and a complete spectrum is rapidly obtained.

A major disadvantage for photobiology application is that the stray light level is high (>10%), although this can be subtracted from the measured spectrum to give a reasonable level of compensation (23).

Calibration of Meters for Phototherapy. It is recommended that radiometers are routinely used in phototherapy practice and regularly calibrated in a manner traceable to a national standards laboratory (24). This can be done in one of two ways: either by reference to a calibrated spectroradiometer (25) or by reference to a calibrated meter (26). Surveys of radiometer accuracy have shown very disappointing results, particularly for the measurement of narrow-band UVB radiation (26,27). This is of particular concern since narrow-band UVB phototherapy is becoming increasingly popular, surpassing traditional PUVA therapy. Also, the increase in erythema with dose is greater with narrow-band UVB than PUVA, and therefore the need for accurate dosimetry is greater.

Chemical and Biological Methods

Detailed description of chemical and biological methods is beyond the scope of this article, but they are mentioned briefly as they may be used in a few medical applications. Polysulphone film (28) changes its optical properties with the absorption of UVR. It is possible to measure this change and relate it to the exposure dose received by the film. These devices are useful for measuring human UV exposure as they are unobtrusive for individuals to wear. Biological dosimeters (29) make use of the inactivation of bacteria or viruses as a function of UVR dose. There are certain applications, such as measuring UV doses in water flowing in a disinfection plant, where biological dosimeters are the only reliable way to measure UV dose.

DIAGNOSTIC USES OF ULTRAVIOLET RADIATION

Fluorescence Techniques in Diagnosis

The principle of fluorescence diagnosis is based on the absorption of radiation by a fluorophore in tissue and the subsequent emission of photons of light at a longer wavelength. Typically, the exciting radiation is short wavelength visible light or UVR and the fluorescence is longer wavelength visible light. Certain tissue types may be distinguished (e.g., neoplastic tissue) from surrounding tissue due to differences in endogenous fluorophore. Other techniques use exogenous photosensitisers that may be selectively accumulated in neoplastic tissues. The most common of these photosensitisers is 5-aminolevulinic acid (5-ALA). Originally, applications were limited to visual observation of fluorescence at easily accessible sites, such as the skin and mouth. However, technological advances in fiber optics, light sources, digital video cameras, and computer enhanced imaging has led to considerable interest in endoscopic fluorescence imaging (30). There is a wide range of excitation and emission wavelengths that can be used in these applications, often these are both in the visible range, but techniques using a Wood's lamp have been used for many years by dermatologists.

Tinea capitis (ringworm) is a fungal infection of the scalp. Hairs that are infected with the fungi *Microsporum audouini* or *Microsporum canis* will fluoresce with a bright blue-green color under Wood's light. The diagnosis can be confirmed by removing the fluorescent hairs for direct microscopic examination and culture. Erythrasma is a superficial bacterial infection of the skin, usually between two surfaces of skin that rub together, such as the groin or toe webs. Erythrasma produces scaling and cracking and may be accompanied by itching. The responsible organism (*Corynebacterium minutissimum*) produces a porphyrin that fluoresces a bright coral-red color on irradiation with a Wood's lamp. Irradiation of the oral cavity with a Wood's lamp will produce fluorescence that may prove useful in the diagnosis of various dental disorders such as early dental caries (tooth decay). Normal teeth fluoresce with a light blue color. The presence of calculus on the teeth will result in a yellow-orange fluorescence under UVA illumination.

Photosensitivity Investigations

Sunlight is capable of inducing or aggravating skin diseases. This group of diseases is collectively known as the photodermatoses (31,32). The classical appearance of a photodermatosis is a rash confined to regions exposed to sunlight; the face, neck, arms, and hands. The appearance may be an exaggerated sunburn, or there may be eczematous changes with polymorphic lesions that can include papules, vesicles, and bullous eruptions. The principal photodermatoses are shown in Table 3.

The accurate diagnosis of a suspected photosensitive patient demands, above all, a clear and detailed history. For example, the season of the year in which the symptoms occur may give some guide to the wavelengths in the sun's spectrum that are responsible. The time between exposure and the appearance of the lesion may also be a helpful guide. However, in equivocal cases, phototesting investigations are desirable (33). The object is twofold: to reproduce the disease so as to confirm the diagnosis, and to ascertain those wavelengths of sunlight that are responsible for the

photosensitivity, so that suitable preventative measures can be taken.

Provocation Testing. The aim of provocation testing is to expose a relatively large area of skin (say 10–20 cm²) to UVR and attempt to reproduce the rash of which the patient complains. The methodology used and the success of the technique is variable. The provocation source used in the past has been a medium pressure mercury arc lamp in combination with optical filters to isolate broad spectral regions (e.g., UVB or UVA). Recently, the use of a filtered xenon arc solar simulator source has been recommended as optimal (34). However, others have had good success with narrow-band UVB and UVA fluorescent lamps (35).

Minimal Erythmal Dose Determination. The MED is usually determined using a radiation monochromator. This enables the patient to be irradiated in narrow wavelength intervals. The patient's back is normally chosen for irradiation since it has a large surface area with reasonably uniform sensitivity to UV radiation. A geometric series of doses at each of several different wavelengths is used (33,36) and the results of irradiation are observed 24 h later. In most centers visual inspection is used, although for quantitative research studies an erythema meter may be used (11). Comparing the response in the UVA and UVB range may be helpful in diagnosis (37). For example, in drug induced photosensitivity the UVA MED may be low, and the UVB MED may be normal, whereas in chronic actinic dermatitis, both UVA and UVB MEDs are low (see Fig. 10).

Photopatch Testing. Certain chemicals when activated by UVR in contact with the skin may cause a photocontact allergic reaction in sensitive individuals. This is not to be confused with a phototoxic reaction that occurs in all subjects when exposed to a photosensitizer such as psoralen. The most common photocontact allergens are sunscreens. When UVR is absorbed by a chemical sunscreen it becomes altered and the patient may have an allergic

Table 3. The Principal Photodermatoses, Main Clinical Features, and Typical Phototest Responses

Condition	Typical Clinical Features	Typical Erythmal Response	Provocation Test Response	Photopatch Test Response
Polymorphic light eruption	Itchy, papular rash occurring after a few days sun exposure. Persists a few weeks	Usually normal MEDs, occasionally low MEDs	Positive response (50–80% sensitivity) at suberythmal doses	n/a
Chronic actinic dermatitis	Eczematous rash on sun exposed site in summer months	Low or very low MEDs within UVA and UVB waveband, sometimes sensitivity to blue light also	n/a	n/a
Solar urticaria	Urticarial reaction a few minutes after sun exposure. Fades rapidly	Rapid reaction (within minutes) that may include UVB, UVA and blue/green light sensitivity	n/a	n/a
Drug induced photosensitivity	Rash on sun exposed sites: features vary with responsible drug	Low MEDs in UVA, normal MEDs in UVB range	n/a	n/a
Photocontact allergy	Confluent rash on exposed sites: may be severe	Normal	n/a	Positive when chemical is exposed to UV



Figure 10. Abnormal erythematous sensitivity test obtained at 24 h after exposure in a patient with chronic actinic dermatitis. Dose increments of 40% were used at the 300, 320, and 350 nm. The \gg marks indicate the direction of decreasing dose. In a normal subject, only a few exposed areas would become red.

reaction to this altered chemical. Photopatch testing (33,38) consists of placing several small patches of chemicals on the back in two duplicate sets. After 24 h, one set is exposed to UVR (usually 5 J/cm^2 UVA from a fluorescent lamp or radiation monochromator). Reactions are examined after a further 48 h. If a chemical causes a reaction at exposed and unexposed sites this is a contact allergic reaction. If the reaction occurs at the UVR exposed site, only then this is a photocontact reaction.

THERAPY OF SKIN DISEASE WITH ULTRAVIOLET RADIATION

The principal application of UVR for therapy is to treat a variety of skin diseases. The UVR may be administered on its own (phototherapy) or in conjunction with adjunctive agents applied topically or photoactive drugs taken systemically (photochemotherapy).

Ultraviolet Phototherapy

In 1994, a survey of UVB phototherapy in the United Kingdom (39) found that an appreciable number of centers were using equipment that was old and was suboptimal in producing a therapeutic effect. A similar survey in 2002 (24) showed that improvements in equipment were being made. Mercury discharge lamps had mostly been replaced by fluorescent tube irradiators and broad-band UVB lamps

had mostly been replaced by narrow-band UVB lamps. The use of narrow-band UVB therapy has increased recently and now surpasses the use of psoralen plus UVA photochemotherapy (PUVA) (40). A recent review and guidance on the use of narrow-band UVB therapy by the British Photodermatology Group (40) concluded that there was good evidence to support its use in the treatment of psoriasis, chronic atopic eczema, vitiligo, and polymorphic light eruption. A range of other diseases have been treated with varying success.

Action Spectrum for Clearance of Psoriasis Using UVR Alone. Psoriasis is a common genetically determined skin disease with a 2% incidence worldwide (41). It is characterized by disfiguring, often distressing, red scaly plaques that may become confluent. The severity varies in any individual at any given time, but it often persists indefinitely. Parrish and Jaenicke (42) studied psoriasis response at different UV wavelengths to derive an action spectrum for psoriasis clearance. Figure 11 shows the clearance action spectrum together with the action spectrum for erythema for one subject in that study; similar results were found for other subjects. It can be seen that radiation with wavelengths $<290 \text{ nm}$ is ineffective at clearing psoriasis, but effective at producing erythema. Broad-band UVB sources (such as the Phillips TL12 lamp) that have traditionally been used for phototherapy tend to have relatively high emissions at wavelengths $<290 \text{ nm}$. The narrow-band UVB lamp (Phillips TL01) has only 2.3% erythemally effective radiation at wavelength $<290 \text{ nm}$. It is therefore not surprising that clinical studies comparing traditional broad- and narrow-band (TL01) UVB lamps for psoriasis treatment (40) have usually shown that narrow band is more effective. There are fluorescent broad-band UVB lamps available (e.g., UV6; Sylvania, Brussels, Belgium) with little or no emission $<290 \text{ nm}$ and the efficacy of these for psoriasis clearance has not yet been established. Spectra of various UVB fluorescent sources are shown in Fig. 7.

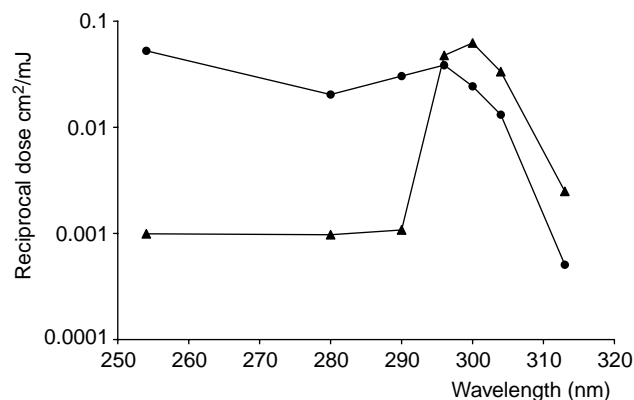


Figure 11. Psoriasis clearance action spectrum and erythema action spectrum in an individual subject. (From Ref. 42). The triangles indicate the reciprocal of the lowest daily dose to clear psoriasis and the circles are the reciprocal of the minimal erythemal dose. Similar responses were seen for the other three subjects in this study.

Psoralen Photochemotherapy

This form of treatment, known as PUVA, involves the combination of the photoactive drug psoralen and UVA irradiation to produce a beneficial effect. Psoralen photochemotherapy has been used to treat many skin diseases (43), although its principal success has been in the management of psoriasis. The psoralen may be applied to the skin either topically (44) or systemically (45); the latter route is generally preferred. The psoralen is usually administered as 8-methoxypsoralen (8-MOP). The patient ingests the 8-MOP tablets and is then exposed to UVA radiation 1 or 2 h later when the photosensitivity of the skin is at a maximum. The mechanism of the treatment is thought to be that psoralen binds to DNA in the presence of UVA, resulting in a subsequent transient inhibition of DNA synthesis and cell division. The UVA lamps used for PUVA therapy (such as the Phillips TL09 or Phillips Cleo Performance lamps) (Fig. 6), typically have a broad spectrum from ~315–400 nm peaking at ~352 nm. Within the UVA range, the action spectrum for psoriasis clearance is thought to be similar to the erythema action spectrum in psoralen sensitised skin and these peak near 320 nm (46). A study using narrow-band UVB and psoralen found no significant difference in response compared to conventional PUVA treatment (47). It is possible that UVA lamps conventionally used for PUVA treatment do not have the optimum spectral characteristics. However, narrow-band UVB phototherapy is now more widely used than PUVA and there is less interest in pursuing optimisation of PUVA treatment.

Treatment Regimes for UV Therapy

It is the custom in phototherapy and photochemotherapy to use the patient as their own biological monitor. The exposure of UVR at a given wavelength required to produce a given degree of erythema in normal Caucasian skin can vary by a factor of 5 or more (48), depending on an individual's susceptibility to sunburn. For photochemotherapy the degree of variability in erythema in sensitized skin may be greater due to variations in psoralen metabolism. Therefore, before embarking upon a course of irradiation, the minimum dose to cause just perceptible erythema should be determined. For phototherapy, this is termed the MED, whereas for photochemotherapy it is termed the minimal phototoxic dose (MPD). Either MED or MPD measurement is made by exposing the skin to increasing doses of UV using a lamp with the same spectral output as the treatment lamp. Devices using a series of filters have been designed to facilitate this (49) (see Figs. 12 and 13). For phototherapy, the MED is usually determined at 24 h. However, erythema in psoralen sensitized skin does not reach a maximum until ~72–96 h, so the MPD is determined at this time. Starting doses for treatment are usually set at ~50–70% of the MED or MPD. Treatment is typically given two or three times weekly until the psoriasis clears. At each treatment, the dose is increased to allow for acclimatisation of the skin. If erythema occurs then the dose may be reduced. The total time until clearance varies considerably from one patient to another, and in some cases complete clearing of the lesions is never

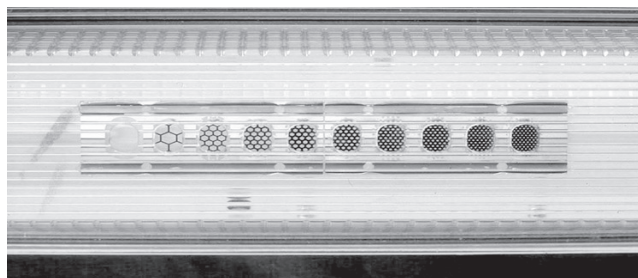


Figure 12. An erythematous skin testing device for use prior to phototherapy, consisting of a TL01 fluorescent lamp and a number of apertures covered by metal grid filters [based on design by Diffey et al. (49)].

achieved. However, typically ~25 treatments are required for clearing of the psoriatic lesion in many patients over a period of ~10 weeks. Various different treatment regimes are used with varying starting doses and number of treatments per week. Examples are given in British Photodermatology Group guidelines (40,44,45) and by Diffey and Hart (50).

UVA1 Therapy

There have been several studies into the use on UVA1 (340–400 nm) therapy (51). As yet this treatment is only available in a few centers and insufficient work has been done to reach a firm conclusion on its efficacy.

Risks of Phototherapy

The major acute risk in phototherapy is that of erythema. In phototherapy, it is the erythematous response in the unaffected skin that limits the treatment dose and frequency. Therefore, it is important to have an understanding of how erythema increases over time and with increasing dose for different therapy sources. Erythema peaks at ~8–12 h after exposure for solar simulated radiation, UVC and UVB (7,52). The degree of erythema varies rapidly with the dose of UVR. When the log of dose is plotted against erythema index, a sigmoid shaped curve is seen with a relatively linear portion above the MED. The steepness of this curve will show variation from individual to individual.

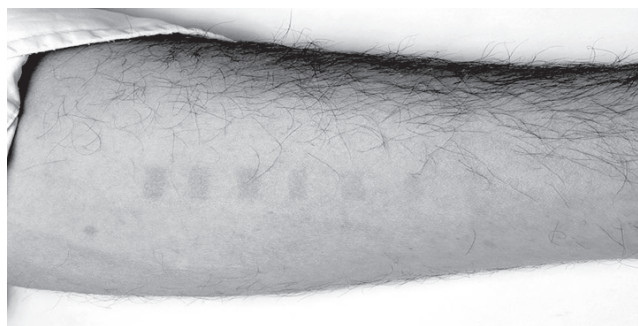


Figure 13. Normal erythematous response at 24 h obtained using the device shown in Fig. 12 in a subject with skin type 2. The MED is seen at site 6 or possibly 7.



Figure 14. A common type of whole body phototherapy therapy supplied by Waldmann (Herbert Waldmann GmbH & Co., Villingen-Schwenningen, Germany.) This illustrates the direct method of measuring UV irradiance; after the door is closed lamps are lit and measurements are made in a number of positions.

However, on average, for UVB sources doubling the dose will cause a change from mild to moderate erythema (53), whereas for psoralen sensitized skin a quadrupling of the dose is required to cause a similar change (54). It is now firmly established that long-term PUVA treatment leads to a rise in SCC incidence (50). A cumulative dose of $<500 \text{ J/cm}^2$ is unlikely to result in significant risk, whereas for a cumulative dose of 2000 J/cm^2 the risk of SCC is $\sim 50\%$. Phototherapy with UVB alone is thought to carry a lower risk of skin cancer than PUVA (40).

Equipment for Ultraviolet Therapy of Skin Diseases

Ultraviolet therapy is usually delivered using large cabins housing $\sim 24\text{--}48$, 6 ft vertical fluorescent lamps Fig. 14. Compact units for treating just the hands and feet or smaller areas are also available (Fig. 15). All treatment devices usually incorporate timers to terminate treatments and most now have in-built UV meters. This enables treatment to be entered in dose units; the machine then terminates the treatment when the dose reaches the preset level. The benefits of using in-built meters are that changes in lamp output over time are compensated for and there is less opportunity for human error in calculating the

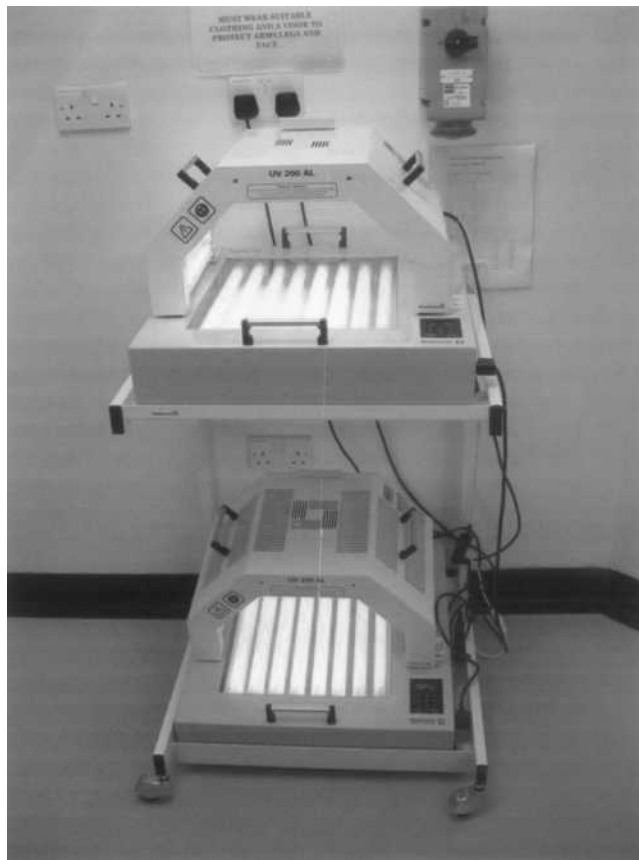


Figure 15. A small PUVA therapy unit suitable for treatment of the hands and feet.

required treatment time for a given dose. The disadvantage is that if the meter sensitivity changes (maybe due to dirt covering the detector) or the sensor becomes covered or shadowed by a large patient, then the patient will receive an overexposure. This problem has been partly overcome in some cabin designs that have 2 m and treatment is terminated if the two readings differ by a set amount. Most systems have safety features to limit the maximum time or dose that can be delivered. An external backup alarm (55) can also help to prevent overexposure.

Equipment Quality Control in Phototherapy

It is recommended that a quality control regime is implemented for phototherapy equipment (24). The approach used is different for cabins with in-built meters to those without. If in-built meters are not used then measurement of irradiance is required for calculation of exposure time for a given prescribed dose. The frequency of measurement required will depend on usage, but typically monthly readings are desirable due to lamp ageing. Daily checks are required to identify any tube failures. If in-built meters are used, compensation is automatically made for changes in cabin irradiance. However, the accuracy and reproducibility of the meter needs to be assessed. It is important to note that the irradiance at the patient surface during treatment is less than that

measured in the same position with the cabin unoccupied, because some of the energy received at the patient surface has been reflected from other parts of the cabin by mirrors behind the fluorescent tubes (56). Without the patient present this reflected fraction is greater. The ratio between “direct” (i.e., with patient in cabin) and “indirect” measurements ranges from ~ 0.65 to 0.90 (57,58), varying with cabin geometry, lamp spectrum, and patient size. The irradiance within a cabin will vary with position, so it is usual to determine the mean irradiance over several positions for an average sized patient; this is termed the Designated Patient Irradiance (DPI). In order to avoid undue staff hazard, the DPI may be measured using the direct method initially (Fig. 14) and subsequently using the indirect method with a correction for the direct–indirect ratio. Automated systems may be employed to move the detector to a number of predetermined positions within the cabin (59,60). The in-built meter reading should agree with the DPI measured using an independent meter to within 20%. Measurement of irradiance from all equipment should be regularly recorded in a standard manner so that trends over time can be assessed. It is important that there is consistency between calibration of equipment used for MED or MPD testing and that used for treatment. The meter used should be correctly calibrated for the lamp being measured and be recalibrated with reference to national standards annually.

SUNBEDS AND TANNING

One could argue that a sunbed is not a medical device, since its primary purpose is cosmetic tanning. However, information on sunbeds is included here for a number of reasons:

- Certain sunbeds could be classed as medical devices according to British Standards.
- There are concerns and considerable media interest regarding the safety of long-term sunbed use that may impact on health services.
- Many people with skin diseases try using sun beds to treat themselves.

Sunbed Lamps

Prior to the mid-1970s, the only artificial way to achieve a tan was using a “sunlamp” at home. These lamps were unfiltered medium or high pressure mercury arc lamps that emitted a wide spectrum including significant UVC and UVB components. Because of the significant short wavelength contribution, exposure times were short (a few minutes) and the incidence of burning was high. In the early 1980s tanning using UVA fluorescent lamps became popular, largely because the tanning industry promoted the idea that UVA tanning was safe (61). Typical UVA lamps used were similar to those used for PUVA therapy (e.g., Philips TL09 lamp) with $\sim 0.7\%$ of the total UVR in the UVB range. In the mid-1980s, a new lamp was introduced for tanning (Philips TL10) that had a spectrum

extending from 340 to 400 nm with a UVB component of only 0.05%, but these lamps are not often used today (62). It has been proposed that the “least bad” way to acquire an artificial tan may be to use a lamp that emits UVB in a similar proportion to natural sunlight (63). Philips have recently introduced the Cleo Natural lamp with this in mind (64). A survey of commercial tanning establishments in Scotland (62) found 13 different models from 7 manufacturers. However, the lamps fell into three categories; 81% were type 1 UVA lamps (e.g., Philips Cleo Performance, emitting $\sim 0.7\%$ UVB); 8% were type 2 UVA lamps (e.g., Philips Cleo Professional, emitting $\sim 1.4\%$ UVB); and 11% were filtered high pressure metal halide lamps. Other new higher UVB emitting lamps are now available such as the Philips Cleo Advantage and Swift (64), but surveys to assess the popularity of these lamps have not been reported. Spectra of UVA lamps used for tanning are shown in Fig. 6.

McGinley et al. (62) found that the UVB irradiance from an average sunbed was similar to that of Glasgow, UK, on a sunny summer day, whereas mean UVA irradiances were three or four times higher. They estimated that the total UVR received during 20 sunbed sessions was similar to that received during a week sunbathing in the Mediterranean.

Tanning

In the tanning process, melanocytes are stimulated to produce melanin that is transferred to keratinocytes. It is widely believed that obtaining a tan protects against sunburn. However, the degree of protection may only be equivalent to a sunscreen with a sun protection factor (SPF) of 3 (65) and there is evidence to suggest that a UVA tan provides even less protection than that due to natural sunlight. The action spectra for tanning and erythema are very similar for people who do not tan easily (42). For those who tan easily it is possible to obtain a UVA tan without burning.

Risks of Sunbed Use

Short-term risks of sunbeds include sunburn, itching, skin rashes, nausea conjunctivitis, and photokeratitis (66). Sunbeds can also provoke the common photodermatoses, polymorphic light eruption. A serious concern over long term sunbed use is that of developing skin cancer (65). The erythema, tanning and squamous cell carcinoma action spectra (see Fig. 3) are all similar, and therefore tanning by any part of the UV spectrum carries an approximately equal risk for SCC. There is also now increasing evidence that sunbed use is associated with increased melanoma risk (65).

Legislation

Sunbeds are used by a significant minority of the population and recently they have become popular with young people. British and European standards (67) classify sunbeds according to the erythemally weighted irradiance in the UVA and UVB range. According to this classification, the typical UVA sunbed is a type 3 device and a sunbed

with Philips Natural lamps is a type 4 device (68). Type 4 devices should be marked with the following warning: "to be used following medical advice only". There is, however, no statutory regulation of sunbed use in the United Kingdom. Some countries or U.S.A. states have regulated access to sunbeds to a certain extent (65). Autier (61) proposes strict controls to minimize and regulate sunbed use. However, Diffey (69) believes that outright prohibition is unnecessary since the excess deaths from sunbed use in the United Kingdom is relatively small compared to that from natural UV exposure and other voluntary risks (e.g., smoking and alcohol).

Sunbeds and Home Phototherapy

A survey has shown that many patients use sunbeds at home to treat psoriasis and believe it to be effective (70). However, a randomized placebo controlled trial of UVA sunbed therapy for psoriasis (71) found only a small degree of improvement. A consensus report of the British Photodermatology Group (72) concluded that home phototherapy represented suboptimal treatment with greater risks compared to hospital based treatment. Das et al. (68) carried out a randomized comparison between conventional UVA sunbed lamps and lamps with higher UVB component (Philips Natural). They found no difference in response for equal erythemal doses. However, when equal exposure times were used (similar to that used in a conventional tanning session) significant improvement in psoriasis were seen. The authors conclude that this home sunbed treatment is likely to be less effective compared to conventional hospital treatment, but may be appropriate for some patients who cannot access hospital therapy. Several groups have investigated installing conventional phototherapy systems in patient's homes. A recent project in Scotland (73) installed narrow-band UVB in patients home and provided them with treatment schedules and training. They concluded that the treatment was effective, safe and cost effective. Overall, home phototherapy probably is worth considering for certain patients and centers with a very rural catchment area.

OTHER USES OF UV IN MEDICINE

Laser Eye Correction

Energy from an ArF1 (argon fluoride) excimer laser at 193 nm is very well absorbed by proteins within the cornea. Molecular bonds are broken as the photon energy is greater than the bond strength within these proteins. The process of tissue removal is not like that of a scalpel, or other lasers. Rather, a broad beam is used that removes tissue by a process called "ablative photodecomposition". Since the laser energy is absorbed very near the surface, little tissue damage beyond this region occurs. The degree of tissue removal is computer controlled so that the cornea is reshaped to give the appropriate degree of refractive correction for the patient. Several different effective techniques are used (74,75), Photorefractive Keratectomy (PRK) uses the laser alone to reshape the cornea, whereas Laser-Assisted *In Situ* Keratomileusis (LASIK) is a combined

surgical and laser technique. During LASIK surgery, a thin flap on the cornea is cut and folded back and the laser is then used to remove a precise amount of corneal tissue. The LASIK method is currently the dominant procedure and has the advantage of maintaining the central corneal epithelium (76). Current evidence on LASIK suggests that it is efficacious in selected patients with mild or moderate myopia. Evidence is weaker for its efficacy in patients with severe myopia or hyperopia, and there are concerns about the safety of its use in the long term (77). The U.K. National Institute of Clinical Excellence is due to issue guidance on the use of PRK during 2005.

Photopheresis

Photopheresis (also known as extra-corporal photochemotherapy or ECP) was first described by Edelson et al. in 1987 (78) and its use has recently been reviewed by McKenna et al. (79) The technique involves the following process: venous blood is removed from the patient, the white cells are separated and mixed with 8-methoxypsoralen, these cells are then irradiated using UVA lamps and then returned to the patient. A treatment session takes ~4 h and this is repeated every few weeks for several months. Currently, the technique is expensive and only used on a relatively small number of patients in a limited number of centers. It has been investigated in a wide range of diseases with varying degrees of success. The most promising results have been in the treatment of cutaneous T-cell lymphoma and graft versus host disease. There is some debate about the mechanism of action; it is not likely to be a direct effect on circulating lymphocytes since only a few percent are treated during each session. It is thought that it may work by stimulating the patient's immune response to the malignant lymphocytes.

Blue Light Phototherapy

Blue light phototherapy (50) is used in the treatment of neonatal jaundice. The condition is caused by the inability to clear bilirubin. Bilirubin is formed as the body scavenges iron from heme molecules released during the destruction of red blood cells. Irradiation of bilirubin by blue light causes its conversion to forms that are less toxic and more easily removed from the body. As the liver matures it becomes more able to excrete bilirubin so that phototherapy is only required for a short time. There is uncertainty about the optimum spectral output for lamps used in phototherapy, although it is thought that the most effective wavelength is ~450 nm. Either fluorescent lamps or metal halide lamps are used for blue light phototherapy. The hazards of blue light are almost entirely to the retina. Note, however, that many blue light phototherapy lamps emit radiation in the UV range. Normally, this is removed using an acrylic sheet. However, it has been shown that lamps may emit ~0.2 mW/cm² UVR with the shield in place and ~10 times this with the shield absent.

Disinfection and Hygiene

Ultraviolet radiation of wounds and ulcers has been used in the past (80). The success of the treatment relies on the

bactericidal properties of UVR and for this reason it is important that lamps that emit UVC are used. Advances in antibiotic therapy meant that the technique became less common. There is, however, renewed interest in this treatment for methicillin resistant *Staphylococcus aureus* (MRSA) infected wounds (81,82).

UVC irradiation is also used to kill bacteria in the air in operating theaters and there has recently been increased interest in this technique as a method of dealing with deliberate terrorist release of pathogens in buildings (83).

Many flying insects are attracted for UVA radiation (~350 nm). Certain insect killing devices attract insects with UVA, and then electrocute them as they fly toward the source. This method of dealing with flying insects is the technique of choice in many hospital kitchens (84).

Stimulation of Vitamin D Production

Production of vitamin D is the only definite beneficial effect of human UVR exposure. Low level UVR has been used to stimulate vitamin D production in residents of an old people’s home (85).

HAZARD ASSESSMENT AND PROTECTION

Exposure Limits and Hazard Assessment of UV Sources

The International Commission on Non-ionizing Radiation Protection has published guidelines on MPE limits for UV sources. The latest revision of the guidelines (6) contains updated information on hazards, but does not change the value of the limits published in 1989 (86). For most individuals, if exposure is below the limits, then they will not experience acute effects and the risks from long-term effects will be at an acceptable level. The limits are based on an envelope action spectra for all acute and chronic deleterious effects on the eye and skin. Certain individuals with disease or sensitized by chemicals or drugs may experience effects at exposure levels below the limits. To be below the MPE there are two criteria that must be satisfied:

1. Total effective spectrally weighted UV radiation (180–400-nm range) incident on the unprotected skin or eye should not exceed 30 J/m² over an 8 h period
2. Total unweighted UV radiant exposure in the spectral region 315–400 nm on the unprotected eye should not exceed 10⁴ J/m² over an 8 h period

The effective spectrally weighted exposure is calculated using the following expression:

$$E_{\text{eff}} = \sum E_{\lambda} \cdot S_{\lambda} \cdot \Delta_{\lambda}$$

Where E_{λ} is the measured irradiance of the source (W/m²), S_{λ} is the relative spectral effectiveness (unitless), and Δ_{λ} is the wavelength measurement interval. The weighting factors used form a “hazard action spectrum” (6), similar to that used when calculating erythemal effective irradiance. Figure 3 shows a comparison of the two spectra.

The hazard weighted effective irradiance can be measured in two ways: A spectroradiometer can be used to measure the absolute spectral irradiance of the source in question, and the relevant weighting factors can be applied accordingly. A less accurate (but simpler) method is to use a radiometer fitted with an appropriately weighted filter. However, attempts to develop filters that have a good match to the hazard action spectrum have not been particularly successful. Measurements of various UV sources in the author’s laboratory with a spectroradiometer and narrow-band meter with hazard weighted filter (International Light, IL1730A UV Actinic radiometer, fitted with an ACT270 filter) are shown in Table 4. While the broad-band meter provides a reasonable approximation for some sources, there are very large errors for others due to the poor matching of the filter characteristics and the hazard action spectrum. It is important to consider both the weighted exposure to the skin and eye and the separate limit for UVA exposure to the eye. As can be seen from the examples in Table 4, the former limit usually dominates, but for the source with high UVA output it is the latter that is more restrictive.

Table 4. UV Hazards Measured in the Author’s Laboratory (all at 10 cm)^a

UV Source	Weighted Irradiance, W/m ² Measured with Spectroradiometer	Weighted Irradiance, W/m ² Measured with Radiometer	UVA Unweighted Irradiance, W/m ² Measured with Spectroradiometer	Time min to Exceed 30 J/m ² Weighted Irradiance, Skin Hazard	Time min to exceed 10 ⁴ J/m ² Unweighted UVA Irradiance, Eye Hazard
Bank of six 4 ft TL01 tubes	0.43	0.40	4	1	45
Bank of eight 2 ft TL09 tubes	0.02	0.14	67	30	2
Single 2 ft TL12 tube	0.73	1.1	2	1	80
Single 6 ft UV6 tube	0.22	0.26	4	2	40
Xe Arc lamp	0.20	0.24	2	3	70

^aThe weighted irradiance was obtained either by using a radiometer with a hazard weighted filter (International Light IL1730 with ACTS filter) or by spectroradiometer measurements combined with INIRPC hazard spectrum (6). The times given in the final two columns are those required to exceed to the maximum permissible exposure for eye and skin calculated from the spectroradiometer measurements.

Protection Measures

The National Radiological Protection Board (NRPB) in the United Kingdom has issued advice on protection against UVR (5). For workers exposed to artificial sources, administrative, engineering, and protective measures should be in place to ensure limits are not exceeded. For other situations, such as members of the public exposed to the sun, strict observance of limits is not practicable. If possible, exposure to artificial sources should be limited by engineering controls. For example, phototherapy cabins are usually fitted with interlocks so that the lamps go out when the door is opened. It is essential that staff working with UV sources are well trained and understand the nature of UV hazards. Access to suitable trained Medical Physics staff for advice on UV hazards is also important. Warning signs and access restrictions should be used where exposure above the MPE level is possible. Personal protective equipment (such as face shield and gloves) should be used if exposure to UVR above the MPE cannot be avoided.

The British Standards Institute provide information on the performance specifications of UV filters for personal eye protection (87). Percentage UV transmission should be <0.0003% at 313 nm and <7% at 365 nm and luminous transmission should be between 43 and 58%.

Patient Safety

Good staff training is vital to minimize risks of patient UV overexposure during phototherapy. There must be adequate protection against electrical hazards. Patients should not come into contact with bare lamps. There is a risk of injury if the patient falls against the lamps and it is now common practice to place a UV transmitting sheet between the lamps and the patient.

Because psoralen is deposited in the lens of the eye, there is a risk of cataract induction if the eye is exposed to UVA in the 12 h following psoralen ingestion. Consequently, the patient should avoid unnecessary sunlight exposure for the remainder of the day following PUVA treatment and they should wear appropriate eye protection. These may either be sunglasses, prescription lenses, or clear safety spectacles. Contact lenses even if marked as UV protective may not afford sufficient protection (88). The suitability of UV protective eyewear may be assessed using a spectrophotometer to fully characterise the transmission and comparison with limits suggested by Moseley et al. (89). Alternatively, a simpler approach (50) is to measure the lens transmission of radiation from PUVA therapy lamps using a hand-held narrow-band meter.

Hazards from Ozone

Ozone is a colorless, toxic irritant gas formed by a photochemical reaction between short wavelength UVR and oxygen in the air. It is possible to find ozone near UV lamps, especially where radiation <250 nm is transmitted through the envelope of the lamp. If ozone production is suspected, then adequate ventilation should be provided to remove the hazard.

ACKNOWLEDGMENTS

This article is based on that in the first edition by Prof. Brian Diffey and certain sections (e.g., the historical introduction) are taken directly from that source. Brian Diffey provided spectra for some lamps and the skin cancer action spectrum. Brian Diffey and Peter Farr were both very helpful in discussing certain parts of the text and providing useful references. Some of the practical measurements were made by John Frame. Steve Burnet and Muzz Hanliffa provided some of the pictures.

BIBLIOGRAPHY

- McGregor JM. The history of human photobiology. In: Hawk JLM, editor. *Photodermatology*. London: Arnold; 1999.
- Sollux publishing. *Actinotherapy techniques*. Slough: Sollux Publishing; 1933.
- Diffey BL. What is light? *Photodermatol Photoimmunol Photomed* 2002;18(2):68–74.
- National Radiological Protection Board, Health effects from ultraviolet radiation. Didcot: NRPB; 2002.
- NRPB. Advice on protection against Ultraviolet radiation. Didcot: NRPB; 2002.
- International Commission on Non-Ionizing Radiation Protection. Guidelines on limits of exposure to ultraviolet radiation of wavelengths between 180 nm and 400 nm (incoherent optical radiation). *Health Phys* 2004;87(2):171–186.
- Harrison GI, Young AR. Ultraviolet radiation-induced erythema in human skin. *Methods* 2002;28(1):14–19.
- McKinley A, Diffey B. A reference action spectrum for Ultraviolet induced erythema in human skin. *CIE J* 1987;6:17–22.
- CIE. Erythral reference action spectrum and standard erythral dose (CIE S 007/E-1998); 1998.
- CIE. Report 138/2: Action spectrum for photocarcinogenesis (non-melanoma skin cancers). Wien: CIE; 2000.
- Diffey BL, Oliver RJ, Farr PM. A portable instrument for quantifying erythema induced by ultraviolet radiation. *Br J Dermatol* 1984;111(6):663–672.
- Gordon PM, Saunders PJ, Diffey BL, Farr PM. Phototesting prior to narrowband (TL-01) ultraviolet B phototherapy. *Br J Dermatol* 1998;139(5):811–814.
- Waterston K, Naysmith L, Rees JL. Physiological variation in the erythral response to ultraviolet radiation and photoadaptation. *J Invest Dermatol* 2004; 123(5):958–964.
- Diffey BL. The future incidence of cutaneous melanoma within the UK. *Br J Dermatol* 2004;151(4):868–872.
- Diffey B. Human exposure to solar ultraviolet radiation. *J Cos Dermatol* 2002;1:124–130.
- Diffey BL. Human exposure to ultraviolet radiation. In: Hawk JLM, editor. *Photodermatology*. London: Arnold; 1999.
- Diffey BL. Sources and measurement of ultraviolet radiation. *Methods* 2002;28(1):4–13.
- World Health Organisation. Intersun, the global UV project; 2003.
- Hersh P, Carr J. Excimer laser photorefractive keratectomy. *Ophthalmic Practice* 1995;13:126–133.
- Anders A, Altheide HJ, Knalmann M, Tronnier H. Action spectrum for erythema in humans investigated with dye lasers. *Photochem Photobiol* 1995;61(2):200–205.
- World Health Organization. Global solar UV index: a practical guide. 2002.
- Wilson AD. Optical radiation detectors. In: Diffey B, editor. *Radiation measurement on photobiology*. London: Academic Press; 1989.

23. Oliver H, Moseley H. The use of diode array spectroradiometers for dosimetry in phototherapy. *Phys Med Biol* 2002;47:4411–4421.
24. Taylor DK, Anstey AV, Coleman AJ, Diffey BL, Farr PM, Ferguson J, Ibbotson S, Langmack K, Lloyd JJ, McCann P, Martin CJ, Menage Hdu P, Moseley H, Murphy G, Pye SD, Rhodes LE, Rogers S. Guidelines for dosimetry and calibration in ultraviolet radiation therapy: a report of a British Photodermatology Group workshop. *Br J Dermatol* 2002;146(5):755–763.
25. Coleman AJ, Collins M, Saunders JE. Traceable calibration of ultraviolet meters used with broadband, extended sources. *Phys Med Biol* 2000;45(1):185–196.
26. Lloyd JJ. Variation in calibration of hand-held ultraviolet (UV) meters for psoralen plus UVA and narrow-band UVB phototherapy. *Br J Dermatol* 2004;150(6):1162–1166.
27. Diffey BL, Roelands R. Status of ultraviolet A dosimetry in methoxsalen plus ultraviolet A therapy. *J Am Acad Dermatol* 1986;15(6):1209–1213.
28. Diffey B. Ultraviolet radiation dosimetry with polysulphone film. In: Diffey B, editor. *Radiation Measurement in Photobiology*. London: Academic Press; 1989.
29. Cabaj A, Sommer R. Measurement of Ultraviolet radiation with biological dosimeters. *Rad Protection Dosimetry* 2000;91:139–142.
30. Ell C. Improving endoscopic resolution and sampling: fluorescence techniques. *Gut* 2003;52(Suppl IV): iv30–iv33.
31. Ferguson J. Diagnosis and treatment of the common idiopathic photodermatoses. *Aust J Dermatol* 2003;44(2):90–96.
32. Ferguson J, Ibbotson S. The idiopathic photodermatoses. *Semin Cutan Med Surg* 1999;18(4):257–273.
33. Bilsland D, Diffey BL, Farr PM, Ferguson J, Gibbs NK, Hawk JL, Johnson BE, Magnus IA, Moseley H, Murphy GM. Diagnostic phototesting in the United Kingdom. British Photodermatology Group. *Br J Dermatol* 1992;127(3):297–299.
34. van de Pas CB, Hawk JL, Young AR, Walker SL. An optimal method for experimental provocation of polymorphic light eruption. *Arch Dermatol* 2004;140(3):286–292.
35. Das S, Lloyd JJ, Walshaw D, Farr PM. Provocation testing in polymorphic light eruption using fluorescent ultraviolet (UV) A and UVB lamps. *Br J Dermatol* 2004;151(5):1066–1070.
36. Farr PM. Irradiation testing of the skin. In: Hawk JLM, editor. *Photodermatology*. London: Arnold; 1999.
37. Diffey BL, Farr PM. The normal range in diagnostic phototesting. *Br J Dermatol* 1989;120(4):517–524.
38. Bruynzeel DP, Ferguson J, Andersen K, Goncalo M, English J, Goossens A, Holzle E, Ibbotson SH, Lecha M, Lehmann P, Leonard F, Moseley H, Pigatto P, Tanew A. Photopatch testing: a consensus methodology for Europe. *J Eur Acad Dermatol Venereol* 2004;18(6):679–682.
39. Dootson G, Norris PG, Gibson CJ, Diffey BL. The practice of ultraviolet phototherapy in the United Kingdom. *Br J Dermatol* 1994;131(6):873–877.
40. Ibbotson SH, Bilsland D, Cox NH, Dawe RS, Diffey B, Edwards C, Farr PM, Ferguson J, Hart G, Hawk J, Lloyd J, Martin C, Moseley H, McKenna K, Rhodes LE, Taylor DK. An update and guidance on narrowband ultraviolet B phototherapy: a British Photodermatology Group Workshop Report. *Br J Dermatol* 2004;151(2):283–297.
41. Green C, Diffey BL, Hawk JL. Ultraviolet radiation in the treatment of skin disease. *Phys Med Biol* 1992;37(1):1–20.
42. Parrish JA, Jaenicke KF. Action spectrum for phototherapy of psoriasis. *J Invest Dermatol* 1981;76(5):359–362.
43. Ortel B, Honigsmann H. Phototherapy and Photochemotherapy. In: Hawk JLM, editor. *Photodermatology*. London: Arnold; 1999.
44. Halpern SM, Anstey AV, Dawe RS, Diffey BL, Farr PM, Ferguson J, Hawk JL, Ibbotson S, McGregor JM, Murphy GM, Thomas SE, Rhodes LE. Guidelines for topical PUVA: a report of a workshop of the British photodermatology group. *Br J Dermatol* 2000;142(1):22–31.
45. British Photodermatology Group, British Photodermatology Group guidelines for PUVA. *Br J Dermatol* 1994;130(2):246–255.
46. Farr PM, Diffey BL, Higgins EM, Matthews JN. The action spectrum between 320 and 400 nm for clearance of psoriasis by psoralen photochemotherapy. *Br J Dermatol* 1991;124(5):443–448.
47. de Berker DA, Sakuntabhai A, Diffey BL, Matthews JN, Farr PM. Comparison of psoralen-UVB and psoralen-UVA photochemotherapy in the treatment of psoriasis. *J Am Acad Dermatol* 1997;36(4):577–581.
48. Mackenzie LA. The analysis of the ultraviolet radiation doses required to produce erythmal responses in normal skin. *Br J Dermatol* 1983;108(1):1–9.
49. Diffey BL, De Berker DA, Saunders PJ, Farr PM. A device for phototesting patients before PUVA therapy. *Br J Dermatol* 1993;129(6):700–703.
50. Diffey BL, Hart G. Ultraviolet and blue-light phototherapy—principles, sources, dosimetry and safety. New York: 1997.
51. Dawe RS. Ultraviolet A1 phototherapy. *Br J Dermatol* 2003;148(4):626–637.
52. Farr PM, Besag JE, Diffey BL. The time course of UVB and UVC erythema. *J Invest Dermatol* 1988;91(5):454–457.
53. Das S, Lloyd JJ, Farr PM. Similar dose-response and persistence of erythema with broad-band and narrow-band ultraviolet B lamps. *J Invest Dermatol* 2001;117(5):1318–1321.
54. Ibbotson SH, Farr PM. The Time-Course of Psoralen Ultraviolet A (PUVA) Erythema. *J Invest Dermatol* 1999;113(3):346–350.
55. Allan W, Diffey BL. A device for minimizing the risk of overexposure of patients undergoing phototherapy. *Photodermatol Photoimmunol Photomed* 2002;18(4):199–200.
56. Langmack KA. An insight into the contributions of self-shielding and lamp reflectors to patient exposure in phototherapy units. *Phys Med Biol* 1998;43(1):207–214.
57. Moseley H. Scottish UV dosimetry guidelines, “ScUViDo”. *Photodermatol Photoimmunol Photomed* 2001;17(5):230–233.
58. Martin CJ, Clouting H, Aitken A. A study of the correction factor for ultraviolet phototherapy dose measurements made by the indirect method. *Br J Dermatol* 2003;149(6):1227–1231.
59. Currie GD, Evans AL, Smith D, Martin CJ, McCalman S, Bilsland D. An automated dosimetry system for testing whole-body ultraviolet phototherapy cabinets. *Phys Med Biol* 2001;46(2):333–346.
60. Evans AL, Martin CJ, Smith DC, Currie GD, McCalman S, Bilsland D, Dunn S. Instrument for scanning the angular variation of irradiance in ultraviolet phototherapy cabinets. *J Med Eng Technol* 2002;26(3):126–131.
61. Autier P. Perspectives in melanoma prevention: the case of sunbeds. *Eur J Cancer* 2004;40(16):2367–2376.
62. McGinley J, Martin CJ, MacKie RM. Sunbeds in current use in Scotland: a survey of their output and patterns of use. *Br J Dermatol* 1998;139(3):428–438.
63. Diffey BL, Farr PM. Tanning with UVB or UVA: an appraisal of risks. *J Photochem Photobiol B* 1991;8(2):219–223.
64. Phillips. Welcome to Phillips Tanning; <http://www.lighting.phillips.com> 2004.
65. Young AR. Tanning devices—fast track to skin cancer? *Pigment Cell Res* 2004;17(1):2–9.
66. Diffey B. Sunbeds: What are they, who uses them and what are the health effect? London: Health Education Authority; 1997.

67. British Standards Institute. Safety of household and similar electrical appliances. Part 2 Section 2.7 Skin Exposure to ultraviolet and infrared radiation (BS EN 60335-2-27: 1997); 1997.
68. Das S, Lloyd JJ, Walshaw D, Diffey BL, Farr PM. Response of psoriasis to sunbed treatment: comparison of conventional ultraviolet A lamps with new higher ultraviolet B-emitting lamps. *Br J Dermatol* 2002;147(5):966–972.
69. Diffey BL. A quantitative estimate of melanoma mortality from ultraviolet A sunbed use in the U.K. *Br J Dermatol* 2003;149(3):578–581.
70. Turner RJ, Farr PM, Walshaw D. Many patients with psoriasis use sunbeds (letter). *Br Med J* 1998;317:412.
71. Turner RJ, Walshaw D, Diffey BL, Farr PM. A controlled study of ultraviolet A sunbed treatment of psoriasis. *Br J Dermatol* 2000;143(5):957–963.
72. Sarkany RP, Anstey A, Diffey BL, Jobling R, Langmack K, McGregor JM, Moseley H, Murphy GM, Rhodes LE, Norris PG. Home phototherapy: report on a workshop of the British Photodermatology Group. December 1996. *Br J Dermatol* 1999;140(2):195–199.
73. Cameron H, Yule S, Moseley H, Dawe RS, Ferguson J. Taking treatment to the patient: development of a home TL-01 ultraviolet B phototherapy service. *Br J Dermatol* 2002;147(5):957–965.
74. Wilson SE. Clinical practice. Use of lasers for vision correction of nearsightedness and farsightedness. *N Engl J Med* 2004;351(5):470–475.
75. Bower KS, Weichel ED, Kim TJ. Overview of refractive surgery. *Am Fam Phys* 2001;64(7):1183–1190.
76. Ambrosio R, Jr., Wilson S. LASIK vs LASEK vs PRK: advantages and indications. *Semin Ophthalmol* 2003;18(1):2–10.
77. National Institute of Clinical Excellence Interventional Procedure Guidance 102: Laser in situ keratomileusis for the treatment of refractive errors. 2004.
78. Edelson R, Berger C, Gasparro F, Jegasothy B, Heald P, Wintroub B, Vonderheid E, Knobler R, Wolff K, Plewig G. Treatment of cutaneous T-cell lymphoma by extracorporeal photochemotherapy. Preliminary results. *N Engl J Med* 1987;316(6):297–303.
79. McKenna KE, Whittaker S, Taylor P, Lloyd J, Ibbotson S, Rhodes LT, Russell-Jones R. Evidence based practice of photopheresis: a report of a workshop of the British Photodermatology Group and UK Skin Lymphoma Group. *Br J Dermatol* 2005; in press.
80. Roelandts R. The history of phototherapy: something new under the sun?. *J Am Acad Dermatol* 2002;46(6):926–930.
81. Thai TP, Houghton PE, Campbell KE, Woodbury MG. Ultraviolet light C in the treatment of chronic wounds with MRSA: a case study. *Ostomy Wound Manage* 2002;48(11):52–60.
82. Conner-Kerr TA, Sullivan PK, Gaillard J, Franklin ME, Jones RM. The effects of ultraviolet radiation on antibiotic-resistant bacteria in vitro. *Ostomy Wound Manage* 1998;44(10):50–56.
83. Brickner PW, Vincent RL, First M, Nardell E, Murray M, Kaufman W. The application of ultraviolet germicidal irradiation to control transmission of airborne disease: bio-terrorism countermeasure. *Public Health Rep* 2003;118(2):99–114.
84. Diffey B, Langley FC. IPEM report 49: Evaluation of Ultraviolet Radiation Hazards in Hospitals. London: Institute of Physical Sciences in Medicine; 1986.
85. Chuck A, Todd J, Diffey B. Subliminal ultraviolet-B irradiation for the prevention of vitamin D deficiency in the elderly: a feasibility study. *Photoderm Photomed Photomed* 2001;17(4): 168–171.
86. International Non-ionizing Radiation Committee of the International Radiation Protection Association. Proposed change to the IRPA 1985 guidelines on limits of exposure to ultraviolet radiation. *Health Phys* 1989;56(6):971–972.
87. British Standards institute. BS EN 270, Personal Eye protection—Ultraviolet filters—transmittance requirements and recommended use. 2002.
88. Anstey A, Taylor D, Chalmers I, Ansari E. Ultraviolet radiation-blocking characteristics of contact lenses: relevance to eye protection for psoralen-sensitized patients. *Photodermatol Photoimmunol Photomed* 1999;15(5):193–197.
89. Moseley H, Cox NH, Mackie RM. The suitability of sunglasses used by patients following ingestion of psoralen. *Br J Dermatol* 1988;118(2):247–253.

See also FLUORESCENCE MEASUREMENTS; SKIN, BIOMECHANICS OF.

UMBILICAL ARTERY AND VEIN MONITORING. See MONITORING, UMBILICAL ARTERY AND VEIN.

VAGINA, EXAMINATION OF. See COLPOSCOPY.

VASCULAR GRAFT PROSTHESIS

KANDICE KOTTKE-MARCHANT
The Cleveland Clinic
Foundation
Cleveland, Ohio

COBY LARSEN
Case Western Reserve
University
Cleveland, Ohio

INTRODUCTION

The system of blood vessels in the body is crucial to transport cells, oxygen, and nutrients to the vital organs. When injured by diseases, such as atherosclerosis, the blood vessels may become occluded, leading to decreased blood flow and organ damage, or may be weakened and rupture due to aneurysmal dilatation. One method of treatment is to use surgery to bypass diseased blood vessels by using an artificial blood vessel substitute, or vascular graft prosthesis. Materials currently used in vascular graft prostheses include polyethylene terephthalate (Dacron) and expanded polytetrafluoroethylene (ePTFE). These devices successfully function as vascular grafts in larger diameter applications, but suitable materials for small diameter prostheses (<5 mm) are still being developed. Current trends in the development of small diameter prostheses include surface modification to prevent thrombosis, incorporation of endothelial cells, and a tissue engineering approach with the development of a biological blood vessel based upon a biodegradable material scaffold.

THE CLINICAL NEED FOR VASCULAR GRAFT PROSTHESES

Normal Blood Vessel Anatomy

The largest artery leaving the left ventricle of the heart is the aorta, which is ~20 mm in diameter. The aorta branches into progressively smaller arteries (<5 mm internal diameter) that feed blood and nutrients to organs, such as the brain (carotid arteries), liver (hepatic artery) and kidneys (renal arteries), and to the extremities of the arms (cephalic arteries) and legs (iliac and femoral arteries). The smaller arteries branch further to become smaller arterioles (<1 mm), which branch even further to become the smallest capillaries (<100 μ m), through which only single cells can pass. The capillaries are the site of most oxygen and nutrient exchange into the tissues. Capillaries then join together to form venules, small veins, and then larger veins that carry deoxygenated blood back to the lungs and heart. The design of replacement blood vessels should take

into account both the structure and function of native vessels for optimal function.

Normal arteries have three different layers: the internal intima, the middle media, and the external adventitia (1). (see Fig. 1). The intima is composed of a layer of endothelial cells, with underlying extracellular matrix proteins delimited by the internal elastic lamina. Endothelial cells are responsible for preventing blood vessel thrombosis under normal physiologic conditions by preventing platelet adhesion, coagulation, and thrombus formation (2,3). (see Fig. 2). This is accomplished by the expression and secretion of antiplatelet agents, such as ecto-ADPase, prostacyclin (PGI₂) and nitric oxide (NO) (2). Endothelial cells express thrombomodulin (TM) and release tissue factor pathway inhibitor (TFPI) to prevent activation of the coagulation cascade. Endothelial cells also express heparan sulfate to bind circulating antithrombin and rapidly inhibit local thrombin formation (2). If a fibrin clot is formed, endothelial cells can release factors that stimulate the fibrinolytic system to degrade the thrombus, such as tissue plasminogen activator (tPA) (2). Upon cytokine or thrombin stimulation, endothelial cells can develop a procoagulant phenotype with expression of tissue factor (TF) and release of vWf, PAI-1, and coagulation factors (6). The intima is present in all vessels, with the smallest capillaries composed of only a single layer of endothelial cells resting upon a thin basement membrane of extracellular matrix proteins.

The media varies in composition from large elastic arteries, such as the aorta, where it is composed of layers of elastic tissue and glycosaminoglycans to the smaller muscular arteries, such as the coronary arteries in the heart, where it is composed predominantly of vascular smooth muscle cells (1). The elastic and mechanical properties of the media allow vessels to contract or relax to maintain the luminal pressure and volumetric flow rate. Arterioles may have only a single layer of smooth muscle cells and capillaries have no media at all.

The external blood vessel layer, or adventitia, serves to supply nutrients to the cells in the blood vessel itself. It consists of a loose layer of connective tissue, with small arterioles and capillaries (vasa vasorum or "vessels of the vessel") feeding the blood vessel. It also contains a network of nerves that can stimulate medial smooth muscle cell contraction or relaxation. The adventitia is prominent in larger vessels, becoming progressively smaller as the vessel diameter decreases. Arterioles and capillaries lack an adventitia completely.

Coronary Atherosclerosis

Atherosclerotic cardiovascular disease, in its many guises, is the leading cause of mortality in the United States, resulting in coronary artery disease, myocardial infarction, stroke, aneurysms, and peripheral vascular disease. Atherosclerosis is a multifactorial disorder of the arterial vascular tree where lipid dysregulation and vascular

Blood Vessel Structure

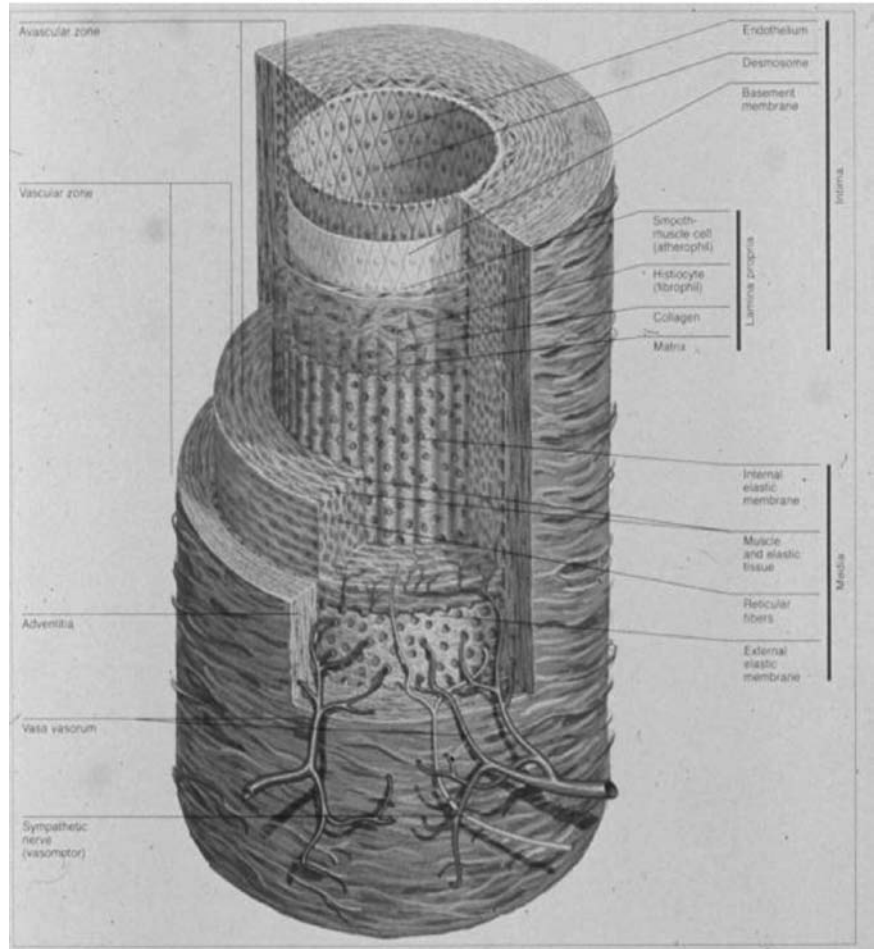


Figure 1. Arteries are composed of three layers: the intima, media, and adventitia. The intima is lined by endothelial cells and bounded by an internal elastic lamina. The media maintains vascular tone and is composed of smooth muscle cells or layers of elastic tissue. The adventitia supports the vessel and provides nutritional and neural stimulation.

inflammation leads to endothelial injury, vascular lipid accumulation, and calcification with the development of lipid-laden deposits, or “plaques” in the vessel wall that eventually lead to stenosis of the vascular lumen (7). The resulting stenosis or narrowing of the blood vessels compromises blood flow through the artery. Rupture of the atherosclerotic plaque with subsequent thrombosis may lead to complete occlusion of the blood vessels. Depending on the location of the artery, this stenosis or occlusion leads to decreased distal blood flow and end-organ dysfunction. Examples include infarction of cardiac muscle with complete coronary artery occlusion or stroke with occlusion of carotid or cerebral arteries.

Despite advances in prevention and early diagnosis, cardiovascular disease claims more lives than the next five leading causes of death, combined (8,9). Coronary atherosclerosis alone, leading to stenosis and occlusion of the small coronary arteries that nourish the cardiac muscle, is the single leading cause of death in American today, responsible for 494,000 deaths in 2002 (8,9). The current modalities for treating coronary artery disease include intravascular angioplasty, stenting, and bypass of the stenotic lesions using saphenous vein or internal mammary artery grafts. Coronary bypass surgery was performed in over one-half million individuals in the United States alone in 2002 (10,11), with the majority of

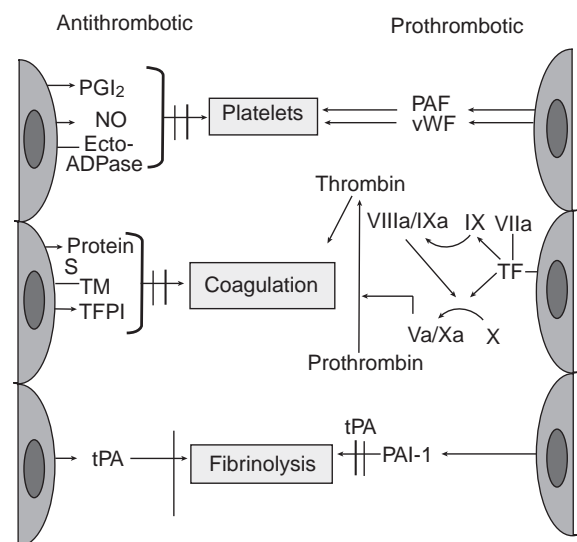


Figure 2. Endothelial cell phenotypes. Endothelial cells can have an antithrombotic (left) or prothrombotic (right) phenotype. (Adapted from Refs. 4 and 5). PGI₂-prostacyclin, NO-nitric oxide, Ecto-ADPase ecto-adenosine diphosphate, TM-thrombomodulin, TFPI-tissue factor pathway inhibitor, tPA-tissue plasminogen activator, PAF-platelet activating factor, vWf-von Willebrand factor, TF-tissue factor, PAI-1-plasminogen activator inhibitor-1.

individuals requiring bypass of three or more vessels. Unfortunately, in a significant number of individuals, autologous veins and arteries are not available, and there is a need for readily accessible, functional, alternative small diameter vascular graft prostheses.

Cerebrovascular Disease

Atherosclerotic vascular disease also affects the cerebral vessels and peripheral arterial tree, leading to stroke, renal failure, and arterial claudication. Stroke is the third leading cause of death, with ~275,000 Americans dying of stroke in 2002 (9). One of the leading causes of stroke is atherosclerotic stenosis of the carotid arteries in the neck, leading to thromboembolic stroke (11). Due to the danger of any further thrombosis or embolism, vascular prostheses have not been used to treat carotid artery disease, and therapy is currently limited largely to carotid endarterectomy, with removal of vascular plaque, and limited use of carotid stents (12). Use of vascular prostheses for cerebrovascular disease will not be feasible until prostheses are developed that pose no risk for distal thromboembolism.

Peripheral Vascular Disease

Peripheral arterial disease, with atherosclerotic occlusion of peripheral arteries predominantly in the lower extremities, leads to claudication and leg pain in ~8–12 million Americans (13,14); while it is not a major cause of mortality, it is associated with significant morbidity (9,13). Current therapies for treating peripheral arterial disease include angioplasty and vascular bypass. Prosthetic grafts have been successfully employed to treat atherosclerosis of the aorto-iliac bifurcation, but saphenous vein bypass is still the most widely used therapy for femoral-popliteal and below-the-knee bypass surgery (15–17). Suitable saphenous veins may not be available in ~15% of patients (17), often due to prior use for coronary artery bypass surgery. There is a clinical need to develop small diameter vascular prostheses that remain patent for long periods to more effectively treat peripheral arterial disease, and provide a viable alternative to saphenous veins.

Aneurysms

An aneurysm is a dilatation of a blood vessel wall that may be circumferential (fusiform aneurysm) or may involve only a portion of the vessel circumference (saccular aneurysm). Aneurysms most frequently affect the abdominal aortic wall and are a significant cause of mortality due to aneurysm rupture, resulting in 1.3% of all deaths (18). Aneurysms can affect other arteries, such as the cerebral arteries, renal arteries, and splenic arteries. Aneurysms usually do not occlude blood flow, but the dilated vessel wall is weakened mechanically and can rupture, leading to extravasation of blood into the surrounding tissues. Infra-renal aortic aneurysms 5.5 and 6.5 cm have annual rupture rates of 11 and 26%, respectively (19). Abdominal aortic aneurysms are usually observed in older patients in association with atherosclerosis, but other factors, such as proteolytic degradation of elastin and collagen by matrix metalloproteinases and plasmin is thought to be contrib-

utory (20–22). Therapy has generally been surgical, with the placement of an intraluminal vascular graft (23,24) or an endoluminal stent-graft (25).

Trauma

Vascular trauma from gun shot wounds, knife wounds, motor vehicle accidents or accidental dismemberment may result in laceration or transection of blood vessels. This injury may necessitate replacement with a vascular graft, if the vessel wall cannot be surgically repaired or an autologous saphenous vein is not available (26). Traumatic injury of large diameter vessels, such as the pulmonary artery or aorta usually requires a synthetic vascular graft or patch, as autologous tissue of large diameter is not available. One unique consideration with vascular trauma is bacterial contamination due to open wounds or colon injury, and remote bypass grafts through clean subcutaneous tissues may be desirable. In cases of dismemberment, the blood vessels usually retract into the severed tissue, and microvascular repair methods are usually required to reestablish blood flow (27). This is also a scenario where an off-the-shelf small diameter vascular prosthesis would be desirable.

THE HISTORY OF VASCULAR REPLACEMENTS

Arterial aneurysms have been noted back to antiquity, when the Roman historian Antyllus first reported their tendency to rupture in the second century AD (28). Treatment dates back to 1684, when Moore attempted to induce thrombosis of an aneurysm by introducing large masses of intraluminal wires (23,29). The first reported surgical therapy was ligation of the abdominal aorta proximal to a leaking iliac artery aneurysm by Sir Astley Cooper in 1817 (28). Aneurysmectomy, with excision of the aneurysm and vascular repair was reported by Cooley and DeBakey in 1952 (4). Early vascular grafts employed aortic wrapping or the use of preserved or freeze-dried homograft replacement (30,31). Use of an intraluminal graft replacement was instituted by Creech in 1966, and this remains the standard treatment today (23). In 1991, Parodi introduced a new technique of treating abdominal aortic aneurysms with a stent-graft that could be inserted endolumenally through an incision in the femoral artery (25).

The use of synthetic materials as vascular substitutes for aneurysms and other vascular bypass grafts was initiated in 1952 with the use of Vinyon-N cloth as an arterial conduit (5). This was followed by the use of polyethylene terephthalate (Dacron) as a vascular conduit in 1957 (32,33). In 1969, Gore patented a microexpanded ePTFE, which was first used as an arterial conduit in 1973 (34,35). Umbilical veins were further developed with use of glutaraldehyde cross-linking and an external wrapping of polyester mesh (36). These synthetic prostheses showed some success in large diameter vascular applications, but were plagued by thrombosis and occlusion due to intimal hyperplasia in small diameter (<5 mm) applications. Lining of synthetic vascular prostheses with endothelial cells to improve thromboresistance was first described by Herring in 1978 (37). Development of a blood

vessel by in vitro culture of biologic tissues (i.e. tissue engineering) was first described in 1986, with a multilayer structure of collagen and vascular cells (38).

CURRENT VASCULAR GRAFT PROSTHESES

Vein Grafts

The saphenous vein is a long vein arising in the calf and thigh with extensive collateral circulation. It has been utilized extensively as a vascular graft due to its long size and diameter that is similar to coronary arteries and smaller arteries in the leg, such as the femoral or popliteal arteries. The use of veins as bypass conduits to surgically treat coronary atherosclerosis was first described by Kunlin in 1949 (39), and this procedure remains the standard method of coronary artery bypass surgery to this day. For coronary bypass, saphenous veins are harvested surgically or endoscopically and the quality of the veins assessed by inflation, with valves stripped prior to use, if necessary (40). Current coronary artery bypass techniques using saphenous veins have an occlusion rate of 10–15% at 1 year after surgery, increasing to 30–40% at 10 years after surgery (41). Improved patency has been observed using arterial conduits, such as the internal mammary artery and radial artery grafts (42).

When used as peripheral vascular bypass grafts for the femoral or popliteal arteries in the leg, saphenous veins can be dissected and the direction reversed to allow flow through the valves, but they can also be utilized *In situ* and the valves stripped to allow retrograde flow (43). It is thought that the *In situ* graft should have lesser disruption to the vein endothelium and adventitia compared to the reversed vein conduit (44), but 2 year clinical patency rates for femoropopliteal bypass are similar at ~82% for both procedures (45). Not all patients have an appropriate saphenous vein for either cardiac or peripheral bypass surgery. This is most frequently due to prior surgical harvest, but also may be from poor vein quality due to insufficiently small diameter or increased branching. In these patients, suitable synthetic vascular conduits are desirable.

Dacron

Dacron is a polyester fabric, polyethylene terephthalate. Clinically utilized Dacron vascular prostheses are manufactured as textiles, either knitted or woven (34) (Fig. 3). The original design of a Dacron vascular graft as a textile versus a solid tube was an attempt to allow tissue healing of the prosthesis through the pores of the material. The knitted prostheses have higher porosity than the more densely fabricated woven prostheses and require preclotting or use of a surface coating, such as gelatin, collagen, fibrin glue, or albumin, to prevent transluminal leaking after implantation (34,46,47) (Fig. 4). The woven grafts are most commonly used in the thoracic aorta and for ruptured abdominal aortic aneurysms to minimize blood loss (43). Most Dacron graft prostheses are crimped like a soda straw to improve graft flexibility and decrease lateral compressibility (Fig. 4). An alternative to crimping that results in a thinner prosthesis wall is the use of an external spiral winding of a stiff polymer, such as poly(propylene). In a



Figure 3. Scanning electron micrograph (SEM) of knitted Dacron vascular graft prostheses. The knitted prostheses are of lower porosity than the woven Dacron prostheses and must be sealed by preclotting prior to implantation. SEM, Original Magnification 1000 \times .

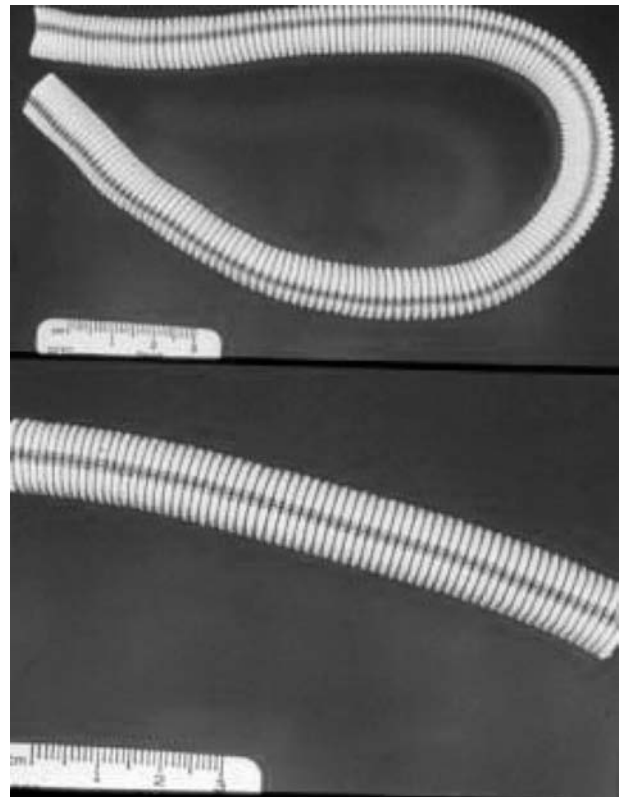


Figure 4. (top) A typical Dacron vascular graft prosthesis. Notice the crimped appearance of the prosthesis, which is utilized for flexibility and lateral compression-resistance. The blue line is to facilitate implantation without twisting, that would lead to kinking (bottom). An albumin-coated Dacron prosthesis. The albumin coating is used to seal the graft interstices and eliminate the need to preclot the graft prior to implantation. This type of coating often fails due to flaking and removal from the surface.

further attempt to foster tissue incorporation and vascular graft healing, some prostheses employ a velour surface, where thousands of individual fiber loops are pulled almost perpendicularly from the material, resulting in a surface that has a markedly increased surface area (48,49). Luminal velours are employed to trap fibrin and platelets, while external velours are used to increase adventitial healing and cellular incorporation (50).

Dacron vascular grafts have most successfully been employed in the surgical treatment of thoraco-abdominal and abdominal aortic aneurysms, where they are manufactured in a both a tubular and bifurcated aorto-bifemoral configuration. They have also been utilized for longer bypass applications, such as axillary-femoral or femoral-femoral surgery. In these longer length applications, the reinforced, eternally wound prostheses are typically utilized to prevent kinking when transversing joints. The successful treatment of abdominal aortic aneurysms with Dacron vascular grafts is proportional to the size of the aneurysm; for smaller aneurysms, the survival rates for surgical and nonsurgical populations is similar (51). The 5 year patency rates are 93% for aortic bifurcation grafts (52), with similar patency for knitted and woven grafts (53).

Dacron has not been widely used for lower extremity bypass, especially when the bypass entails vessels below the knee, due to poor patency rates. However, a recent study utilizing heparin-bonded Dacron showed favorable patency rates for femoro-popliteal bypass grafting for modified Dacron versus ePTFE (55% vs. 42%; $p < 0.044$) at 3 years, with similar patency rates at 5 years (45% vs. 35%; $p < 0.055$) (54).

Expanded Polytetrafluoroethylene

Polytetrafluoroethylene (PTFE-Teflon) is a chemically inert, hydrophobic polymer that resists long-term *In vivo* degradation. As mentioned above, PTFE has been manufactured in a microporous fabric for vascular grafts (ePTFE). The ePTFE is manufactured by a heating, stretching, and extruding process to form solid nodes of PTFE separated by many thin fibrils, with an internodal distance of $\sim 30 \mu\text{m}$ (34) (see Fig. 5). This same material, known as Goretex, is also popular for waterproof sporting applications, such as jackets and shoes, as the material hydrophobicity prevents gross water flux while allowing air transport. The same material properties that make ePTFE attractive for sporting materials make it attractive as a vascular graft. The hydrophobicity is designed to prevent surface thrombosis, while the microporous texture prevents transluminal graft leaking, while encouraging cell incorporation and healing. In applications where increased structural stability is required, an external winding or wrap of polypropylene is often employed.

Patency rates of ePTFE for aorto-bifemoral bifurcation grafts are 95% at 5 years, similar to Dacron (52,55). In extraanatomic bypass, ePTFE has shown a cumulative patency of 83% for femoro-femoral bypass and 75% for axillo-femoral bypass at 5 years (56). In lower extremity revascularization, saphenous vein bypass remains the

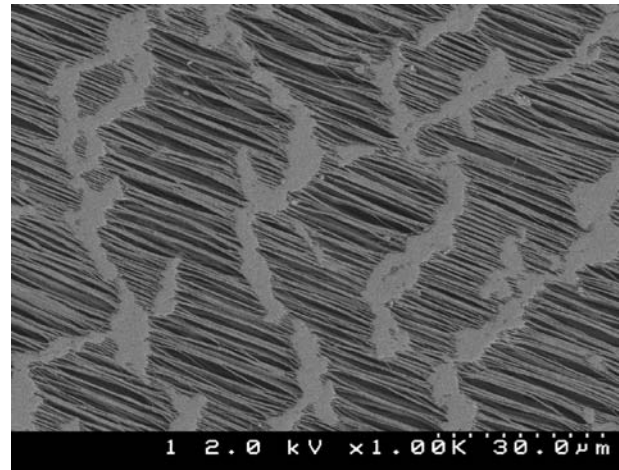


Figure 5. An SEM of expanded ePTFE vascular graft. Note the solid nodes of ePTFE separated by thin fibrils. The typical internodal distance of most clinical prostheses is $30 \mu\text{m}$. Due to the hydrophobic nature of ePTFE, this structure makes the material microporous, allowing limited cell migration, but limiting bulk water leakage. Original magnification $1000\times$.

graft of choice for below-knee applications. However, there is still controversy regarding the optimal material for above-knee femoro-popliteal bypass. Many studies report five year patency rates $< 50\%$ for ePTFE in femoro-popliteal bypass (54,57), with 5 and 10 year saphenous vein patency rates at 77 and 50% (58,59). However, in a retrospective study, the primary patency rate at 4 years was not significantly better for vein bypass (82.2%) versus ePTFE (80.6%) (60).

A somewhat unique use for ePTFE grafts is the arteriovenous graft used to provide vascular access to patients with renal failure undergoing hemodialysis. These grafts are interposed between the radial artery and cephalic vein in the wrist (61). Other options are the creation of a direct arterio-venous fistula without the use of a graft, and a central venous catheter. Primary patency of arteriovenous grafts is inferior to that of fistulas, irrespective of graft material or postoperative treatment with anticoagulants (62-64).

Endovascular Stent-Grafts

Historically, surgical repair of abdominal aortic aneurysms required major abdominal surgery with its attendant morbidity. For this reason, a procedure to repair aneurysms endolumenally from within the vasculature was attractive. In 1991, Parodi (25) reported the first successful use of a combined stent-graft that could be placed into the aortic aneurysm endolumenally through a femoral artery catheter without open surgery. During the past decade, many different stentgrafts have been developed for this application. They typically are composed of a fabric graft (Dacron or ePTFE) and an expandable metallic skeleton (23,65). They are usually inserted in a closed configuration and then opened via a balloon catheter. The stent is typically held in place with proximal and distal hooks. The metallic

skeletons employed are stainless steel, Elgiloy (nickel, cobalt, and chromium alloy) and Nitinol (nickel titanium alloy).

The stent-grafts may have a lower operative mortality and less frequent complications than open repair (18). However, two large European registries indicate a 3% yearly failure rate for endovascular repair versus 0.3% for open repair (66,67). Failure mechanisms unique to endovascular stent-grafts include endoleak, with blood flow between the exterior of the graft and the lumen of the aneurysm. There are also difficulties with stent-graft migration due to movement of the anchoring hooks (65). Some devices have problems with durability, due to repeated rubbing of the graft material against stent metal, with either fabric degradation or corrosion or stress cracking of the metal (68). In other patients, the aneurysm may continue to expand despite the presence of the stent-graft, termed endotension, due to transmission of lateral wall pressure through the graft.

Vascular Graft Healing

Implantation of synthetic vascular grafts that are not preclotted is associated with rapid protein adsorption, followed by platelet adhesion, inflammatory cell adhesion and variable degrees of fibrin formation (69,70). The fibrin formation is usually limited to a thin layer on the luminal surface, but this is variable depending on the diameter of the graft. The fibrin layer also includes entrapped macrophages and neutrophils, and reaches a maximal thickness ~2 weeks postimplantation (43). In preclotted Dacron grafts, there is already a fibrin-platelet layer in the graft interstices created by the preclotting process. This may be associated with increased platelet adhesion upon implantation (71). Later cellular healing may come from blood-borne cells, migration of cells from the anastomosis,

or in-growth of cells from the adventitia. This may or may not be accompanied by endothelial growth and development of a neo-intima, either due to migration from the anastomoses or implantation of circulating endothelial progenitor cells from the blood stream (72). In the absence of an endothelial layer, the fibrin matrix and inflammatory cells forms a pseudointima. There also may be an in-growth of smooth muscle cells and fibroblasts from the adventitia of the graft, depending on the porosity of the graft material. The adventitia is often associated with fibrosis and a foreign body giant cell response (see Fig. 6).

Most ePTFE grafts implanted in humans do not develop an endothelialized lumen spontaneously (73), and endothelialized pannus in-growth from the anastomosis is usually limited to only a few centimeters. Similarly, Dacron grafts usually do not re-endothelialize, but typically maintain a compact fibrin layer at the lumen. In lower porosity grafts, the graft interstices usually remain acellular. In ePTFE grafts and higher porosity Dacron prostheses, the interstices become populated with fibroblasts, macrophages, and less frequently, smooth muscle cells. Neovascularization, with new capillary in-growth, is rare.

Failure Mechanisms

The healing of implanted vascular grafts is not optimal due to lack of endothelialization or neovascularization. This may lead to clinical complications, such as thrombosis, intimal hyperplasia, and infection. Failure may be acute, midterm, or late (74). Early or acute failure is usually due to technical surgical problems or acute thrombosis. Early infection may be tied to contamination during surgery. Midterm failure (3 months to 2 years) may be due to cellular proliferation and intimal hyperplasia or infection. Late failure may be due to development of atherosclerotic lesions in the graft (75).

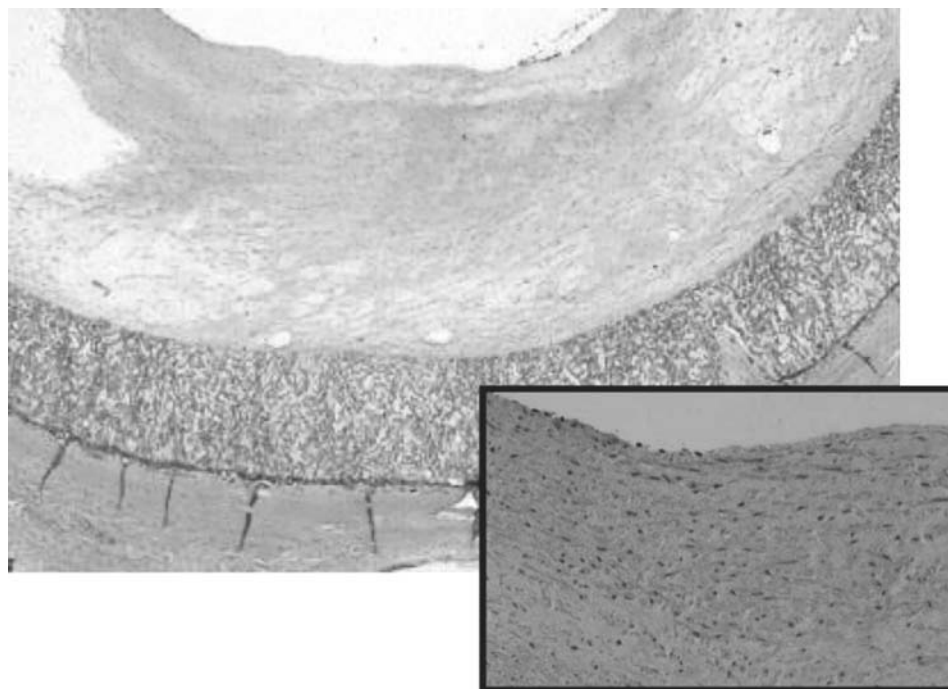


Figure 6. Healing of an ePTFE vascular graft. Note the cellular pseudointima without an obvious endothelial lining. The graft itself is sparsely populated with cells. The adventitial layer is dense compact fibrous tissue with a lining of multi-nucleated giant cells. Hematoxylin and eosin. Original magnification, 10 \times . Inset, 20 \times .

Thrombosis of Vascular Grafts

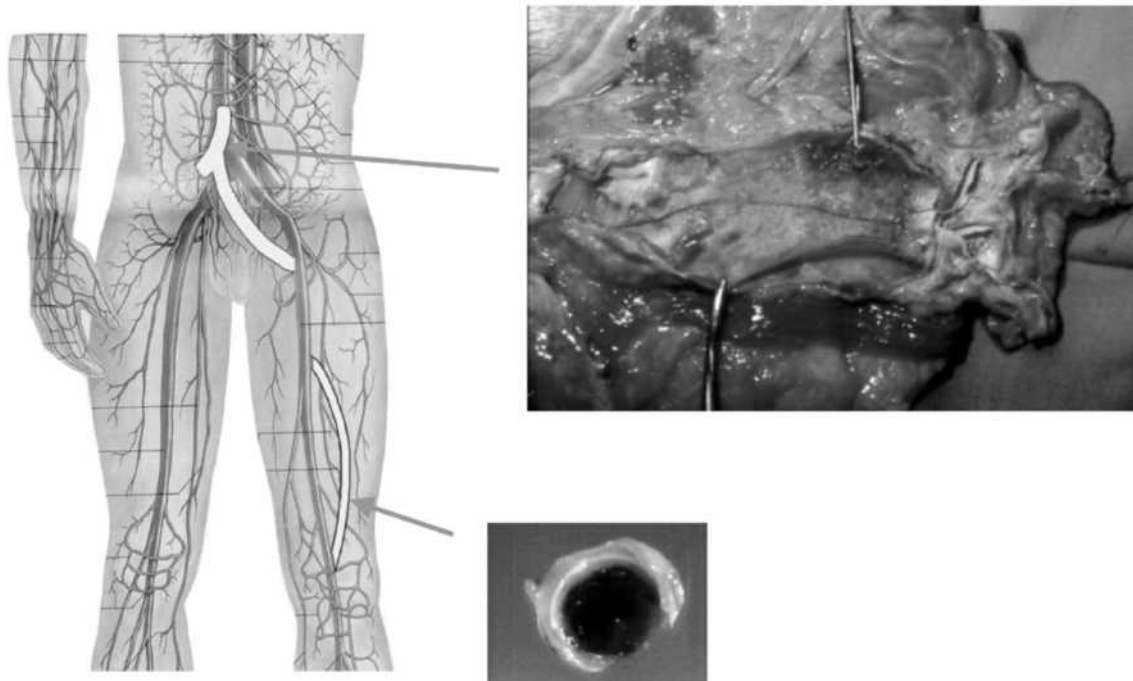


Figure 7. Thrombosis of a Dacron vascular graft in the aorta. In such a large diameter prosthesis, this degree of luminal thrombosis was of little clinical consequence. A similar degree of thrombosis in a small diameter vessel could result in complete occlusion.

The rapid adhesion of platelets and development of a luminal fibrin layer can lead to thrombosis (Fig. 7). Thrombosis is a leading cause of vascular graft failure, especially in smaller diameter prostheses, where it leads to decreased flow or occlusion. It is usually an acute or subacute failure mechanism, but may be a cause of late failure due to thrombosis super-imposed upon stenosis due to other causes of vessel narrowing, such as atherosclerosis or intimal hyperplasia. Antithrombotic therapy, particularly antiplatelet therapy, has been shown to be beneficial in decreasing graft occlusion up to 2 years after surgery (76–78).

The majority of stenotic vascular graft failures are due to hyperplasia of the tissues near the anastomosis, particularly the distal anastomosis (79,80), termed anastomotic intimal hyperplasia (81,82). This is particularly a problem with arteriovenous grafts used for dialysis access where the distal intimal hyperplasia may lead to significant stenosis of the lumen (83). The development of intimal hyperplasia is multifactorial, and its etiology is not completely known. It evolves from the tissue healing response at the anastomosis, but inciting factors include chronic inflammation, platelet adhesion with release of platelet derived growth factor (PDGF), vessel wall injury and mechanical factors, such as disturbed hemodynamic flow and compliance mismatch between the native vessel and the more rigid prosthesis (43,83–85). Alteration in cell phenotype due to interaction with the prosthesis or due to disturbed flow may lead to smooth muscle cell prolifera-

tion from altered production of basic fibroblast growth factor (bFGF) or PDGF by endothelial cells (86,87) or tumor necrosis factor-alpha (TNF α) by inflammatory cells (88). Indeed, antibodies to bFGF have been shown to decrease smooth muscle cell proliferation in ePTFE grafts experimentally (89). Recent gene-transfer experiments have suggested that increased expression of tPA and increased fibrinolysis may be associated with increased intimal hyperplasia, while increased expression of nitrous oxide synthase (eNOS) and increased expression of the platelet inhibiting molecule nitrous oxide may decrease the hyperplastic response (90).

Infection occurs in 1–6% of arterial vascular graft prostheses (43,91). The sources of infections may be contamination of the prosthesis during implantation or hematogenous seeding from bacteremia. The earlier infections are often due to *Staphylococcus aureus*, while the later infections are often due to a lower virulence organism, such as *Staphylococcus epidermidis* (92). Clinical symptoms of graft infections include fever, leukocytosis, and bacteremia; newer imaging techniques that are able to detect acute inflammation may be useful in diagnosis of graft infections (93). These infections may be resistant to antibiotics due to the lack of vascularity in the graft interstices. Complete or partial graft excision is often necessary for treatment (94). Graft modifications to decrease the risk of infection have included bonding antibiotics to graft surfaces or including antibiotics in the blood used to preclot the graft.

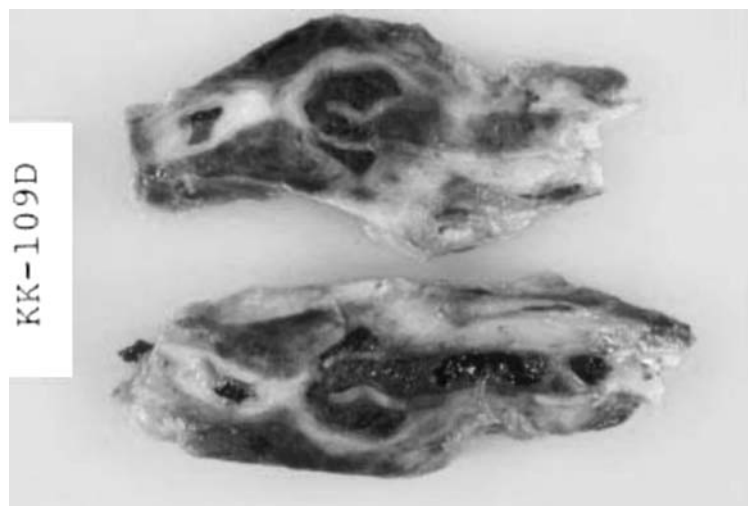


Figure 8. Gross photograph of a pseudoaneurysm at a graft anastomosis. This is not a true aneurysm as there is no failure of the graft material, but a defect between the graft and the native vessel, leading to blood leaking in the interstitial tissues. This type of pseudoaneurysm may cause graft occlusion due to external pressure from the leaking blood.

Material failure of Dacron or ePTFE grafts has been quite uncommon. Neither are substantially susceptible to *In vivo* degradation, although Dacron grafts may expand slightly *In vivo* due to flattening of the crimps or relaxation of the knitted or woven structures. Suture failure and blood leakage at the anastomosis with pseudoaneurysm formation has also been described (Fig. 8).

NEW DEVELOPMENTS IN VASCULAR GRAFT PROSTHESES

Surface Modification

Attachment of an anticoagulant, such as heparin or dipyridamole, to the graft lumen has been investigated as a method to improve small-diameter graft patency (76,95). As an example, Lin et al. (96) modified the luminal surface of a standard ePTFE with an aldehyde-modified heparin bound to layers of crosslinked polyethyleneimine (PEI), Carmeda bioactive surface (CBAS), and dextran sulfate. This graft (4 mm ID), as an exteriorized baboon femoral arteriovenous shunt, showed an 80% reduction in 1–4 h platelet deposition. When compared to contralateral uncoated grafts in a 4 week baboon aortoiliac bypass model, there was significant reduction in neointimal cell proliferation and area at both proximal (0.26 vs. 0.56 mm²) and distal (0.29 vs. 0.63 mm²) anastomoses for the heparin-coated grafts (96). One major concern with this strategy is the duration of heparin function. Delamination of portions of the layered coating or the presence of physical barriers deposited on the graft surface in the form of blood components would lead to failure for this type of approach. A heparin-bonded Dacron graft is currently available on the European market, with the heparin attached using a tridodecyl-methyl-ammonium chloride (TMAC) and the outer portion of the graft coated with collagen to decrease porosity (97). This heparin-bonded Dacron had similar 5 year patency rates to ePTFE for femoral-popliteal bypass grafting. Other surface modifications for ePTFE grafts have included carbon coating to decrease thrombogenicity (98,99), or incorporation of

growth factors to facilitate the healing process (100,101) and coating with polyurethane to enhance endothelialization (102).

Newer Biomaterials

There has been some interest in polyurethane elastomers as vascular graft materials, in an effort to develop radially compliant vascular grafts (34,85). Initial vascular grafts with polyester and polyether-based polyurethanes resulted in failure due to material degradation (103). Vectra, a polyetherurethaneurea vascular access graft, showed 12 month patency similar to ePTFE (104), and received U.S. Food and Drug Administration (FDA) clearance in 2000. More recently developed polyurethanes for vascular grafts utilize polycarbonate-based polyurethanes to impart hydrolytic and oxidative stability (85,105). One of these prostheses, Myolink, is available commercially for hemodialysis access. Some experimental hybrid polyurethane grafts have been developed in which the polyurethane is coupled with a gelatin–heparin matrix and reinforced with Dacron mesh (106). Other hybrid polyurethane grafts include a polyurethane with a biodegradable poly(ethylene glycol)–poly(lactic acid) copolymer coating (107). Polyurethanes have recently been modified to incorporate a diazeniumdiolate-modified nitric oxide (NO)-producing peptide, which can release NO, a potent platelet-inhibiting compound (108). A similar NO-producing polyurethane has recently been shown to decrease thrombus formation in a sheep model (109).

Endothelial Incorporation

Current clinically used large-diameter vascular graft materials, such as polyethylene terephthalate (PET) and ePTFE, suffer from early occlusion and thrombosis or late intimal hyperplasia. When employed in small-diameter applications, these failure modes are more prominent and have largely prevented the use of these materials in applications where the diameter is <5 mm. Since these prostheses characteristically do not develop a luminal

endothelial layer spontaneously, there has been considerable effort expended to develop surface modifications of ePTFE in particular, in order to encourage growth of a confluent endothelial layer, subsequent to the initial report by Herring in 1978 (37).

Two different procedures have been described, endothelial seeding, and sodding. Endothelial seeding involves treating the graft with a low concentration of endothelial cells prior to implantation with *In vivo* expansion of the cells. A two-stage endothelial seeding technique involves treating the graft with a low concentration of cells, followed by extended *In vitro* culture time to expand the cells to a confluent monolayer (110,111). Conversely, endothelial sodding entails coating the graft with a high concentration of endothelial cells to quickly form a confluent endothelial layer. The ePTFE is a very hydrophobic material and few, if any, endothelial cells attach directly to the ePTFE with either direct seeding or sodding techniques. Without modification, hydrophobic ePTFE grafts show attachment of only $10 \pm 7\%$ of applied cells, with EC retention only $4 \pm 3\%$ (112,113). Surface modification of the ePTFE is necessary to encourage endothelial attachment and growth. The source of autologous human endothelial cells can range from harvested veins, adipose tissue capillaries, and endothelial progenitors (75,114,115).

Many different surface modifications of ePTFE to promote endothelialization have been studied, such as attaching cell-binding peptides (116) or fibrin glue (117). A few of these have reached human clinical trials. One such modification involved the application of fibrin glue to the luminal surface of a small-diameter ePTFE graft to which autologous endothelial cells can attach (110,118). These coated grafts were seeded with autologous endothelial cells, harvested 4–6 weeks prior to graft implantation and expanded *In vitro*. The seeded graft constructs were allowed to mature for 8–10 days before implantation as either one of 21 coronary artery bypass grafts (110) or 153 infrainguinal grafts (118). After a mean postoperative follow-up of 27.7 months, the 4 mm diameter coronary artery graft patency was 90.5%. Angiograms of all 19 patent aortocoronary bypass grafts showed smooth luminal borders without stenotic regions. Percutaneous transluminal angioscopic evaluation showed a glossy white and smooth endoluminal graft surface without any fibrin, platelet, or erythrocyte deposits. The EC seeded grafts showed an improvement in patency versus unseeded ePTFE aortocoronary bypass grafts (110). For the 6–7 mm diameter infrainguinal reconstructions, Kaplan–Meier analyses revealed a primary patency rate of 84% after 4 years and 63% after 7 years, comparable to primary patency rates for vein grafts (118).

These clinically applied ePTFE modifications were not without incident. In one trial, *In vitro* flow experiments revealed a washout of ECs between 10–15% in the first minute after the application of a physiologic pulsatile flow. Additionally, one patient had a perioperative MI caused by the immediate occlusion of the EC-seeded ePTFE graft because of poor runoff and possible culture contamination with fibroblasts (110). In the other trial, 5% of patients had EC that failed to grow in culture despite rescue serum treatment. At implantation, 5% of seeded grafts were sub-

confluent and 5% had patches devoid of EC (118). These clinical trials highlight common difficulties faced in many ePTFE endothelialization attempts. Subconfluent EC coatings allow for possible thrombus formation. Extended and technically demanding *In vitro* EC culture is required to generate enough cells for confluent seeding. An additional surgery is required for vessel harvest to isolate primary EC.

To circumvent these difficulties, a potential way to encourage autologous cellular healing *In vivo* is to alter ePTFE pore size. The ePTFE is a hydrophobic material that is microporous due to thin fibrils of material stretched between nodes of solid PTFE; the internodal distance controls the porosity of the ePTFE material. In standard grafts with an internodal distance of 30 μm , endothelialization typically proceeds slowly, if at all, due to EC migration from the anastomoses. Animal studies employing larger pore sizes have shown improved cell in-growth and tissue healing for prostheses with larger internodal distances (119–121). Some (122,123) have suggested that an internodal distance of 60 μm allows optimal cell in-growth from the adventitia while not being as macroporous as prostheses with internodal distances greater than 90 μm (119).

Neovascularization has been demonstrated in animal models with vessels penetrating the larger graft pores (124), however, similar neovascularization and endothelialization is not observed in humans (125). Pore size alone, due to the hydrophobic nature of ePTFE, may be insufficient to influence cell in-growth into ePTFE prostheses.

A significant improvement in ePTFE graft design would be the development of *In vivo* endothelialization without the need for EC preseeding, due to capture of circulating endothelial progenitor cells *In vivo* or facilitating rapid neovascularization from the adventitia, also termed transmural endothelialization. A confluent luminal monolayer of endothelial cells would serve to eliminate the primary thrombotic mechanism of vascular graft failure. In this regard, the important consideration with ePTFE is how to facilitate endothelial cell attachment. Methods to improve endothelial cell retention include fibronectin preadsorption, covalent fibronectin attachment or glow discharge modification to facilitate protein adsorption. In a study of amide–amine glow-discharge modification, endothelial cell surface adhesion was four-fold higher in treated versus untreated ePTFE grafts (113). Additionally, the endothelial population on treated grafts remained shear stable, whereas untreated grafts showed a >90% decrease in endothelial cell population after application of shear stress (113). An animal implantation model showed improved graft healing after covalent binding of fibronectin to the graft materials (126). In a different approach, Kidd et al. attempted to promote *In vivo* endothelialization in a 1 mm ID ePTFE rat abdominal aortic implant model by facilitating transmural endothelialization (127). Squamous epithelial cell lines were allowed to attach and elaborate extracellular matrix (ECM) on the abluminal surface of the ePTFE graft for 8 days prior to implantation. The cells were removed, leaving an ePTFE graft with incorporated ECM for

implantation. Upon removal at 5 week, all samples were patent with ECM-modified grafts exhibiting extensive abluminal vascularization and tissue incorporation compared to nonmodified samples. Additionally, ECM modified grafts possessed a luminal cellular lining, while nonmodified grafts were void of a cellular lining except for limited pannus in-growth. These results indicate that luminal EC coverage can be derived from in-growth of capillaries through porous grafts (127).

Tissue Engineering

The recent thrust in developing a successful small-diameter vascular graft has been the use of a tissue engineering approach, instead of starting with an established vascular graft material (128,129). Strategies have included developing degradable polymeric biomaterials (130), utilizing decellularized biological conduits (131), and fabrication of a totally tissue-engineered vessel (38). Typically, a tissue engineering approach utilizes some type of degradable scaffold in which layers of endothelial and smooth muscle cells are grown to recapitulate an arterial architecture. These approaches have unique advantages and limitations. Ideally, tissue engineered vascular constructs should be nonthrombogenic, vasoreactive, and biostable, but should also be biocompatible and not prone to excessive inflammation, immune response, or infection. The cells should also be able to regenerate and angiogenesis should be supported. One challenge for tissue engineered vascular grafts is to combine degradability of the scaffold with sufficient mechanical strength to withstand pulsatile arterial pressures without developing aneurysmal dilatation. From an operative standpoint, tissue engineered grafts should be stored in a readily implantable form, be individualized to particular implant requirements of size and length and not prone to leaking.

Scaffolds. Synthetic polymer scaffolds for tissue engineering applications have utilized bioresorbable polymers, such as polyglycolic acid, polylactic acid, and polydioxanone (34). Polylactic acid is semicrystalline with high mechanical strength, with a naturally occurring degradation product, L-lactic acid (34). In one example, Niklason et al. (132) utilized polyglycolic acid (PGA) scaffolds chemically modified with sodium hydroxide. These were seeded with bovine smooth muscle cells and allowed to incubate under conditions of pulsatile radial stress for 8 weeks. Bovine aortic endothelial cells were seeded onto the lumen of the constructs and continuous perfusion was applied for the final 3 days of culture. These vessels cultured under pulsed conditions with appropriate supplements had average rupture strengths of 2150 ± 708 mmHg (286.6 ± 94.3 kPa) after 8 weeks with 50% dry weight collagen content. These values were very similar to native vessel rupture strength [1680 ± 307 mmHg (223.9 ± 40.9 kPa)] and collagen content (45% dry weight). The engineered vessels also displayed measurable contractions in response to vasoactive substances, such as serotonin, endothelin-1, and prostaglandin F_{2a}.

Another approach is to use two or more bioresorbable polymers with different degradation rates to optimize mechanical properties. Matsumura et al. (133) fabricated two types of bioresorbable polymer for use in a tissue engineering vascular autograft for pediatric patients. One polymer system was comprised of a 50:50 copolymer of lactide and ϵ -caprolactone [P(CL/LA)] that was reinforced by a nonwoven fabric made with PGA. Endothelial cells were harvested from saphenous vein 1–2 months prior to surgery, expanded *In vitro* and seeded onto the scaffold 10 days prior to surgery. This graft construct was used in a limited clinical study in the pulmonary artery, a low pressure application, with some success. Another hybrid scaffold was similar, but the reinforcing material was changed to a woven fabric made from poly-L-lactic acid (PLLA) to increase durability and the polymer scaffolds were seeded with bone marrow cells aspirated at the time of cardiac surgery and allowed to incubate with the scaffold in culture medium for 2–4 h before implantation. This BMC seeded polymer scaffold was used in 22 patients without thrombogenic complications, stenosis, or obstruction of tissue-engineered autografts. The benefits of this type of tissue engineered vascular autograft is that it has growth potential, reduced incidence of calcifications, and no risk of rejection due to use of autologous cells, but long term *In vitro* cell culture may increase the risk of infection and changes in cell phenotype could lead to thromboembolic complications.

The degradable polymer scaffolds have also been modified to include cell-binding peptides to further encourage cell growth and incorporation, such as arginine-glycine-aspartic acid (RGD), a well-characterized amino acid peptide that is a ligand to several cell surface integrin receptors (134–136). In one study, a biotinylated PLA-polyethylene glycol copolymer was reacted with streptavidin-RGD to facilitate endothelial cell attachment (137). Controlling the density of RGD groups on the polymer surface has been shown to modulate endothelial cell migration rate and shear stability (138,139).

A different approach to synthetic materials is the use of deoxyribonucleic acid (DNA) technology to engineer synthetic proteins. These have been used with some success to incorporate both RGD functionality and elastin-based peptides to facilitate endothelial cell growth and function, while allowing cross-linking and control of mechanical properties (140). Degradation can be controlled by incorporating sequences susceptible to degradation by enzymes, such as collagenase. Incorporation of cell binding amino acid sequences into these matrices has been shown to control cell functionality (141).

Biological materials have also been used as scaffolds, including fibrinbased or collagenous matrices and decellularized materials, such as extracellular matrix, arteries, small intestinal submucosa, and peritoneum-fascia (75,142–145). Cross-linked collagen is an attractive scaffold material, as it can have significant mechanical strength, depending on cross-link density, and contains several cell-binding peptides to encourage cell incorporation. A collagen scaffold was used in the first tissue engineered prosthesis (38), but this failed due to suboptimal mechanical properties. A fibrin gel-based graft has recently

shown some success, albeit in a low pressure jugular vein model (142). Methods to improve the scaffold mechanical properties have included mechanical preconditioning and use of fabricating techniques, such as electrospinning (146,147). Matsuda and co-workers used a rodlike mandrel around which a solution of collagen and smooth muscle cells was formed (148). The mandrel was then removed and the lumen seeded with endothelial cells. This was unsuccessful due to low burst pressures, and external reinforcement was needed to impart suitable mechanical properties (149,150). A collagen micro sponge has been combined with a biodegradable polymeric scaffold to produce an implantable graft material for *In vivo* cell repopulation (151). Recently, a hyaluronan-based material (Hyaff-11) has been tested as a scaffold for endothelial cells (152).

Use of biological tissues with a preexisting extracellular matrix, such as arteries (153) or ureters (154), would be attractive, but antigenic differences between xenogenic species and humans preclude their use due to immunoreactivity and rejection. For this reason, there have been attempts to remove the cells from these scaffolds by enzymatic and detergent treatment and repopulate the decellularized scaffolds with autologous human cells (155). The xenogenic matrix still is potentially immunogenic (156), which may be alleviated by cross-linking (157) or population with autologous cells. Decellularized porcine carotid arteries treated with heparin and repopulated with canine cells have been shown to function for up to 18 weeks in a canine aortic graft model (153). Decellularized porcine small intestinal submucosa has also been studied as a tissue engineering scaffold. These constructs have been tested in animal models with some success, but intimal hyperplasia and incomplete endothelial coverage have been observed (158).

Source of Cells. In most tissue engineering constructs, the endothelial and smooth muscle cells are isolated from autologous human tissue, such as abdominal fat aspirates or veins, then cultured *In vitro* together with the construct for up to several weeks prior to implantation. There has been some recent interest in the use of bone marrow-derived endothelial progenitor cells as a novel source of endothelial cells (154). Cho et al. (159) examined tissue engineering of small diameter vascular grafts using bone marrow cells (BMCs) and decellularized arteries. Bone marrow mononuclear cells (BMMNCs) were isolated and two distinct fractions, one containing smooth muscle (SM) α -actin/SM myosin heavy-chain (SMMHC) positive cells and one containing vWF/CD31 positive cells, were cultured for 3 weeks using different culture media and supplements. Other studies have shown that when using BMCs, sufficient cells for seeding can be obtained on the day of surgery, thus obviating the need for extra vein harvesting surgery and prolonged cell culture.

SUMMARY

Vascular prostheses are vital in the treatment of many vascular diseases, from coronary artery disease and aortic

aneurysms to peripheral arterial disease. Several synthetic vascular graft materials, particularly Dacron and ePTFE, have been successful as arterial substitutes in large diameter applications. However, these devices can fail due to thrombotic occlusion, intimal hyperplasia, and infection. There is a need to develop novel vascular prostheses that function well in small diameter applications. Progress has been made on methods to facilitate the incorporation of endothelial cells into vascular grafts. The tissue engineering approach, with advances in new materials that can control graft mechanical properties plus cell attachment and proliferation, holds promise for the next generation of these crucial medical devices.

BIBLIOGRAPHY

1. Stehbens WE. General features, structure, topography and adaptation of the circulatory systems. In: Stehbens WE, Lie JT, editors. *Vascular Pathology*. New York: Chapman and Hall; 1995. p 1–18.
2. Wu KK, Thiagarajan P. Role of endothelium in thrombosis and hemostasis. *Ann Rev Med* 1996;47:315–331.
3. Gertler JP, Abbott WM. Prothrombotic and fibrinolytic function of normal and perturbed endothelium. *J Surg Res* 1992;52(1):89–95.
4. Cooley DA, DeBakey ME. Surgical consideration of intrathoracic aneurysms of the aorta and great vessels. *Ann Surg* 1952;135:660–680.
5. Voorhees AB Jr, Jaretzke AL III, Blakemore AH. The use of tubes constructed from Vinyon-‘N’ cloth in bridging arterial defects. *Ann Surg* 1952;135:332–336.
6. Sporn L, Huber P. Endothelial cell biology, in Hemostasis and Thrombosis: Basic principles and Clinical Practice. In: Colman RW, editor. Philadelphia: Lippincott, Williams, and Wilkins; 2001. p 615–623.
7. Hansson GK. Inflammation, atherosclerosis, and coronary artery disease. *N Engl J Med* 2005;352(16):1685–1695.
8. Cardiovascular Disease Statistics. Available at <http://www.americanheart.org/presenter.jhtml?identifier=4478>: Accessed 2005 April 30.
9. American Heart Association. Heart Disease and Stroke Statistics—2005 Update. Dallas: American Heart Association; 2005.
10. Open-Heart Surgery Statistics. Available at <http://www.americanheart.org/present.jhtml?identifier=4674>: Accessed 2005 April 30.
11. Benavente O, Hart RG, Sherman DG. Primary Prevention of Transient Ischemic Attack and Thromboembolic Stroke. In: Verstraete M, Fuster V, Topol EJ, editors. *Cardiovascular Thrombosis: Thrombocardiology and Thromboneurology*, 2nd ed. Philadelphia: Lippincott-Raven Publishers; 1998. p 585–595.
12. Ferro JM. Cardioembolic stroke: an update. *Lancet Neurol* 2003;2:177–188.
13. Hirsch AT. Peripheral Arterial Disease Detection, Awareness, and Treatment in Primary Care. *JAMA* 2001;286:1317–1324.
14. Dunbard RL, Mohler ER 3rd. The unsung perils of peripheral arterial disease: a malady in search of a patient. *Prev Cardiol* 2005;8(2):108–113.
15. Londrey GL, et al. Infrapopliteal bypass for severe ischemia: comparison of autogenous vein, composite and prosthetic grafts. *J Vasc Surg* 1991;13:631–636.
16. Veith FJ, et al. Six-year prospective multicenter randomized comparison of autologous saphenous vein and expanded

- polytetrafluoroethylene grafts in infrainguinal arterial reconstructions. *J Vasc Surg* 1986;3:104–114.
17. Sala F, et al. Long-term outcome of femoral above-knee popliteal artery bypass using autologous saphenous vein versus expanded polytetrafluoroethylene grafts. *Ann Vasc Surg* 2003;17:401–407.
 18. Sakalihasan N, Limet R, Defawe OD. Abdominal aortic aneurysm. *Lancet* 2005;365:1577–1589.
 19. The UK Small Aneurysm Trial Participants. Mortality results for randomized controlled trial of early elective surgery or ultrasonographic surveillance for small abdominal aortic aneurysms. The UK Small Aneurysm Trial Participants. *Lancet* 1998;352:1649–1655.
 20. Sakalihasan N, et al. Modification of the extracellular matrix of aneurysmal abdominal aortas as a function of their size. *Eur J Vasc Surg* 1993;7:633–637.
 21. Dobrin PB, Mrkvicka R. Failure of elastin or collagen as possible critical connective tissue alterations underlying aneurysmal dilatation. *Cardiovasc Surg* 1994;2:484–488.
 22. Shah PK. Inflammation, metalloproteinases, and increased proteolysis: an emerging pathophysiological paradigm in aortic aneurysm. *Circulation* 1997;96:2115–2117.
 23. Kaminen R, Heuser RR. Abdominal aortic aneurysm: A review of endoluminal treatment. *J Intervent Cardiol* 2004;17:437–445.
 24. Moore WS, Kashyap VS, Vescera CL, Quinones-Baldrich WJ. Abdominal aortic aneurysm: A 6-year comparison of endovascular versus transabdominal repair. *Ann Surg* 1999; 230:298–306.
 25. Parodi JC, Palmaz JC, Barone HD. Transfemoral intraluminal graft implantation for abdominal aortic aneurysms. *Ann Vasc Surg* 1991;5:491–499.
 26. Perry MO. Arterial Injuries. In: Bell PRF, Jamieson CW, Ruckley CV, editors. *Surgical Management of Vascular Disease*. London: W. B. Saunders; 1992. p 905–925.
 27. Lumley JSP. Microvascular Surgery. In: Bell RF, Jamieson CW, Ruckley CV, editors. *Surgical Management of Vascular Disease*. London: W.B. Saunders; 1992. p 941–954.
 28. Millis JM, Brown SL, Busuttill RW. Thoracic and Abdominal Aneurysms. In: Bell PRF, Jamieson CW, Ruckley CV, editors. *Surgical Management of Vascular Disease*. London: W. B. Saunders; 1992. p 797–828.
 29. Keen KW. *Surgery: Its Principles and Practice*. Philadelphia: W.B. Saunders; 1921.
 30. Rea CE. Surgical treatment of aneurysm of the abdominal aorta. *Minn Med* 1948;31:153.
 31. Dubost C, Allary M, Oeconomos N. Resection of an aneurysm of the abdominal aorta: Reestablishment of the continuity by a preserved human arterial graft, with result after five months. *AMA Arch Surg* 1952;64:405–408.
 32. Cooley DA. Early development of Surgical treatment for aortic aneurysms: Personal recollections. *Texas Heart Institute J* 2001;28:197–199.
 33. Hess F. History of (micro)vascular surgery and the development of small-caliber blood vessel prostheses. *Microsurgery* 1985;6:59–69.
 34. Xue L, Greisler HP. Biomaterials in the development and future of vascular grafts. *J Vasc Surg* 2003;37:472–480.
 35. Matsumoto H, Hasegawa T, Fuse K. A new vascular prosthesis for a small caliber artery. *Surgery* 1973;74:518–523.
 36. Dardik H, et al. Glutaraldehyde-tanned human umbilical vein grafts. In: Stanley JC, et al. editors. *Biologic and Synthetic Vascular Prostheses*. New York: Grune & Stratton; 1982. p 433.
 37. Herring M, Gardner A, Glover J. A single-staged technique for seeding vascular grafts with autogenous endothelium. *Surgery* 1978;84:498–504.
 38. Weinberg CB, Bell E. A blood vessel model constructed from collagen and cultured vascular cells. *Science* 1986;231:397–400.
 39. Kunlin JL. Le traitement de l'arterite oblitérante par la greffe veineuse. *Arch Mal Coeur* 1949;42:371.
 40. Yun KL, et al. Randomized trial of endoscopic versus open vein harvest for coronary artery bypass grafting: six-month patency rates. *J Thorac Cardiovasc Surg* 2005;129:496–503.
 41. Lytle BW. Prolonging patency—choosing coronary bypass grafts. *N Eng J Med* 2004;351:2262–2264.
 42. Desai ND, Cohen EA, Naylor CD, Fremes SE, the Radial Artery Patency Study Investigators. A randomized comparison of radial-artery and saphenous-vein coronary bypass grafts. *N Engl J Med* 2004;351:2302–2309.
 43. Chervu A, Morre WS. *Vascular Grafts and Sutures*. In: Bell RF, Jamieson CW, Ruckley CV, editors. *Surgical Management of Vascular Disease*. London: WB. Saunders; 1992. p 367–389.
 44. Boyd JH, Stevens R, Havey A, Silver D. Intimal integrity and fibrinolytic potential of reversed and in situ vein grafts. *J Vasc Surg* 1987;5:614–621.
 45. Lawson JA, Tangelder MJ, Algra A, Eikelboom BC. The myth of the in situ graft: superiority in infrainguinal bypass surgery? *Eur J Vasc Endovasc Surg* 1999;18:149–157.
 46. Snyder RW, Botzko KM. Woven knitted and externally supported Dacron vascular prostheses. Stanley JC, et al. editors. *Biologic and Synthetic Vascular Prostheses*. New York: Grune & Stratton; 1982. p 485.
 47. Cziperle DJ, et al. Albumin impregnated vascular grafts: albumin resorption and tissue reactions. *J Cardiovasc Surg* 1992;33:407–414.
 48. Hall CW, et al. Velour fabrics applied to medicine. *J Biomed Mater Res* 1967;1:179–196.
 49. Lindenauer SM, Lavanway JM, Fry WJ. Development of a velour vascular prosthesis. *Curr Top Surg Res* 1970;2: 491.
 50. Sauvage LR, et al. An external velour surface for porous arterial prostheses. *Surgery* 1971;70:940–953.
 51. Norman PE, Semmens JB, Lawrence-Brown MM. Long-term relative survival following surgery for abdominal aortic aneurysm: a review. *Cardiovasc Surg* 2001;9:219–224.
 52. Friedman SG, et al. A prospective randomized comparison of Dacron and polytetrafluoroethylene aortic bifurcation grafts. *Surgery* 1995;117:7–10.
 53. Quarmby JW, et al. Prospective randomized trial of woven versus collagen-impregnated knitted prosthetic Dacron grafts in aortoiliac surgery. *Br J Surg* 1998;85: 775–777.
 54. Devine C, McCollum C. for the North West Femoro-Popliteal Trial Participants. Heparin-bonded Dacron or polytetrafluoroethylene for femoropopliteal bypass: Five-year results of a prospective randomized multicenter clinical trial. *J Vasc Surg* 2004;40:924–931.
 55. Davidovic L, et al. Aortobifemoral grafting: factors influencing long-term results. *Vascular* 2004;12:171–178.
 56. Gupta SK, Ascer E, Veith FJ. Expanded polytetrafluoroethylene arterial grafts: An eight-year experience. In: Sawyer PN, editor. *Modern Vascular Grafts*. New York: McGraw-Hill; 1987. p 181.
 57. Berglund J, Björck M, Elfstrom J. for the SWEDVASC Femoro-popliteal study group. Long-term results of above knee femoro-popliteal bypass depend on indication for surgery and graft-material. *Eur J Vasc Endovasc Surg* 2005;29:412–418.
 58. Taylor LM, Edwards JM, Porter JM. Present status of reversed vein bypass grafting: five-year results of a modern series. *J Vasc Surg* 1990;10:220–225.

59. Donaldson MC, Mannick JA, Whittemore AD. Femoro-distal bypass with in situ greater saphenous vein. *Ann Surg* 1991; 213:457–465.
60. Sala F, et al. Long-term outcome of femoral above-knee popliteal artery bypass using autologous saphenous vein versus expanded polytetrafluoroethylene grafts. *Ann Vasc Surg* 2003;17:401–407.
61. Wood RFM. Vascular access in dialysis. In: Bell RF, Jamieson CW, Ruckley CV, editors. *Surgical Management of Vascular Disease*. London: W.B. Saunders; 1992. p 1049–1067.
62. Francis DMA. More vein, less plastic. *Nephrology* 2005;10: 10–14.
63. Bosman PJ, et al. A comparison between PTFE and denatured homoogous vein grafts for haemodialysis access: A prospective randomized multi-center trial. *Eur J Vasc Endovasc Surg* 1998;16:126–132.
64. Crowther MA, et al. Low-intensity Warfarin is ineffective for the prevention of PTFE graft failure in patients on hemodialysis: A randomized controlled trial. *J Am Soc Nephrol* 2002;13:2331–2337.
65. Towne JB. Endovascular treatment of abdominal aortic aneurysms. *Am J Surg* 2005;189:140–149.
66. Laheij RJ, et al. Need for secondary interventions after endovascular repair of abdominal aortic aneurysms: intermediate-term follow-up results of a European collaborative registry (EUROSTAR). *Br J Surg* 2000;87:1666–1673.
67. Hallett JW Jr, et al. Graft-related complications after abdominal aortic aneurysm repair: reassurance from a 36-year population-based experience. *J Vasc Surg* 1997;25:277–284.
68. Jacobs TS, et al. Mechanical failure of prosthetic human implants: A 10 year experience with aortic stent graft devices. *J Vasc Surg* 2003;37:16–26.
69. Anderson JM, Kottke-Marchant K. Platelet interaction with biomaterials and artificial devices. *CRC Crit Rev Biocompat* 1985;1:111–204.
70. Greisler HP, et al. Biointeractive polymers and tissue engineered blood vessels. *Biomaterials* 1996;17:329–336.
71. Kottke-Marchant K, Anderson JM, Rabinovitch A. The platelet reactivity of vascular graft prostheses: An In vitro model to test the effect of preclotting. *Biomaterials* 1986;7:441–448.
72. Lin Y, Weisdorf DJ, Solovey A, Hebbel RP. Origins of circulating endothelial cells and endothelial outgrowth from blood. *J Clin Invest* 2000;105:71–77.
73. Davids L, Dower T, Zilla P. The lack of healing in conventional vascular grafts. In: Zilla P, Greisler HP, editors. *Tissue engineering of vascular prosthetic grafts*. Austin, TX: R.G. Landes Co.; 1999. p 3–44.
74. Conte MS, et al. Genetic interventions for vein bypass graft disease: a review. *J Vasc Surg* 2002;36:1040–1052.
75. Rashid ST, et al. Engineering of bypass conduits to improve patency. *Cell Prolif* 2004;37:351–366.
76. Kidane AG, et al. Anticoagulant and antiplatelet agents. Their clinical and device application(s) together with usages to engineer surfaces. *Biomacromolecules* 2004;5:798–813.
77. Collins TC, Soucek J, Beyth RJ. Benefits of antithrombotic therapy after infrainguinal bypass grafting: a meta-analysis. *Am J Med* 2004;117:93–99.
78. Dorffler-Melly J, et al. Antiplatelet agents for preventing thrombosis after peripheral arterial bypass surgery. *Cochrane Database Syst Rev* 2003;3:CD000535.
79. LoGerfo FW, et al. Anastomotic hyperplasia: A mechanism of failure in Dacron arterial grafts. *Ann Surg* 1983;197:479–483.
80. Rotmans JI, et al. Rapid, arteriovenous graft failure due to intimal hyperplasia: A porcine, bilateral, carotid arteriovenous graft model. *J Surg Res* 2003;113:161–171.
81. Clowes AW, Gown AM, Hanson SR, Reidy MA. Mechanisms of arterial graft failure. Role of cellular proliferation in early healing of PTFE prostheses. *Am J Pathol* 1985;118:43–54.
82. Clowes AW, Kirkman TR, Clowes MM. Mechanisms of arterial graft failure. II. Chronic endothelial and smooth muscle cell proliferation in healing polytetrafluoroethylene prostheses. *J Vasc Surg* 1986;3:87–884.
83. Haruguchi H, Teraoka S. Intimal hyperplasia and hemodynamic factors in arterial bypass and arteriovenous grafts: a review. *J Artif Organs* 2003;6:227–235.
84. Purcell C, Tennant M, McGeachie J. Neo-intimal hyperplasia in vascular grafts and its implications for autologous arterial grafting. *Ann R Coll Surg Engl* 1997;79:164–168.
85. Tiwari A, Salacinski H, Seifalian AM, Hamilton G. New prostheses for use in bypass grafts with special emphasis on polyurethanes. *Cardiovasc Surg* 2002;10:191–197.
86. Zubilewicz T, et al. Injury in vascular surgery — the intimal hyperplastic response. *Med Sci Monit* 2001;7:316–324.
87. Sapienza P, et al. Release of PDGF-BB and bFGF by human endothelial cells seeded on expanded polytetrafluoroethylene vascular grafts. *J Surg Res* 1998;74:24–29.
88. Mattana J, Effiong C, Kapasi A, Singhal PC. Leukocytepolytetrafluoroethylene interaction enhances proliferation of vascular smooth muscle cells via tumor necrosis factor-alpha secretion. *Kidney Int* 1997;52:1478–1485.
89. Randone B, et al. Suppression of smooth muscle cell proliferation after experimental PTFE arterial grafting: a role for polyclonal anti-basic fibroblast growth factor (bFGF) antibody. *Eur J Vasc Endovasc Surg* 1998;16:401–407.
90. Yu H, et al. Neointimal hyperplasia on a cell-seeded polytetrafluoroethylene graft is promoted by transfer of tissue plasminogen activator gene and inhibited by transfer of nitric oxide synthase gene. *J Vasc Surg* 2005;41:122–129.
91. Goldstone J, Moore WS. Infection in vascular prosthesis: Clinical manifestations and Surg management. *Am J Surg* 1974;128:225–233.
92. Shell DH, et al. Comparison of small-intestinal submucosa and expanded polytetrafluoroethylene as a vascular conduit in the presence of gram-positive contamination. *Ann Surg* 2005;241:995–1001.
93. Stadler P, Bilohlavek O, Spacek M, Michalek P. Diagnosis of vascular prosthesis infection with FDG-PET/CT. *J Vasc Surg* 2004;40:1246–1247.
94. Hart JP, et al. Operative strategies in aortic graft infections: is complete graft excision always necessary?. *Ann Vasc Surg* 2005;19:154–160.
95. Aldenhoff YBJ, et al. Performance of a polyurethane vascular prosthesis carrying a dipyridamole (Persantin) coating on its luminal surface. *J Biomed Mater Res* 2001;54:224–233.
96. Lin P, et al. Small-caliber heparin-coated ePTFE grafts reduce platelet deposition and neointimal hyperplasia in a baboon model. *J Vasc Surg* 2004;39(6):1322–1328.
97. Lambert AW, et al. Experience with heparin-bonded collagen-coated grafts for infrainguinal bypass. *Cardiovasc Surg* 1999;7:491–494.
98. Walpoth BH, et al. Improvement of patency rate in heparin-coated small synthetic vascular grafts. *Circulation* 1998;98:II 319–II 323.
99. Akers DL, Du YH, Kempczinski RF. The effect of carbon coating and porosity on early patency of expanded polytetrafluoroethylene grafts: An experimental study. *J Vasc Surg* 1993;18:10–15.
100. Greisler HP, et al. Endothelialization of expanded PTFE grafts by heparin binding growth factor-type1 pretreatment. *Surgery* 1992;112:244–255.

101. Gray JL, et al. FGF-1 affixation stimulates ePTFE endothelialization without intimal hyperplasia. *J Surg Res* 1994; 57:596–612.
102. Wang C, Zhang Q, Uchida S, Kodama M. A new vascular prosthesis coated with polyamino-acid urethane copolymer (PAU) to enhance endothelialization. *J Biomed Mater Res* 2002;62:315–322.
103. Zhang Z, et al. Vascugraft polyurethane arterial prosthesis as femoro-popliteal and femoro-peroneal bypasses in humans: pathological, structural and chemical analysis of four excised grafts. *Biomaterials* 1997;18:113–124.
104. Glickman MH, et al. Multicenter evaluation of a polyurethane vascular access graft as compared with the expanded polytetrafluoroethylene vascular access graft in hemodialysis applications. *J Vasc Surg* 2001;34:465–472.
105. Jeschke MG, Hermanutz V, Wolf SE, Koveker GB. Polyurethane vascular prostheses decreases neointimal formation compared with expanded polytetrafluoroethylene. *J Vasc Surg* 1999;29:168–176.
106. Wilson GJ, et al. The composite Corethane/Dacron vascular prosthesis. Canine *In vivo* evaluation of 4 mm diameter grafts with 1 year follow-up. *ASAIO Trans* 1991;37:M475–M476.
107. Izhar U, et al. Novel synthetic selectively degradable vascular prostheses: A preliminary implantation study. *J Surg Res* 2001;95:152–162.
108. Jun H-W, Taite LJ, West JL. Nitric oxide-producing polyurethanes. *Biomacromolecules*, 2005;6:838–844.
109. Flester PS, et al. Nitric oxide-releasing biopolymers inhibit thrombus formation in a sheep model of arteriovenous bridge grafts. *J Vasc Surg* 2004;40:803–833.
110. Laube H, Duwe J, Rutsch W, Konertz W. Clinical experience with autologous endothelial cell-seeded polytetrafluoroethylene coronary artery bypass grafts. *J Thorac Cardiovasc Surg* 2000;120(1):134–141.
111. Deutsch M, et al. Clinical autologous *In vitro* endothelialization of infrainguinal ePTFE grafts in 100 patients: a 9 year experience. *Surgery* 1999;126:847–855.
112. Kent K, Oshima A, Whittemore A. Optimal seeding conditions for human endothelial cells. *Ann Vasc Surg* 1992;6(3): 258–264.
113. Tseng DY, Edelman ER. Effects of amide and amine plasma-treated ePTFE vascular grafts on endothelial cell lining in an artificial circulatory system. *J Biomed Mater Res* 1998;42: 188–198.
114. Sharp WV, Schmidt SP, Meerbaum SO, Pippert TR. Derivation of human microvascular endothelial cells for prosthetic vascular graft seeding. *Ann Vasc Surg* 1989;3:104–107.
115. Boyer M, et al. Isolation of endothelial cells and their progenitor cells from human peripheral blood. *J Vasc Surg* 2000;31:181–189.
116. Chan BP, et al. *In vivo* performance of dual ligand augmented endothelialized expanded polytetrafluoroethylene vascular grafts. *J Biomed Mater Res Part B: Appl Biomater* 2005; 72B:52–63.
117. Kumar TRS, Krishnan LK. A stable matrix for generation of tissueengineered nothrombogenic vascular grafts. *Tissue Eng* 2002;8:763–770.
118. Meinhart J, et al. Clinical autologous *In vitro* endothelialization of 153 infrainguinal ePTFE grafts. *Ann Thorac Surg* 2001;71(5 Suppl):S327–331.
119. Hazama K, et al. Relationship between fibril length and tissue ingrowth in the healing of expanded polytetrafluoroethylene grafts. *Surg Today* 2004;34:685–689.
120. Golden MA, et al. Healing of polytetrafluoroethylene arterial grafts is influenced by graft porosity. *J Vasc Surg* 1990; 11:838–845.
121. Hirabayashi K, et al. Influence of fibril length upon ePTFE graft healing and host modification of the implant. *J Biomed Mater Res* 1992;26:1433–1447.
122. Kuzuya A, et al. Healing of implanted expanded polytetrafluoroethylene vascular access grafts with different inter-nodal distances: A histologic study in dogs. *Eur J Vasc Endovasc Surg* 2004;28:404–409.
123. Contreras MA, Quist WC, LoGerfo FW. Effect of porosity on small diameter vascular graft healing. *Microsurgery* 2000; 20:15–21.
124. Clowes AW, Kirkman TR, Reiday MA. Mechanisms of graft healing. Rapid transmural capillary ingrowth provides a source of intimal endothelium and smooth muscle in porous PTFE prosthesis. *Am J Pathol* 1986;123:221–230.
125. Sauvage LR, et al. Interspecies healing of porous arterial prosthesis. Observations, 1960 to 1974. *Arch Surg* 1974;109: 6989–6705.
126. Shimada T, et al. Improved healing of small-caliber, long-fibril expanded polytetrafluoroethylene vascular grafts by covalent bonding of fibronectin. *Surg Today* 2004;34:1025–1030.
127. Kidd KR, Patula VB, Williams SK. Accelerated endothelialization of interpositional 1-mm vascular grafts. *J Surg Research* 2003;113:234–242.
128. Matsuda T. Recent progress of vascular graft engineering in Japan. *Artif Organs* 2004;28:64–71.
129. Nerem RM, Seliktar D. Vascular tissue engineering. *Ann Rev Biomed Eng* 2001;3:225–243.
130. Pachence JM, Kohn J. Biodegradable polymers. In: Lanza RP, Langer R, Vacanti J, editors. *Principles of tissue engineering*. 2nd ed. San Diego: Academic Press; 2000. p 263–277.
131. Wilson GJ, et al. Acellular matrix: a biomaterial approach for coronary artery bypass and heart valve replacement. *Ann Thorac Surg* 1995;60(2 Suppl):S353–S358.
132. Niklason LE, et al. Functional arteries grown *In vitro*. *Science* 1999;284:489–493.
133. Matsumura G, et al. Successful application of tissue engineered vascular autografts; clinical experience. *Biomaterials* 2003;24:2303–2308.
134. Drumheller PD, Hubbell JA. Polymer networks with grafted celladhesion peptides for highly biospecific cell adhesive substrates. *Anal Biochem* 1994;222:380–388.
135. Hersel U, Dahmen C, Kessler H. RGD modified polymers: biomaterials for stimulated cell adhesion and beyond. *Biomaterials* 2003;24:4385–4415.
136. Shin H, Jo S, Mikos AG. Biomimetic materials for tissue engineering. *Biomaterials* 2003;24:4353–4364.
137. Patel N, et al. Spatially controlled cell engineering on biodegradable polymer surfaces. *FASEB J* 1998;12:1447–1454.
138. Sagnella S, et al. Human Microvascular Endothelial Cell Growth and Migration on Biomimetic Surfactant Polymers. *Biomaterials* 2004;25:1249–1259.
139. Sagnella S, Kligman F, Marchant RE, Kottke-Marchant K. Biomimetic Surfactant Polymers Designed for Shear Stable Endothelialization on Biomaterials. *J Biomed Mater Res* 2003;67A(3):689–701.
140. Urry DW, Pattanaik A. Elastic protein-based material in tissue reconstruction. *Ann NY Acad Sci* 1997;831:32–46.
141. Richman GP, Tirrell DA, Asthagiri AR. Quantitatively distinct requirements for signaling-competent cell spreading on engineered versus natural adhesion ligands. *J Controlled Rel* 2005;101:3–12.

142. Swartz DD, Russell JA, Andreadis ST. Engineering of fibrin-based functional and implantable small-diameter blood vessels. *Am J Physiol Heart Circ Physiol* 2005;288:H1451–H1460.
143. Bader A, et al. Engineering of human vascular aortic tissue based on a xenogeneic starter matrix. *Transplantation* 2000; 70:7–14.
144. Clarke DR, et al. Transformation of nonvascular acellular tissue matrices into durable vascular conduits. *Ann Thorac Surg* 2001;71:S433–S436.
145. Sarac TP. *In vivo* and mechanical properties of peritoneum/fascia as a novel arterial substitute. *J Vasc Surg* 2005; 41:490–497.
146. Seliktar D, Black RA, Vitro RP, Nerem RM. Dynamic mechanical conditioning of collagen-gel blood vessel constructs induces remodeling in vitro. *Ann Biomed Eng* 2000;28:351–362.
147. Matthews JA, et al. Smooth muscle cell migration in electropun poly(lactic acid) and collagen/elastin. *Cardiovasc Pathol* 2002;11:13–18.
148. Hirai J, Matsuda T. Venous reconstruction using hybrid vascular tissue composed of vascular cells and collagen: tissue regeneration process. *Cell Transplant* 1996;5:93–105.
149. He H, Matsuda T. Newly designed compliant hierarchical hybrid vascular graft wrapped with microprocessed elastomeric film — II: Morphogenesis and compliance change upon implantation. *Cell Transplant* 2002;11:75–87.
150. He H, Shirota T, Yasui H, Matsuda T. Canine endothelial progenitor cell-lined hybrid vascular graft with non-thrombogenic potential. *J Thorac Cardiovasc Surg* 2003;126:455–464.
151. Iwai S, et al. Biodegradable polymer with collagen micro-sponge serves as a new bioengineered cardiovascular prosthesis. *J Thorac Cardiovasc Surg* 2004;128:472–479.
152. Turner NJ, Kieley CM, Walker MG, Canfield AE. A novel hyaluronan-based biomaterial (Hyaff-11) as a scaffold for endothelial cells in tissue engineered vascular grafts. *Biomaterials* 2004;25:5955–5964.
153. Tamura N. New acellular vascular prosthesis as a scaffold for host tissue regeneration. *Artif Organs* 2003;26:783–792.
154. Shirota T, He H, Yasui H, Matsuda T. Human endothelial progenitor cell-seeded hybrid graft: Proliferative and antithrombogenic potentials *In vitro* and fabrication processing. *Tissue Eng* 2003;9:127–136.
155. Wilson GJ, et al. Acellular matrix: a biomaterials approach for coronary artery bypass and heart valve replacement. *Ann Thorac Surg* 1995;60(2 Suppl):S353–S358.
156. Allaire E, et al. The immunogenicity of the ECM in arterial xenografts. *Surgery* 1997;122:73–81.
157. Courtman DW, Errett BF, Wilson GJ. The role of cross-linking in modification of the immune response elicited against xenogenic vascular acellular matrices. *J Biomed Mater Res* 2001;55:576–586.
158. Nemcova S, et al. Evaluation of a xenogeneic acellular collagen matrix as a small-diameter vascular graft in dogs — preliminary observations. *J Invest Surg* 2001;14:321–330.
159. Cho S-W, et al. Small-diameter blood vessels engineered with bone marrow-derived cells. *Ann Surg* 2005;241:506–515.

See also **BIOCOMPATIBILITY OF MATERIALS; BIOMATERIALS, SURFACE PROPERTIES OF; POROUS MATERIALS FOR BIOLOGICAL APPLICATIONS.**

VASCULAR MEASUREMENTS. See **PERIPHERAL VASCULAR NONINVASIVE MEASUREMENTS.**

VENOUS SHUNT. See **VASCULAR GRAFT PROSTHESIS.**

VENTILATION, HIGH-FREQUENCY. See **HIGH FREQUENCY VENTILATION.**

VENTILATORS, ACUTE MEDICAL CARE

PAUL C. TAMUL
KENNETH SCOPE
LEONARD CRAIG
Feinberg School of Medicine of
Northwestern University
Chicago, Illinois

VENTILATOR USE

Introduction

A ventilator, like most critical care devices, is simply a machine. However, it is not a “simple machine.” Over the past five decades, the means of delivering gases to patients have become more complex. This has led to an evolution of more than 30 critical care ventilators being used in the United States. The numerous modes of ventilation and our unfortunate attempt to reach a consensus on the terminology used when referring to mechanical ventilation presents a challenge to even the most skilled clinician. Mushin’s description of ventilators in his archetypal text (1) is appreciated by many and addresses the classification scheme that can be adopted to the less-sophisticated mechanical ventilator. An innumerable count of authors has used this classification method when writing about mechanical ventilator support (2–4). Today’s mechanical ventilators are designed to provide life support using an array of computerized logic systems, software utilities, and even artificial intelligence models to deliver gas to patients. This evolution of ventilators has led to a modernistic classification system (5) to describe today’s microprocessor-controlled mechanical ventilator. A classification system that is specific and well defined is intended to enhance communications between clinicians (6). However, some respected organizations (7) and clinicians (8) have been reluctant to accept this new system for understanding mechanical ventilation despite its acceptance by leading members in the pulmonary and critical care medicine community (9–11).

Basic Concepts

A mechanical ventilator is a system designed to alter, transmit, and direct applied energy in a predetermined manner to perform useful work (12). This work was generally accomplished with some pneumatic component in earlier (first-generation) ventilators. Current systems employ electronic circuitry that is microprocessor controlled.

The energy we put into ventilators is in the form of electricity or compressed gas. Mechanically speaking, a

ventilator will take the power input and convert or transmit that energy using the control scheme circuit to arrive at a desired output. Put more plainly, that energy is transmitted or transfigured by the ventilator in a predetermined manner to provide partial or full support to the patient's respiratory muscles.

Input Power

The input power or power source is the energy needed to perform the work required to ventilate the patient.

Generally, two forms of input energy are used to operate mechanical ventilators:

Electric

AC (alternating current)

DC (direct current)

Pneumatic. Many critical care ventilators in use today incorporate electric and pneumatic power sources to function properly because they employ advanced mechanical systems in combination with microprocessor control.

Electric

Critical care ventilators for use in the United States use common electrical voltage (110–115 V) to power them. Some manufacturers have adapted their machines with rechargeable batteries as an alternative in case of transport from one location to another or in case a power failure erupts. Some disadvantages to the type of batteries (lead-acid) used is that they supply limited power, lasting approximately 1 h and generally require a 12–24 h recharging time.

Pneumatic

The availability of compressed gases of oxygen and air in many hospital intensive care units (ICUs) makes it an ideal energy source (energy = pressure × volume) to power today's mechanical ventilator. Although the standard regulated source gas pressure is 50 pounds per square inch (psi), periodic fluctuations in gas pressure may occur (13). Ordinarily, ventilators have internal regulators that work at a lesser pressure (30–45 psi) than the source pressure. This will inherently maintain normal function despite inconsistent source gas pressure. Aside from unpredictable gas pressures, a pneumatically powered ventilator is useful in environments using magnetic resonance imaging (MRI) or when the need for transport arises. To date, only a few machines are capable of functioning on pneumatics only. Most ICU ventilators that are pneumatically powered still require electricity to support their control functions.

Power Conversion and Transmission

Power conversion and transmission refer to the mechanism(s) used to provide gas delivery. Sometimes this is referred to as the drive mechanism. This is only *partially*

true. The drive mechanism generates the actual force (force = pressure/area) to deliver gas under pressure. This is usually done with a compressor (internal or external) and/or motor linkage. Power transmission and conversion systems also consist of an output control mechanism. This is usually in the form of one or a series of valves that is used to regulate gas flow to the patient. In more complex machinery, a computerized, closed-loop feedback system is established or provides consistent gas delivery in the event of system disturbances. More recent developments of power transmission and conversion mechanisms are described here. A more detailed list can be found elsewhere (14,15).

Drive Mechanisms. This section will investigate the various drive mechanisms required to accomplish lung inflation. The post-polio epidemic introduced an upsurge in the development of many types of drive mechanisms. Although a pressure gradient (required for lung inflation) can be accomplished by one method, many forms of drive mechanisms were developed and used. Whereas each drive mechanism had a unique way of generating a particular pressure and flow pattern during a positive pressure breath, the ideal mechanism should mimic the physiologic breath and restrain adverse effects. Unfortunately, that is still not accomplished today.

There are two categories of drive mechanisms: (1) direct application of compressed gas via a pressure reducing valve and (2) indirect application via an electric motor or compressor.

Compressed Gas and Pressure Reducing Valve. When compressed gas is used for the drive mechanism, its force is adjusted via a pressure-reducing valve. A pressure-reducing valve (PRV) reduces high input pressure to a lower constant output pressure. This pressure may be set as high as 50 psi or be adjusted to a few cmH₂O. The functioning principle of an adjustable PRV can be described by

$$P_L = \frac{(F_{S1} - F_{S2}) - (P_H \times a)}{A}$$

where

P_L = Low or reduced pressure (generated pressure)

F_{S1} = Force of the large spring (adjusting spring)

F_{S2} = Force of the small spring (sealing spring)

P_H = High input pressure (source pressure)

a = Area of the small seat

A = Area of the large diaphragm

After gas is reduced from the PRV, it is delivered to the patient by one of the following approaches:

1. Directly inflate the lung, as the Puritan-Bennett 7200.
2. Prime a spring weighted bellows as in the Servo 900C.

Electric Motor and Compressor. The drive system in electrically powered ventilators consists of an electrical motor that drives a compressor. The compressor may use either a rotating crank and piston, rotary vane, or linear drive motor. Of particular interest is the type of motor and linkage system employed, because this determines the waveform the ventilator will produce.

Electric Motor/Rotating Vane

Referred to as nonlinear pistons (16), eccentric wheel (17), and rotary pistons (18), these units have a constant speed-rotating wheel with a piston rod connecting a piston to the outer edge of the wheel. As the wheel turns, the piston's actual movement is not constant. Consequently, the pressure and flow developed by this mechanism varies with the motion of the piston.

Electric Motor/Rotating Vane

A rotary vane compressor consists of an electric motor with vanes (blades) attached to the motor's shaft, which rapidly spin inside a cylinder. This action generates a constant level of pressure to serve as a drive mechanism. A series of electrically controlled switches (solenoids) serves to assure that the patient is connected to the breathing circuit only during inspiration. During exhalation, the pressure generated by the blower is vented to the atmosphere.

Electric Motor/Linear Piston

Linear piston drive systems may be a rack-and-pinion gear system, a high-tension spring driven system, or a high-pressure pneumatic drive system. This system may be constructed as a single-circuit or double-circuit scheme. In a single circuit, gas is sent directly to the patient from the device's power circuit. The piston's movement is constant throughout its stroke. Therefore, the flow of air being delivered to the patient is also constant, resulting in a square flow wave and linearly increasing pressure curve. A double-circuit design is constructed so that gases from a power circuit are used only to compress a bag-type apparatus (bellows) containing separate gases, which are sent to the patient through a patient circuit. This results in a progressively increasing flow pattern and initially an accelerating pressure pattern. Once the pressures start to equilibrate, a drop in circuit pressure occurs producing a notch in the pressure curve.

Output Control Systems. An output control system is designed to control the flow of gas to a patient. Most ventilators used in intensive care units employ a series of valves to accomplish this. In earlier models (Puritan Bennett MA-1), it was a simple ON/OFF exhalation valve. Current technology enables a valved component to be controlled with microcomputer intelligence. This allows for precise control over gas flow delivery that was not feasible with the compressor-bag delivery system used in earlier generations of mechanical ventilators. For example, the Puritan Bennett 7200 ventilators can precisely manage the movement of its output control system allowing for 3095 movements within 0.003 of an inch (19). The most common valves are discussed below.

Pneumatic Poppet Valve. Sometimes referred to as electromagnetic valves, Plunger valves, and Poppet valves (20,21). This type of device uses a magnetic force to control gas pressure in an ON/OFF fashion. A given valve regulates flow by opening or closing a particular opening. By digitally controlling the open/close sequence of the valve, flow output can be varied.

Proportional Solenoid Valves. Proportional solenoid valves can control flow incrementally instead of turning flow "ON and OFF." This allows for shaping of the inspiratory gas flow being delivered to the patient. Most current generation ICU ventilators use some type of proportional solenoid technology.

Control Scheme

Mechanical ventilators coordinate the delivery of gases based on four variables: flow, volume, time, and pressure. Therefore, the control variable becomes the variable that the ventilator manipulates to create inspiration. A single breath does not reflect significant changes in compliances and resistance. However, flow, volume, time, and pressure can shift and are regulated by the ventilator. In general, if flow, volume, time, or pressure are individually preprogrammed, the others must respond. The control variable will remain constant in the presence of changes in impedance (resistance and compliance). For example, a ventilator is a volume controller if the volume is measured and used to control the volume signal. Many of today's volume ventilators are truly *flow controllers*. They derive a volume measurement from a flow signal. The signal is measured and calculated and displayed as a volume unit of measure on the ventilator display window.

Now that we have discussed *what* is controlled, let us examine *how* it is controlled. In the open-loop system, the user selects the variable (volume, pressure, and flow), for input and observes the output variable (volume, pressure, and flow), which mirrors the actual valve. The operator handles any needed adjustments. The closed-loop system analyzes the data using a feedback signal and automatically modifies the input in an attempt to match the reference value.

Phase Variables. Control of the breath requires the ventilator to perform four phases during a breathing cycle (22):

1. End of expiratory phase (beginning of inspiration)
2. Inspiration
3. End of inspiration (beginning of expiratory phase)
4. Expiration

During these phases, the mechanical ventilator assumes all or part of the work of breathing. During each stage, a certain variable is used to start (trigger), sustain (limit), and end (cycle) the breath.

Trigger. Inspiration is initiated when one of these variables reaches a preset value. This describes

the “trigger” phase of the ventilator. With patient triggering, the ventilator senses the patient effort to initiate gas delivery. The breath is flow, volume, or pressure triggered if the patient initiates the breath. Most modern adult ventilators have a frequency, breathing rate, or respiratory rate control knob that serves as the time trigger. Manual trigger is available on most ventilator control panels as well.

Limit Variable. A limit variable is a parameter that can reach a preset level before inspiration ends, but it does not terminate the inspiration phase (23). This is often confused with a cycle variable. A limit variable does not terminate inspiration; it sets the upper boundaries for pressure, volume, or flow. An example present in many critical care ventilators is the ability for an operator to perform an “inflation hold” of plateau maneuver. The ventilator is allowed to deliver a preset volume, but inspiration does not end until a preset time interval has elapsed. The volume is limited (held and maintained) until the preset pause time is reached.

Cycle. Inspiration will end when either time, pressure, flow, or volume reaches its preselected value. Unlike the “limit” variable, the inspiration phase will end (cycle off).

Baseline Variable. The variable controlled during the expiratory phase is the baseline variable (24). Although flow or volume can serve as the baseline variable, pressure is the most commonly used. Expiratory pressure can be set to atmospheric—Zero. This is called ZEEP (zero end expiratory pressure). ZEEP is obtained when the ventilator’s exhalation valve opens to the atmosphere, exposes the patient’s airway to the atmosphere, and exposes the patient’s airway to a relative pressure of zero. Pressures exceeding atmospheric are referred to as PEEP (positive end expiratory pressure) or CPAP (continuous positive airway pressure). Two well-documented phenomena applicable to PEEP/CPAP therapy are reduced functional residual capacity (FRC) and extravascular lung water (25–27).

Conditional Variables. Specific patterns of breath delivery are examined by the microprocessors control logic system. The ventilator may be set to keep the breath constant or perform more complex maneuvers. The ventilator algorithm for a particular set of breath delivery patterns makes this determination.

Modes of Ventilation. A mode of ventilation represents a combination of control, phase, and conditional variables that establish a set pattern of mandatory and/or spontaneous breaths (28).

Output. Most ventilator waveforms can be classified as one of the following:

1. Exponential
2. Ramp
3. Rectangular
4. Sinusoidal

No ventilator is an ideal controller, and ventilators are designed to only approximate a particular waveform (29,30).

Alarm Systems. Ventilator alarm systems have increased in number and complexity (31) and have recently begun to draw national attention on its proper use in and out of the critical care area (32).

VENTILATOR OPERATION

Introduction

The scope of medical practice expanding to accommodate the needs of those with respiratory diseases is immense. The field of respiratory therapy alone encompasses over 35,000 persons in the United States (33). As one may glean from the previous section of this article, not unlike other areas of medicine, when the need for new technology and therapeutic interventions is required, along with it will come a vast array of specialists, technical support, ancillary staff, and commercial entrepreneurs. The use of ventilators is expensive and represents approximately 10% incidence and contributes significantly to hospital costs (34). Although these costs are impressive, so are the diseases from which the patients suffer, and the true cost is patient mortality, which is invariably high (35). To that end, we have experienced increased understanding of the pathophysiology of their diseases as well as have enjoyed a significant increase in the technologic advances to support them.

The initial aim of ventilatory support has been justified to aid those persons suffering from neuromuscular disorders such as poliomyelitis or tetanus. The practice of medicine precludes one from solely defining a disease and indicating an intervention. But, rather, it identifies a derangement from normal and determines whether it causes harm. Should this be the case, then one should implement a therapeutic change. In the case of mechanical ventilatory support, we refer to the concept of work of breathing (WOB) (36).

$$\text{Work} = \text{Force} \times \text{Distance}$$

$$\text{Force} = \text{Pressure} \times \text{Area}$$

$$\text{Distance} = \text{Volume}/\text{Area}$$

$$\text{Work} = (\text{Pressure} \times \text{area})(\text{Volume}/\text{Area})$$

$$\text{Work} = \text{Pressure} \times \text{Volume}$$

$$\text{WOB} = \text{Tidal Volume} \times \text{Transpulmonary Pressure}$$

The tidal volume that is normally generated results from a relatively small pressure gradient from the intrapleural space to the atmosphere drawing gas inward. Exhalation is then passive when the pressure gradient is reversed, and the pressure then exceeds atmospheric. The relationship between pressure and volume is known as compliance. Pulmonary parenchymal (C_p) compliance is one component of total lung compliance (C_T) that would

also include the chest wall and rib cage (36).

$$C_T(L/cm H_2O) = \Delta V(L)/\Delta P(cm H_2O)$$

$$1/C_T = 1/C_L + 1/C_{CW}$$

Suffice it to say that most disease processes that impart ictus on patients and require ventilatory support alter pulmonary compliance, usually decreasing it. For example, a patient with pneumonia will generally have areas of alveoli, which are normally air-filled, collapsed, and filled with a pus-like edema matter. This fluid decreases lung compliance in the patient and increases the work (WOB) one must do to generate the same tidal volume to breathe. When this tidal volume is sacrificed, minute ventilation is diminished unless respiratory rate (RR) is increased.

$$\text{Minute ventilation} = \text{RR} \times \text{tidal volume}$$

Minute ventilation is a universal measure of the amount of gas that moves in and out of the lungs and sometimes indicates their global ability to remove carbon dioxide. Blood is delivered to the alveoli, and over a tremendous surface area, CO₂ diffuses freely and expires. As CO₂ diffusion is efficient, this process is uncommonly compromised. However, ventilatory failure is common. When a disease process, either of pulmonary etiology or otherwise, overcomes the normal compensatory mechanisms of the body, such that the patient can no longer maintain a normal pH and a normal PaCO₂, acute ventilatory failure is present. In this circumstance, the blood pH falls below a normal level and the PaCO₂ of arterial blood rises. The imposed, detrimental, WOB requires that the lungs and muscles of ventilation (i.e., chest wall, intercostals, sternocleidomastoid, and diaphragm) consume the oxygen that is delivered to them, hence, complicating matters because this oxygen is intended for the rest of the patient.

The remaining physiologic systems will begin to falter and malfunction as the acidemia worsens due to the fact that many proteins in the body operate within a narrow pH range (heart, brain, kidneys). This process typically will reveal a patient who is breathing rapidly and shallowly. The associated signs often include diaphoresis and the use of accessory muscles. However, of most concern, patients will describe a feeling of breathlessness, even that they are about to die. These are the patients with a clinical diagnosis of impending ventilatory failure.

Respiration Versus Ventilation

The terms “respiration” and “ventilation” are often incorrectly used interchangeably. It is imperative to offer a brief discussion depicting their differences. Respiration refers to the to and fro movement of a gas across a membrane. In the specific instance of breathing, we will consider the gases oxygen and carbon dioxide. Whereas oxygen is inhaled and taken up by the capillary endothelium and removed by the pulmonary venous system, carbon dioxide is delivered to the lungs by the pulmonary artery where it is eliminated by freely diffusing across the alveolar membrane. Typically, diffusion of carbon dioxide is very rapid, in contrast to oxygen, which requires approximately one third of the course of the capillary (37). The cardiopulmonary system maintains oxygen and carbon dioxide tensions within

the normal range (and a normal pH) despite changing metabolism. Any derangement from the normal compensatory homeostatic mechanisms (i.e., disease) may be life-threatening (38).

Ventilation simply defines the movement of a gas in and out of the pulmonary system (38). Clinical medicine makes a distinction and refers to carbon dioxide as a marker for ventilation. This is essentially because the tension of CO₂ inspired is usually zero, and is therefore a better indicator for the cardiopulmonary system’s ability to excrete it (38). Modern ventilators, although complex, are designed simply to improve oxygenation, ventilation, or both. Germane to the following, they are not always mutually exclusive.

Mechanical Ventilatory Support: Indications

Many have classified respiratory distress and the need for mechanical ventilatory support into categories that are useful. In addition, a thorough understanding of the physiologic derangement allows selection of the most optimal mode of support. However, simple definitions that are clear allow straightforward clinical judgment so that an intervention can be made and later fine tuned, (Table 1).

If one was to obtain an arterial blood gas and find a pH of less than 7.30 and a PaCO₂ greater than 60 mm Hg, they would meet criteria for acute ventilatory failure. This is because, at both this pH and PaCO₂, the CSF hydrogen ion concentration creates a maximal stimulus to breathe and represents an acute process. Impending ventilatory failure refers to the patient who is clearly in respiratory distress and frequently is found to be diaphoretic, tachypneic, using accessory muscles, and claims to be short of breath as described above. Apnea is intuitive, and the clearest example is a patient suffering from cardiac arrest. Although intermittently controversial, yet clinically important, hyperventilation for those patients with intractable intracranial hypertension will benefit from lowering carbon dioxide tensions with mechanical ventilatory support. Lastly, the concept of total ventilatory support is reserved for those persons with acute lung injury or adult respiratory distress syndrome at risk for ventilator-associated injury and severely decreased pulmonary compliance.

Mechanical Ventilatory Support: Goals

It seems appropriate to ascertain salient endpoints once a therapeutic intervention has been implemented such as mechanical ventilatory support. In this circumstance, if the ventilator can assume some or all of the patient’s WOB and alleviate the detrimental component, it would be helpful to classify the level of support provided to the patient

Table 1. Indications for Mechanical Ventilatory Support

Indication	Comments
Acute ventilatory failure	ABG diagnosis
Impending ventilatory failure	Physical diagnosis
Apnea	Cardiac arrest
Therapeutic hyperventilation	Neurosurgery
Total ventilatory support	Lung protective strategy

Table 2. Levels of Ventilatory Support

Support Level	Work of Breathing	Eucapnia and Normal pH
Partial ventilatory support	Significant	Yes
Full ventilatory support	Not required	Yes
Total ventilatory support ^a	Not required	No

^aFor use in patients with abnormal lung compliance. Reprinted with permission from Reference 39.

based on the amount of work the *patient* is *required* to provide to achieve a normal pH and a normal carbon dioxide tension in arterial blood (Table 2).

Full ventilatory support is likely the most common mode of ventilatory support employed in the postoperative period or immediately after intubation for one of the above indications for institution of mechanical ventilatory support. The aim of this intervention is to alleviate all of the patient's WOB until whatever medical or surgical disease has begun to resolve. Once a patient's respiratory status has stabilized, it seems prudent to allow one to resume spontaneous ventilatory effort and a significant amount of WOB.

It seems intuitive that once a breathing tube has been placed and ventilatory support has been instituted, it would be wise to work toward retuning the patient to a state free from those devices. Interestingly, there is evidence to support the concept that patients may not demonstrate the hypotensive effects of positive pressure ventilation in the presence of hypovolemia or decreased pulmonary compliance (40). In most circumstances, it is commonplace to transition to partial ventilatory support using various volume- or pressure-limiting modes (see below). As one is "weaned" from support, it is imperative that several items are taken into consideration. It has been suggested that some protocol be followed to standardize the routine and enhance safety. One such protocol, involves a daily screen and a spontaneous breathing trial (41).

Advantages that have been suggested include shorter days on mechanical ventilation, shorter intensive care unit stays and days in the hospital, and lower cost. To that end, safety still seems to be an appropriate concern. Patients must be monitored appropriately for any sudden deterioration in clinical status. Aside from standard intensive care unit monitoring, recommended additional parameters would include evidence of adequate ventilation by arterial

blood gas analysis. Should this not be available, continuous end-tidal carbon dioxide capnometry is an alternative. However, intermittent arterial pH and PaCO₂ should be obtained to correlate with the EtCO₂ values. The best monitor for the adequacy of oxygenation and ventilation remains the patient. Even while ventilated, it is imperative to engage in frequent assessments.

Modes of Support: Volume Versus Pressure

Despite all best efforts, at times there is a need for postoperative mechanical ventilation for a period of time. This may be due to residual anesthetic effects or the effects of surgery, especially abdominal or thoracic, on an already compromised respiratory status. When providing this support, several different modes of mechanical ventilation can be used, depending on the clinical situation at hand. Generally, during a period of acute stabilization, it is best to use a volume preset–pressure variable mode of ventilation (i.e., synchronized intermittent mechanical ventilation) and then to change to a volume variable–pressure preset mode (i.e., pressure support) when the patient is ready for withdrawal of mechanical ventilation (42). However, in postoperative patients, even in those with severe lung disease, the mode of ventilation is not as important as the endpoints of ventilation that are chosen (42).

The flexibility of modern mechanical ventilators has allowed for other considerations such as the degree of patient triggering, whether the breath is limited by pressure or flow, and the degree to which these pressures and volumes support the patient (Table 3).

An understanding of the limitations of the various modes is essential to success. For example, if one was to place a patient on assist/control mode and subsequently decrease the respiratory rate in an effort to wean the patient from support, the patient would continue to receive the preset tidal volume at his or her intrinsic respiratory rate and would not gain any intrinsic WOB. To the contrary, this method of weaning is perfectly acceptable in the synchronized intermittent mandatory ventilation mode (SIMV). In such a case, decreasing the SIMV rate would leave the patient to generate whatever tidal volume he or she could as determined by their transpulmonary pressure and pulmonary compliance (Fig. 1).

An alternative approach would be to change a patient from a volume-targeted mode of ventilation (i.e., SIMV or A/C) and convert to pressure support ventilation (PSV) for weaning purposes, perhaps. Having some discrete

Table 3. Characteristics of Volume-Targeted Ventilation (VTV) and Pressure-Targeted Ventilation (PTV)

Variable	Volume TV	Pressure TV
Trigger	Patient or time	Patient or time
Limit	Flow	Pressure
Cycle	Volume	Time or flow
Tidal volume	Constant	Variable
Peak pressure	Variable	Constant
Modes	Assist/control (synchronized) intermittent Mandatory ventilation	Assist/control (synchronized) intermittent Mandatory ventilation Pressure support

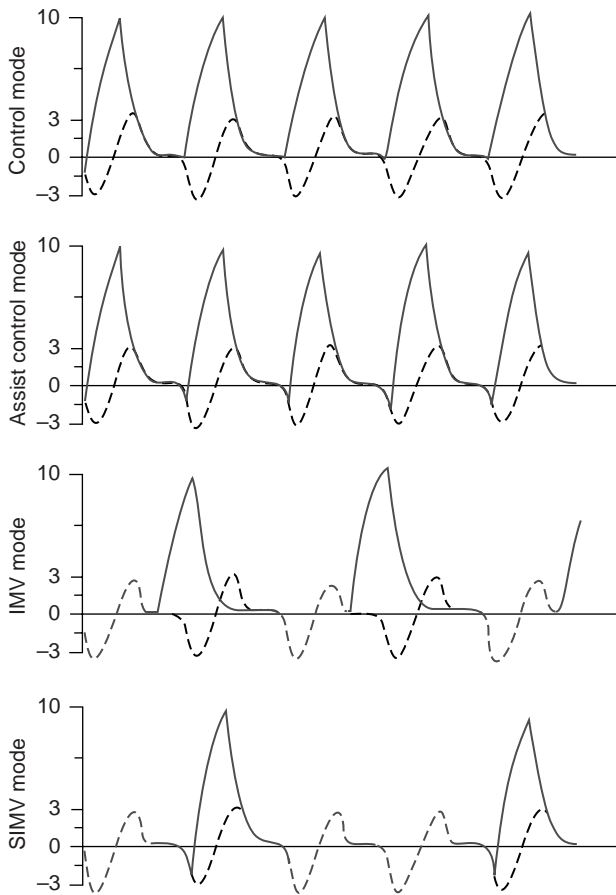


Figure 1. Airway pressure tracings of the four standard volume-preset modes. *Thick solid lines* represent ventilator breaths; *thick dotted lines* represent spontaneous breaths; and *thin dotted lines* refer to what the spontaneous pattern would have been without the ventilator breaths. IMV, intermittent mandatory ventilation; SIMV, synchronized IMV. Reprinted with Permission from Reference (36).

knowledge of the patient's arterial blood gas and a minute ventilation set on the ventilator commensurate with adequate oxygenation and ventilation, alternating between either, mode can mode be done easily. For example, if one has a ventilator set to deliver an SIMV respiratory rate of 10 by 700 mL, this then equates to about a 7 L minute volume. If the plateau pressure (point B in Fig. 2) generated is roughly 20 cm H₂O, a reasonable pressure support setting would then be 20 cm H₂O. Provided the patient respire about 10 times a minute, the same degree of oxygenation and ventilation should be anticipated. The minute volume would roughly be similar, and the volume-pressure relationship remains in tact provided there is no abrupt change in the patient's clinical condition.

As a patient's pulmonary disease regresses, lung compliance should improve. Using the ventilator monitor, the tidal volume that is generated from a pressure support breath should increase or the plateau pressure generated from a volume cycled breath should decrease (point B in Fig. 2). This can be a valuable piece of information often overlooked in the daily management of critically ill

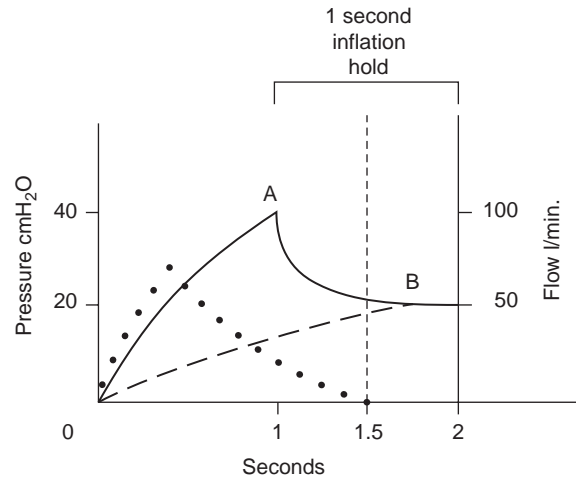


Figure 2. Schematic illustration of inspiratory cycle for volume-preset (1000 mL) tidal volume delivered by square-wave generator in 1 s, followed by 1-s inflation-hold maneuver. Note that the inspiratory cycle is 2 s, although tidal volume is delivered by the ventilator in 1 s. Normal compliance and resistance are assumed. Proximal airway pressure is shown as a *solid line*, alveolar pressure as a *dashed line*, and gas flow in airways as a *dotted line*. Peak airway pressure (A) is achieved near the end of tidal volume delivery, at which point there is considerable gradient between peak airway pressure and alveolar pressure that results in continued flow to the lung. Airway pressure rapidly diminishes as flow continues toward alveoli. Note that measurable flow in the system has essentially ceased 0.5 s after the ventilator has delivered tidal volume (B); however, a gradient remains between airway and alveolar pressures. An elevated peak pressure (A) in the presence of normal plateau pressures indicates increased airway resistance, which may include bronchospasm. To attain true plateau pressure, in which the airway pressure reflects the average peak alveolar pressure, an additional 0.5 s of inflation hold is required after the absence of measurable flow in the system. Reprinted with permission from Reference 42.

patients. Any difference between the peak and plateau pressures (points A and B in Fig. 2) would indicate an increase in airway resistance from bronchospasm, inspissated secretions, kinking of the tracheal tube, or some other form of incomplete obstruction.

PEEP/CPAP

Many patients suffer from a deficit in arterial oxygenation. The use of CPAP or PEEP is an effective means to remedy the malady. To clarify, the physiologic effect of CPAP/PEEP is identical; however, the nomenclature will vary depending on whether the patient is receiving positive pressure ventilation. One should use the term CPAP if the patient is not receiving positive pressure ventilation and is spontaneously ventilating and PEEP if the patient is receiving positive pressure.

Increasing the fraction of inspired oxygen, although simple and seemingly effective, to treat arterial hypoxemia may not be entirely benign. There is evidence that exposure of alveolar membranes to prolonged periods of high concentrations of oxygen will induce an oxygen-free radical species and subsequently lung injury (38). In addition,

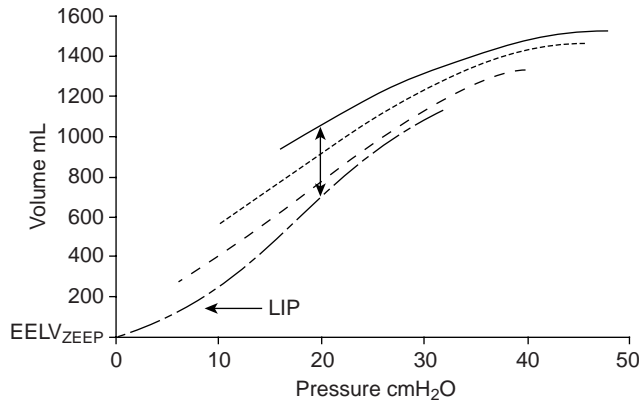


Figure 3. Multiple pressure/volume (P/V) curves of the respiratory system recorded from different levels of PEEP (---: PEEP 5; =: PEEP10; —: PEEP15) and from zero end-expiratory pressure (ZEEP; -.-), and related to the elastic equilibrium volume of the respiratory system at ZEEP ($EELV_{ZEEP}$). The curves are aligned on the same P/V diagram after correcting the starting points of each PEEP P/V curve for the volumes measured during passive expired spiroms performed at each PEEP level. The upward shift of PEEP P/V curves indicates alveolar recruitment. The recruited volume with 15 cm H_2O PEEP (double-headed arrow), compared with ZEEP, is quantified by the volume difference between the curves, for the same level of airway pressure (20 cm H_2O in this example). In other terms, when PEEP is applied, for the same level of airway pressure, lung volume is greater than without PEEP, suggesting the reopening of some alveolar units previously collapsed at ZEEP. Note that recruitment continues far above the value of pressure at the lower inflection point (LIP) and above to the upper inflection point. The different P/V curves tend to join at higher lung volumes, suggesting that the maximal lung volume is approached. Reprinted with Permission from Reference 43.

even shorter periods of high levels of inspired oxygen fractions limit the amount of nitrogen that may “stent” open; an alveolar unit can contribute to alveolar collapse, a phenomenon termed *denitrogenation absorption atelectasis* (38). The Law of LaPlace states that the pressure that tends to keep an alveolus open is twice the wall tension (T) divided by the radius and is represented by

$$P = 2T/r$$

To that end, many employ CPAP/PEEP therapy to improve arterial oxygenation and decrease WOB. WOB may be decreased by shifting a patient to a more favorable portion of the pulmonary compliance curve (Fig. 3).

On the flatter part of the curve, less tidal volume is generated with a substantial change in pressure. According to Fig. 3, on the steeper portion of the curve, less pressure is needed to generate a larger volume, which results in less work. By definition, the alveolar units would remain open at end-expiration, continually participating in gas exchange. It is this very concept of “open lung” ventilation that is thought to be the mainstay of avoiding ventilator-induced injury and ARDS (44,45).

CPAP/PEEP in most instances will have a significant impact on arterial oxygenation. Bringing alveoli into closer contact with the pulmonary capillaries will provide a

greater surface area for gas exchange, and as such, the pulmonary venous blood leaving the lungs will have a higher oxygen content. If one could make existing alveoli larger, opening previously closed alveoli, arterial oxygen tension would improve. At CPAP/PEEP levels less than 10 cm H_2O , existing open alveoli are made larger. When CPAP/PEEP levels exceed 15–50 cm H_2O , the collapsed alveoli are then opened and can participate in gas exchange, and thus they are recruited.

This fluid typically does not move out of the lung to some other area of the body. Rather, it shifts within the small spaces of the alveolar unit to the loose connective tissue-filled peri-bronchiolar areas that do not affect gas exchange. Perfusion will better match ventilation, thus increasing the oxygen content of arterial blood.

Mechanical Ventilatory Support: Adverse Reactions

Just as with any drug or medication that is prescribed, ventilators are not without iatrogenic potential. Typical side effects may include inadvertent hyperventilation or patient dysynchrony. These issues can be avoided with careful monitoring of the patient. Meticulous attention to artificial airway maintenance is mandatory. Relative to mechanical ventilatory support, the usual culprit of iatrogenic injury is excessive positive pressure termed *barotrauma* and large tidal volume in specific patients at risk for acute lung injury or ARDS referred to as *volu-trauma*.

Barotrauma is simply the disruption of the large conduction airways when exposed to very high pressures. When this occurs, gas leaks into the tissues of the neck

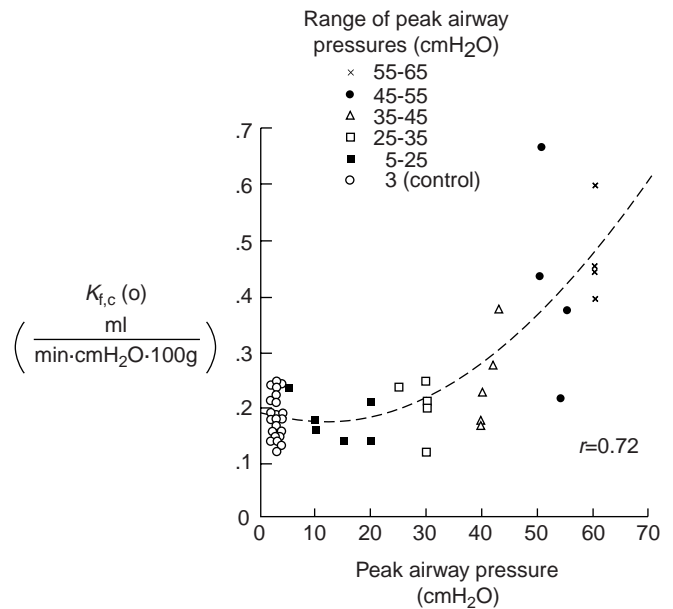


Figure 4. Changes in the capillary filtration coefficient ($K_{f,c}$) of isolated blood-perfused lobes of dog lungs given 20 min of intermittent positive pressure ventilation. Moderate (up to 30 cm H_2O) increases in peak airway pressure did not affect $K_{f,c}$, whereas higher peak pressure produced a steep increase in $K_{f,c}$. Reprinted with Permission from Reference 47.

and chest such as the mediastinum (pneumomediastinum) or the pleural space (pneumothorax). The most life-threatening of these is the tension pneumothorax, which allows gas to fill and accumulate in the pleural space. If there is no mechanism to remove it, the accumulated gas compressed the mediastinum and cardiac chambers compromising cardiac output to the point where venous return ceases and left ventricular output approaches zero.

ARDS, also referred to as acute lung injury, has been studied at great length. It has been noted that more patients survive in certain circumstances if lower tidal volumes are used to ventilate their lungs. In addition, it has been proven in several instances that chemical mediators of inflammation as released and adversely affect the body of large tidal volumes that are used (46) (Fig. 4). Recently, much has been learned about ARDS, and patient populations at risk for ventilator-induced injury have been identified.

BIBLIOGRAPHY

- Mushin M, Rendell-Baker W, Thompson PW, Mapleson WW. *Automatic Ventilation of the Lungs*. Oxford: Blackwell Scientific Publications; 1980. 62–166.
- Dupuis YG. *Ventilators—Theory and Clinical Application*. St. Louis, MO: CV Mosby Company; 1986.
- McPherson SP. *Respiratory Therapy Equipment*, 3rd ed. St. Louis, MO: CV Mosby Company; 1985.
- Cairo JM, Pilbeam SP. *Respiratory Care Equipment*, 6th ed. St. Louis, MO: CV Mosby Company; 1999.
- Chatburn RL. Classification of mechanical ventilator. *Respiratory Care* 1992;37:1009–1025.
- Kacmarek RM. *Critical Care Ventilators, Foundations of Respiratory Care*.
- ECRI. <http://www.ecri.org>.
- Blanch PB, DeSautels DA. Chatburn's ventilator classification scheme—poor substitute for the classic approach. *Respiratory Care* 1994;39:762.
- Consensus Statement on the essentials of mechanical ventilators—1992. *Respiratory Care* 1992;37:1000–1008.
- Chatburn RL. Classification of mechanical ventilators. *Respiratory Care* 1992;37:1009–1025.
- Branson RD, Chatburn RL. Technical description and classification of modes of ventilator operation. *Respiratory Care* 1992;37:1026–1044.
- Morris W. *The American Heritage Dictionary of the English Language*. Boston, MA: American Heritage and Houghton Mifflin; 1975.
- McPherson SP. *Respiratory Therapy Equipment*, 3rd ed. St. Louis, MO: CV Mosby Company; 1985.
- Dupuis YG. *Ventilators—Theory and Clinical Application*. St. Louis, MO: CV Mosby Company; 1986.
- McPherson SP. *Respiratory Therapy Equipment*. 3rd ed. St. Louis, MO: CV Mosby Company; 1985.
- Dupuis YG. *Ventilators—Theory and Clinical Application*. St. Louis, MO: CV Mosby Company; 1986.
- Chatburn RL. Classification of mechanical ventilators. In: Branson RD, Hess DR, Chatburn RL, editors. *Respiratory Care Equipment*. Philadelphia, PA: JB Lippincott Company; 1995.
- Spearman CB, Sheldon RL, editors. *Egans Fundamentals of Respiratory Therapy*. St. Louis, MO: CV Mosby Company; 1982.
- <http://www.puritanbennell.com>.
- Chatburn RL. Classification of mechanical ventilators. In: Branson RD, Hess DR, Chatburn RL, editors. *Respiratory Care Equipment*. Philadelphia, PA: JB Lippincott Company; 1995.
- Chatburn RL, Scanlan CL. Ventilator modes and functions. In: Scanlan CL, Wilkins RL, Stoller JK, editors. *Egan's Fundamentals of Respiratory Care*. 7th ed. St. Louis: Yearbook Mosby; 1999.
- Mushin M, Rendell-Baker W, Thompson PW, Mapleson WW. *Automatic Ventilation of the Lungs*. Oxford: Blackwell Scientific Publications; 1980.
- Chatburn RL. Classification of mechanical ventilators. *Respiratory Care* 1992;37:1009–1025.
- Kacmarek RM. *Critical Care Ventilators, Foundations of Respiratory Care*.
- Daly BDT, Edmonds CH, Norman JC. In vivo alveolar morphometrics with positive end expiratory pressure. *Surg Forum* 1973;24:217.
- McIntyre RW, Laws AK, Ramachandran PR. Positive expiratory pressure plateau: Improved gas exchange during positive pressure ventilation. *Can Anesth Soc J* 1969;16:477.
- Pare PD, Wariner B, Baile M, et al. Redistribution of pulmonary extravascular water with positive end expiratory pressure in canine pulmonary edema. *Am Rev Respir Dis* 1983;127:590.
- Chatburn RL. Classification of mechanical ventilators. *Respiratory Care* 1992;37:1009–1025.
- Kacmarek RM. *Critical Care Ventilators, Foundations of Respiratory Care*.
- Chatburn RL. Classification of mechanical ventilators. In: Branson RD, Hess DR, Chatburn RL, editors. *Respiratory Care Equipment*. Philadelphia, PA: JB Lippincott Company; 1995.
- Chatburn RL. Mechanical ventilators: Classification and principles of operation. In: Hess DR et al., editors. *Respiratory Care, Principles and Practice*. Philadelphia, PA: WB Saunders Company; 2002.
- JCAHO Alarm event. <http://www.jcaho.org>.
- <http://www.aarc.org>.
- Neddham DM, Bronskill SE, Sibbald. Mechanical ventilation in Ontario, 1992–2000: Incidence, survival, and hospital bed utilization of non-cardiac surgery adult patients. *Crit Care Med* July 2004.
- Depuydt PO, Benoit DD, Vandewoude KH. Outcome in non-invasively and invasively ventilated hematologic patients with respiratory failure. *Chest* October 2004.
- Miller. *Anesthesia*. 5th ed. Philadelphia, PA: Churchill Livingstone.
- Guyton. *Textbook of Medical Physiology*. 10th ed.
- Shapiro BA, Peruzzi WT, Templin R. *Clinical Application of Blood Gasses*, 5th ed.
- Peruzzi WT, Shapiro BA. Changing practices in ventilator management: A review of the literature and suggested clinical correlations. *Surgery* 1995;117:121–133.
- Stock C, Perel A. *Handbook of Mechanical Ventilatory Support*. 2nd ed. Philadelphia, PA: Williams and Wilkins; 1997.
- Murray J, Coursin DB, Pearl RG. *Critical Care Medicine*. Philadelphia, PA: Lippincott Williams, Wilkins.
- Tamul PC, Peruzzi WT. Assessment and management of patients with pulmonary disease. *Crit Care Med* April 2004.
- Maggiore SM, Richard JC, Brochard L. What has been learned from P/V curves in patients with acute lung injury/acute respiratory distress syndrome. *Eur Respir J* 2003.
- Grasso S, Mascia L, Del Turco M. Effects of recruiting maneuvers in patients with acute respiratory distress syndrome ventilated with protective ventilatory strategy. *Anesthesiology* April 2002.

45. Schreiter D, Reske A, Stichert B. Alveolar recruitment in combination with sufficient positive-end expiratory pressure increases oxygenation and lung aeration in patients with severe chest trauma. *Crit Care Med* April 2004.
46. Amato MB, Barbas CS, Medeiros DM. Effect of a protective-ventilation strategy on mortality in the acute respiratory distress syndrome. *NEJM* Feb 1998.
47. Dreyfuss D, Sauman G. Ventilator-induced lung injury: Lessons from experimental studies. *Am J Respir Crit Care Med* 1998; 157.

See also **CARDIOPULMONARY RESUSCITATION; CONTINUOUS POSITIVE AIRWAY PRESSURE; VENTILATORY MONITORING.**

VENTILATORY MONITORING

STEVEN R. KNOPER
STUART F. QUAN
University of Arizona College of
Medicine

INTRODUCTION

One of the most important aspects of the clinical care of patients with known or suspected cardiopulmonary disease is monitoring the adequacy or status of their ventilatory function. However, in order to understand why various parameters of lung function are monitored, a few aspects of basic respiratory physiology require review. Although the lungs perform important metabolic and endocrine functions, their primary function is gas exchange. They provide the vehicle by which oxygen (O_2) in inspired air moves into venous blood and carbon dioxide (CO_2) in venous blood is released into expired gas. The mechanisms by which these processes occur are complex, but can be subdivided into four phases. The first phase involves the movement of gas into and out of the lungs themselves. This aspect of respiratory physiology is termed lung mechanics. During spontaneous breathing, contraction of the diaphragm and intercostal muscles increases negative pressure within the intrapleural space, leading to lung expansion. This results in a pressure gradient along the airways causing gas to flow into the lungs and to eventually reach the alveoli. With relaxation of the inspiratory muscles, the elastic recoil of the lungs reverses the pressure gradient resulting in gas being expired through the mouth into the atmosphere. In the second phase, gas exchange occurs in the alveoli. By the process of diffusion across the alveolar capillary membrane, O_2 enters and CO_2 exits the vascular system. The third phase involves the transport of O_2 and CO_2 in the blood, and is intimately related to adequate cardiovascular and hematologic functioning. Oxygen is carried in the blood primarily in a complex with hemoglobin in the red blood cells. A small amount of O_2 is also dissolved in plasma. The total amount of O_2 delivered throughout the body is a function of the volume of O_2 in the blood, and the rate of delivery or cardiac output. Transport of CO_2 in the body is more complex, but also is related to the properties

of hemoglobin and cardiac function. The fourth phase is cellular respiration (defined as the metabolic consumption of O_2 and production of CO_2 and H_2O), and involves the mechanisms by which cells in the body utilize O_2 and excrete CO_2 . All of these phases are related to each other, and are essential for normal respiration to occur. Ventilatory monitoring encompasses methods by which the functional status of the aforementioned aspects of lung function can be determined. This article reviews the indications, principles, and techniques used in clinical ventilatory monitoring. A detailed discussion of pulmonary function testing is not included, and is reviewed elsewhere (1).

INDICATIONS FOR VENTILATORY MONITORING

Ventilatory monitoring usually is performed for four general clinical indications: (1) diagnostic, (2) therapeutic, (3) prophylactic, and (4) safety considerations. With respect to diagnostic use, the goal of monitoring is to obtain information pertinent to making a specific etiologic diagnosis of a disease entity or process. For example, in myasthenia gravis patients who are having an acute deterioration in ventilatory function, vital capacity (VC) measurements (the maximum volume excursion of which the lungs are capable by voluntary effort) are obtained after injection of a short-acting cholinesterase inhibitor (edrophonium, Tensilon). Documentation of an increase in VC suggests the need for an increase in anticholinesterase medication, whereas a deterioration is evidence for excessive effect from medication. Another frequently encountered situation where ventilatory monitoring is used diagnostically is the continuous monitoring of airflow and ventilatory effort in sleep disorders laboratories to detect evidence of sleep disordered breathing (2).

A number of ventilatory parameters are useful in monitoring the therapeutic effect of specific treatment modalities. The most frequently encountered situation is where arterial O_2 and CO_2 gas tensions are sampled following institution of supplemental O_2 or mechanical ventilation. Measurement of respiratory system compliance following application of positive end-expiratory pressure (PEEP) in mechanically ventilated patients with acute respiratory failure is another example of where ventilatory monitoring is used to determine the effects of a treatment modality (3). In both the aforementioned examples, therapy frequently is adjusted using results obtained from the ventilatory parameters monitored.

Ventilatory monitoring commonly is performed prophylactically with the goal of detecting a perturbation in normal or baseline respiratory function so that treatment can be instituted early. For example, apnea monitors are employed in infants suspected to be at risk for sudden infant death syndrome. In this clinical situation, alarms are activated if life-threatening apnea or bradycardia occurs so that the infant can be resuscitated before a full cardiorespiratory arrest ensues.

Finally, safety considerations require ventilatory monitoring when certain categories of biomedical life support devices are used. In patients with pulmonary disease, a

Table 1. Levels of Ventilatory Monitoring

Visual observation
Intermittent objective measurements
Continuous or automated measurements
Computerized analysis with treatment algorithms

commonly encountered example is the use of mechanical ventilators. With mechanical ventilation, a number of ventilatory parameters with appropriate alarms are monitored so as to avoid inadvertent disconnection and dangerously high airway pressures. Another example is patients undergoing general anesthesia. Inspired gas concentrations delivered by anesthesia machines frequently may be monitored to insure that the concentrations of anesthetic gases and O_2 are appropriate (4).

LEVELS OF VENTILATORY MONITORING

As emphasized elsewhere (5), there are several levels of intensity with respect to ventilatory monitoring (Table 1). On the most basic level is visual observation and examination by experienced personnel. Although inherently simple, examination of patients is an attribute of clinical medicine that is irreplaceable. The clinical "gestalt" obtained by an experienced healthcare practitioner from observing the work of breathing and the breathing frequency of a patient in respiratory distress has not yet been duplicated by any automated monitoring device. Furthermore, physical examination of the respiratory system is an "art" that should not be lost on future physicians, physician's assistants, nurses, and respiratory care practitioners. In many patients with pulmonary disease, little monitoring other than periodic observation and examination is required.

In some patients with pulmonary disease, more objective data than can be obtained with visual observation and examination are needed. Therefore, the next level of monitoring intensity is the use of intermittent objective measurements of ventilatory function. Probably the most common ventilatory parameter obtained at this level of monitoring are arterial blood gases. Periodic sampling of arterial blood to determine pO_2 , pCO_2 , and pH is used in almost all patients with the presence or suspicion of significant ventilatory dysfunction. Arterial blood gases are invaluable in following the clinical course and the effect of ventilatory therapy rendered in these patients. Arterial blood gas tensions currently are considered the "gold standard" against which any other measurement of gas exchange is compared. Abnormalities in pO_2 and pCO_2 now are used as definitions of the degree of respiratory impairment. Other frequently used intermittent monitors of ventilatory function are measurements of the VC, the volume of gas exhaled in the first second during a forced expiration (FEV_1), the peak expiratory flow rate during a forced expiration (PF), and the maximum negative airway pressure generated during inspiration against an occluded airway (maximum inspiratory force, MIF). All of these parameters may be useful in following the clinical progress of patients with certain types of pulmonary disease.

Patients who are critically ill may require the third level of monitoring intensity, which is continuous measurement of various ventilatory parameters over time. In most intensive care units, devices are available to continuously track breathing frequency. Patients receiving mechanical ventilation have continuous monitoring of their airway pressures and exhaled tidal volumes for safety considerations. Recent technological advances now allow for continuous monitoring of arterial and venous oxygen saturations. In the most sophisticated intensive care units, on-line monitoring of exhaled gases from ventilated patients and complex measurements of lung mechanics, such as resistance and compliance, can be performed.

Finally, the highest level of monitoring is the use of computer algorithms to not only monitor various ventilatory parameters, but to also adjust therapy according to the values obtained without human intervention. The technology required for such monitoring is undergoing continuous development, and there is increasing use in patient care applications. This is particularly evident on newer microprocessor ventilators where some ventilator modes allow for automatic adjustments to the level of ventilator support according changes in patient lung mechanics.

PARAMETERS USED IN MONITORING LUNG MECHANICS

Breathing Frequency

The most common measurement of ventilatory function is an assessment of the patient's breathing frequency. Usually, this is performed by visually counting the breathing frequency over a 30–60 s period of time, although biomedical devices also can be used for this purpose. Normal individuals breathe at a frequency of ~ 12 min. However, this rate is markedly altered in patients with lung disease. Reductions in spontaneous respiratory rate are diagnostic of ventilatory dysfunction and suggest central depression of ventilatory control centers in the brain stem. Examples of clinical situations in which this might occur are narcotic overdose, incomplete recovery from a general anesthetic, and primary central alveolar hypoventilation. Whereas reductions in breathing frequency indicate ventilatory dysfunction, an increase in spontaneous breathing frequency (tachypnea) is more nonspecific since individuals can voluntarily increase their breathing frequency in response to physical or psychological stress, or an increase in metabolic demands. However, if a nonpulmonary basis for an increase in breathing frequency can be excluded, tachypnea generally indicates the presence of lung disease. Common examples are pneumonia, congestive heart failure with pulmonary edema, and asthma.

Lung Volume and Airflow. An assessment of the volume of air inspired and expired with each breath (tidal volume, V_T) is an important ventilatory parameter since adequacy of ventilation is determined not only by the breathing frequency, but also by the V_T . The spontaneous V_T in normal individuals varies according to their size, but in an average adult is ~ 500 mL. Ventilatory dysfunction, whether from processes such as parenchymal pulmonary disease or a reduction in central respiratory drive, generally results in a

decrease in V_T . In mechanically ventilated patients, exhaled V_T is often measured to monitor for inadvertent disconnection or leaks in the ventilator circuit. Sudden declines in exhaled V_T herald disconnection or large ventilator circuit leaks, and on most ventilators trigger audible alarms.

The VC and FEV₁ are frequently monitored intermittently in ambulatory patients with diseases characterized by airflow limitation such as asthma, asthmatic bronchitis, and emphysema. In these diseases, a reduction in comparison to predicted norms (6) is observed in both these parameters. In addition, the ratio of the FEV₁ to the forced VC (FVC), which is normally >0.7, is decreased. The VC by itself is a useful parameter to follow in patients with acute respiratory failure. In spontaneously breathing patients, one criterion for intubation and mechanical ventilation is a VC <10 mL/kg body weight. Conversely, in mechanically ventilated patients, a VC <10 mL/kg body weight is used as a criterion to wean mechanical ventilatory support. Obviously, many other factors are involved in a decision to remove or supply mechanical ventilation, but VC definitely is one key determinant.

The rate of inspiratory air flow is an important ventilatory parameter during mechanical ventilation. It determines the duration of inspiration, and influences airway pressures and the distribution of ventilation in the lungs. Higher peak inspiratory flow rates result in a shortening of the duration of inspiration, higher peak airway pressures, and less uniform intrapulmonary distribution of inspired gas. However, if inspiratory flow rates are too low, inspiration becomes longer than expiration. This may lead to incomplete exhalation, gas trapping in the lung, and an increase in the functional residual capacity (volume of the lung at end-expiration, FRC). These physiologic effects may have deleterious consequences with respect to a decline in cardiac output and an increase in the incidence of pulmonary barotrauma (i.e., pneumothorax, pneumomediastinum, and subcutaneous emphysema).

Monitoring of expiratory flow rates is useful in following the clinical course of patients with airflow limitation. Measurement of flow rates during the middle of a FVC maneuver is a frequently performed pulmonary function test and will not be discussed further. However, periodic determination of PF is a useful monitoring technique to follow the progress of patients with acute and chronic asthma, and also is used to monitor pulmonary function in outpatient studies of patients with suspected bronchospasm resulting from occupational or environmental exposures. Normal values are <500 L/min, but measurements <120 L/min frequently are observed during acute exacerbations of asthma.

Airway and Esophageal Pressures. In patients receiving mechanical ventilation, perhaps one of the most important ventilatory parameters that can be monitored is airway pressures. With volume constant ventilators, a feedback loop is present so that the peak airway pressure cannot exceed a manually preset level. When such a situation occurs (e.g., a patient coughing while inspiratory flow is being delivered from the ventilator), inspiration is terminated when the preset pressure is reached, and an alarm is activated. In this case, peak airway pressure monitoring

acts as a safety feature to prevent overpressurization of the airway and the possibility of barotrauma. In contrast, when pressure limited ventilators are used, peak airway pressure is set as a ventilator control parameter and is one of the factors that determines the V_T . Airway pressures are also important in monitoring for inadvertent disconnection or unrecognized leaks in the ventilator circuit during mechanical ventilation. A sudden marked reduction in airway pressure is indicative of a leak in the ventilator circuit, whereas a reduction in the airway pressure to atmospheric pressure would signal a disconnection. On most ventilators, auditory alarms are activated when such events occur. Inspection of the airway pressure curve during mechanical ventilation also can yield important information. As examples, a marked downward concavity at the onset of inspiration is a sign of an insufficient inspiratory flow rate to meet patient demand. During expiration, the failure of the airway pressure curve to return to baseline with an end-expiratory occlusion is indicative of auto-PEEP or air trapping. Finally, airway pressure measurements are an important component of derived ventilatory parameters such as airway resistance and lung compliance.

Monitoring of the maximum inspiratory force (MIF) in patients who are receiving mechanical ventilation is an important ventilatory parameter in determining whether they can be weaned from mechanical ventilation. The MIF can be considered an indicator of the underlying strength of the inspiratory muscles. Values more negative than -25 cm H₂O usually indicate that a patient will be able to maintain adequate ventilation without mechanical ventilatory support, although other factors obviously must also be taken into account.

Measurement of esophageal pressure is used as a reflection of intrapleural pressure (7). It is therefore a component of calculated ventilatory parameters such as lung compliance and the work of breathing. Esophageal pressure catheters also are used to quantify ventilatory effort during sleep studies (polysomnography). However, although there are commercially available devices that incorporate esophageal pressure monitoring to measure lung mechanics, they have not found common acceptance, because placement of an esophageal catheter is invasive and technical factors make accurate determinations difficult (7).

Resistance, Compliance, Volume-Pressure Curves. The total pulmonary resistance to the flow of gas into the lungs is comprised of two factors: (1) friction between the molecules of flowing gas and the airways (airways resistance, R_{aw}), and (2) resistance of the tissues (lungs, rib cage, diaphragm, abdominal contents) as a result of their own displacement (tissue resistance) (8). Elevations of airways resistance are observed in patients with diseases characterized by airflow limitation such as asthma and emphysema. In addition, superimposition of external devices such as narrow endotracheal or tracheostomy tubes can increase airways resistance. Tissue resistance can be increased in patients with such disorders as kyphoscoliosis or pulmonary fibrosis. Total pulmonary resistance is not a frequently used clinical measurement since it requires the simultaneous measurement of lung volume, airway, and esophageal (as a reflection of pleural) pressures (8). Similarly,

airways resistance is also not performed on most patients since it is necessary to use a body plethysmograph (8). No direct measurement of tissue resistance is available, but it is generally calculated as the difference between total pulmonary resistance and airways resistance. Normally, the pulmonary tissue resistance represents ~20% of the total pulmonary resistance. In spite of the difficulty in obtaining a precise measurement of pulmonary resistance in mechanically ventilated patients, an estimate of the pressure needed to overcome airways resistance can be obtained by measuring the difference between the peak airway pressure (p_{\max}) and the airway pressure observed after a 1 s inspiratory hold (p_{st}). If airflow measurements (V) at the point of peak airway pressure are available, the equation:

$$R_{\text{aw}} = (p_{\max} - p_{\text{st}})/V$$

is an estimate of airways resistance. In general, however, estimates of resistance have not been shown to provide important information relating to the care of patients.

Measurement of compliance or elasticity of the lung is of more importance to the care of patients than resistance. The compliance of a biological system can be defined as the change in volume occurring in response to a change in distending pressure. In addition, compliance measurements can be obtained during static conditions (no flow), or during dynamic conditions (continuous flow). With respect to the respiratory system, static respiratory system compliance (C_{st}) is defined as the change in lung volume occurring with a change in airway pressure over the tidal volume range. Similarly lung compliance (C_{L}) is defined as the change in lung volume with respect to the change in pleural pressure over the tidal volume range. The relationship between C_{st} and C_{L} is defined by the following equation:

$$\frac{1}{C_{\text{st}}} = \frac{1}{C_{\text{L}}} + \frac{1}{C_{\text{t}}}$$

where C_{t} is the compliance of the thorax.

Dynamic lung compliance (C_{dyn}) is the change in lung volume occurring with a change in pleural pressure between two points of instantaneous zero flow during breathing (9).

In the respiratory system, changes in compliance are a function of two factors: the elasticity of the tissues and lung volumes. In the upright position, the normal value for C_{st} in adults is ~100 mL/cm H₂O, and the normal value for C_{L} is 2.0 L/kPa. However, many types of lung pathology such as pulmonary edema and pneumonia result in a reduction of both C_{st} and C_{L} . In addition, normal values decline with decreasing lung volume. Thus, the normal C_{st} of a small animal or infant is less than an adult human. Dynamic lung compliance and C_{L} are virtually identical in the absence of significant parenchymal lung disease. However, with airways disease, all alveoli do not ventilate evenly during dynamic conditions, with some gas exchanging areas ventilating slower than others. This phenomenon also has been described as time constant disparity. Because of this disparity in time constants, these slow ventilating areas do not participate fully in gas exchange, and in effect a smaller lung volume is ventilated with each breath. In

contrast, during static conditions, a disparity in time constants does not influence the distribution of ventilation. Therefore, C_{dyn} is $<C_{\text{L}}$ in the presence of significant airways disease. Another measure of lung elasticity called the "dynamic characteristic" (C_{dch}) (10) also has been described. It is calculated in patients receiving mechanical ventilation as V_{T} divided by the difference between P_{\max} and PEEP. However, C_{dch} is not just a reflection of lung compliance, but also is influenced by changes in airways resistance.

Measurements of both C_{dyn} and C_{L} require an estimate of intrapleural pressure that, in most cases, is obtained using an esophageal balloon. Because accurate esophageal pressures are difficult to obtain, especially in critically ill patients, clinical measurement of these ventilatory parameters are uncommon. However, C_{st} only requires measurement of V_{T} and airway pressure, and thus is easily obtained in most patients receiving mechanical ventilation. Since, in most patients, changes in C_{st} are primarily a result of changes in C_{L} , serial measurements of C_{st} frequently are used to follow changes in lung elasticity. For example, maximizing C_{st} during application of PEEP has been shown to correspond to the best levels of oxygen transport (product of the cardiac output and the arterial oxygen content) (3). Serial measurements of both C_{st} and C_{dch} are useful in many patients with acute respiratory failure who are receiving mechanical ventilation. A change in both these parameters usually signifies an alteration in the properties of the lung parenchyma such as increasing pulmonary edema or pneumonia. However, a change in C_{dch} without a change in C_{st} is indicative of an increase in airways resistance such as might occur with the onset of bronchospasm or mucous plugging of the airways (10).

Measurements of compliance and resistance summarize pressure–volume relationships in the lung. These can be depicted graphically with two-dimensional (2D) plots of volume versus pressure. A decrease in compliance will be represented by a decrease in slope of this relationship, whereas bowing of the volume–pressure relationship is observed when airways resistance is increased. Analysis of the inspiratory volume–pressure curve may aid in determining the optimum level of PEEP for patients with the acute respiratory distress syndrome who require mechanical ventilation. The optimum level of PEEP is suggested by the inspiratory airway pressure at which the slope of the volume–pressure curve appears to increase (11). In some intensive care units, on-line plots of volume versus pressure can be obtained, but their value in the care of patients has not been proved.

Work of Breathing. In physics, linear work is defined as the product of force times distance. The analogous situation when applied to the movement of air into the lungs is described by the product of pressure times volume. Therefore, the work of breathing is equivalent to the cumulative product of the transpulmonary pressure and the volume of air moved at each instant in time (12). Work of breathing measurements have been proposed as a method of determining whether mechanically ventilated patients are able to breathe spontaneously (13). Unfortunately, accurate determinations require simultaneous measurements of

pleural pressure (usually estimated from esophageal pressure), and airflow. Therefore, work of breathing measurements have not found great utility in the clinical management of patients.

Other Parameters. The rapid shallow breathing index (RSBI) is frequently used to assess the ability of a patient to breath without ventilator support (14). It is calculated by measuring the breathing frequency and dividing it by the V_T in liters. High values (>105) indicative of low V_T and a high breathing frequency suggest that a patient cannot be removed from the mechanical ventilator. However, there has been some controversy concerning the predictive value of this parameter in all clinical situations.

Ventilatory drive is another parameter that can be monitored. Although measurement of the minute ventilation in response to hypoxia or hypercapnia are indexes of central drive in the absence of impairment of neuromuscular or mechanical pulmonary disease, measurement of the airway pressure generated within the first 100 ms after airway occlusion ($p_{0.1}$) is generally recognized as the best indicator of the central drive (15). Unfortunately, this is difficult to accomplish without specialized equipment.

PARAMETERS USED IN THE ASSESSMENT OF GAS EXCHANGE

Oxygenation

The amount of oxygen contained in arterial blood (oxygen content, CaO_2) is determined by the arterial oxygen tension (PaO_2), the percent saturation of hemoglobin (Hgb) with oxygen (S_aO_2), and the Hgb concentration in arterial blood:

$$CaO_2 = 1.34 \times (\text{Hgb}) \times (S_aO_2) + 0.003 \times (PaO_2)$$

where $CaO_2 = \text{mL } O_2/100 \text{ mL blood}$, $1.34 = \text{mL } O_2/\text{g Hgb}$, $\text{Hgb} = \text{g}/100 \text{ mL blood}$, $S_aO_2 = \% \text{ saturation}$, $0.003 = \text{mL } O_2 \text{ dissolved in blood}/\text{mmHg}$, $PaO_2 = \text{mmHg}$.

In most circumstances, adequacy of oxygenation is monitored by measurement of the PaO_2 or the S_aO_2 . Although the PaO_2 and S_aO_2 are intimately related to each other, their relationship is not linear (Fig. 1), but rather is described by a sigmoid curve. At levels of oxygen tension greater than $\sim 60 \text{ mmHg}$, there is little increase in S_aO_2 with increasing values of PaO_2 . In contrast, when the PaO_2 falls $<60 \text{ mmHg}$, S_aO_2 declines rapidly with decreases in PaO_2 . Therefore, monitoring of S_aO_2 as an estimate of PaO_2 is of little value when the PaO_2 is $>70\text{--}80 \text{ mmHg}$, since significant alterations in PaO_2 will not be reflected by corresponding changes in S_aO_2 . At sea level, the normal PaO_2 is $\sim 100 \text{ mmHg}$ and the S_aO_2 is 97%. However, normal values may vary according to alterations in barometric pressure and increasing age (16). Occasionally, arterial oxygen content (normal value = $9.5 \text{ mL } O_2/100 \text{ mL blood}$) is monitored as an indicator of oxygenation. Because most of the oxygen in blood is carried in combination with Hgb, this may be of value when Hgb levels are severely abnormal or fluctuating rapidly.

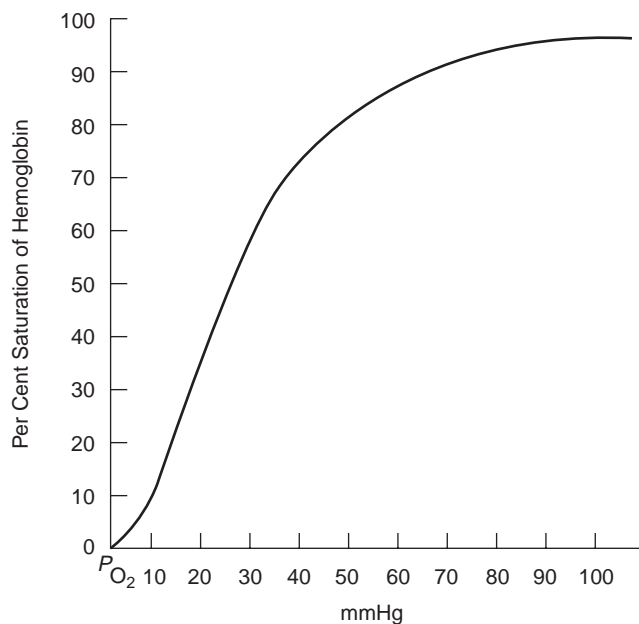


Figure 1. Oxyhemoglobin dissociation curve. The graph shows the relationship between hemoglobin saturation and oxygen tension. Above 60 mmHg, the curve becomes progressively flatter so that little additional saturation occurs when the PO_2 is $<60\text{--}70 \text{ mmHg}$.

Mixed venous oxygen tension ($P_{\bar{v}}O_2$) or saturation ($S_{\bar{v}}O_2$) generally is regarded as an indicator of cardiac output or metabolic consumption of oxygen in the tissues. However, during continuous monitoring of $S_{\bar{v}}O_2$ in critically ill patients, an abrupt fall in $S_{\bar{v}}O_2$ may herald a sudden deterioration in arterial oxygenation.

There are several parameters derived from measurements of P_aO_2 , S_aO_2 , $P_{\bar{v}}O_2$, and $S_{\bar{v}}O_2$ that are occasionally used in monitoring the status of oxygen gas exchange. Higher inspired oxygen concentrations (FIO_2) result in higher P_aO_2 values. In order to compare P_aO_2 values obtained at differing FIO_2 values, the P_aO_2/FIO_2 ratio or its reciprocal, the FIO_2/P_aO_2 ratio, can be computed. It is assumed, however, that these ratios remain constant with changing FIO_2 in the presence of lung disease, and that the P_aCO_2 is constant. Since changes in P_aCO_2 result in approximately reciprocal changes in P_aO_2 , the status of oxygen gas exchange is sometimes determined by calculating the alveolar–arterial oxygen tension difference ($AaDO_2$; A-a gradient), which controls for changes in oxygenation resulting from changes in the P_aCO_2 . This is performed by calculating the alveolar oxygen tension (P_AO_2) using the alveolar air equation:

$$P_AO_2 = (P_B - PH_2O) \times FIO_2 - P_ACO_2/R + [P_ACO_2 \times FIO_2 \times (1 - R)/R]$$

where P_B is barometric pressure, PH_2O is water vapor pressure, P_ACO_2 is alveolar PCO_2 tension, and R is the respiratory gas exchange ratio. It is generally assumed that the $P_ACO_2 = P_aCO_2$ and that R is ~ 0.8 . After the P_AO_2 is computed using the alveolar air equation, the $AaDO_2$ is

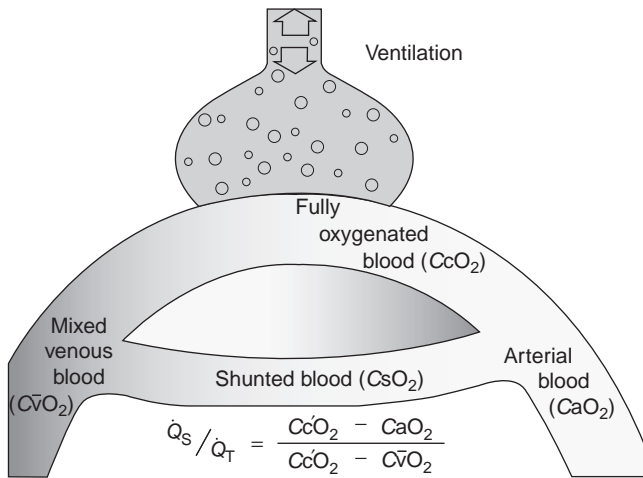


Figure 2. Two-compartment lung model and shunt equation. Conceptually, mixed venous blood either perfuses normal lung and is completely oxygenated or is shunted past the lungs and receives no oxygen. Both oxygenated blood and shunt blood combine to form arterial blood.

the difference between the $P_{A}O_2$ and the $P_{a}O_2$. $AaDO_2$ changes with age, $F_{I}O_2$, and position. In a young, healthy person breathing air at sea level, this value is usually 10 mmHg, and 100 mmHg on 100% $F_{I}O_2$.

The principal disadvantage to the aforementioned indicators of oxygen gas exchange is that they all fail to account for changes in cardiac output, confounding the interpretation of changes in these indicators in the presence of lung disease. In general, impairments in oxygen gas exchange can be quantified using a two-compartment model. As shown in Fig. 2, mixed venous blood can be thought of as either becoming fully oxygenated by the lungs or “shunted” through the lungs without receiving any oxygen. The percentage of the total cardiac output (Q_t) that is “shunted” (Q_s) is termed the shunt fraction (Q_s/Q_t). Since changes in Q_t may result in changes in the oxygen content of mixed venous blood ($C_{\bar{v}}O_2$), arterial oxygenation may be altered without any change in Q_s/Q_t (Fig. 3). Therefore, a change in arterial oxygenation may be misconstrued as resulting from an alteration in gas exchange in the lungs when a change in cardiac output is actually responsible. To

circumvent this problem, Q_s/Q_t occasionally is calculated:

$$Q_s/Q_t = C'_cO_2 - C_aO_2 / C'_cO_2 - C_{\bar{v}}O_2$$

where C'_cO_2 is the pulmonary capillary oxygen content. Usually C'_cO_2 is calculated by assuming that the hemoglobin in pulmonary capillary blood is fully saturated with oxygen and that the $p_{A}O_2$ reflects the amount of dissolved oxygen in pulmonary capillary blood. Unfortunately, computation of Q_s/Q_t requires the presence of a pulmonary artery catheter for sampling of mixed venous blood, which makes it unfeasible as a ventilatory parameter in many patients. In the vast majority of patients, however, measurement of the $p_{a}O_2$ and, in some situations, the S_aO_2 , will provide satisfactory indicators of oxygen gas exchange. Calculation of the aforementioned derived parameters should be reserved for selected critically ill patients in whom such information is necessary for their care.

Ventilation. Adequacy of ventilation is defined by the levels of arterial pCO_2 (p_aCO_2) and pH. These parameters are usually obtained from arterial blood gas samples, and are normally 40 and 7.40 mmHg, respectively. Although hypoventilation is defined as a p_aCO_2 significantly >40 mmHg and hyperventilation is present when the p_aCO_2 is <40 mmHg, the appropriateness of the lung’s ventilatory response is determined by the arterial pH. For example, in the presence of a metabolic acidosis, hyperventilation is appropriate if the pH approaches 7.40. However, the hyperventilatory response would be considered inadequate if the pH remained significantly <7.40.

Precise determination of the p_aCO_2 and arterial pH can only be obtained by sampling of arterial blood. Venous pCO_2 and pH occasionally can be used as indirect indicators of arterial values, but this is not recommended since these parameters also reflect cellular metabolism. Of more use as a reflection of p_aCO_2 are measurements of end tidal pCO_2 ($p_{et}CO_2$). In normal individuals, the composition of the gas at the end of a normal exhalation primarily represents gas originating from alveolar areas of the lung. Therefore, the pCO_2 of the gas at the end of a normal tidal breath should approximate the alveolar pCO_2 (p_ACO_2). Since the alveolar–arterial pCO_2 gradient across the alveolar capillary membrane is only 1–2 mmHg, $p_{et}CO_2$ is a good estimate of the p_aCO_2 in normal individuals. In the presence of lung disease, however, the distribution of

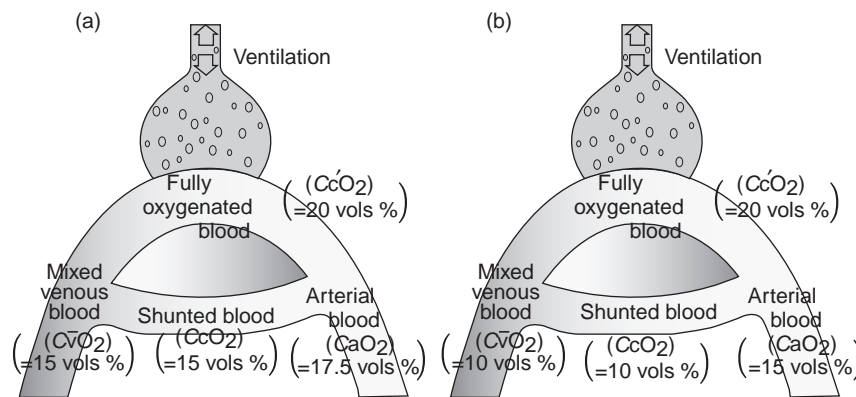


Figure 3. Effect of changes in mixed venous oxygen content on arterial oxygen content in a two-compartment lung model. (1) Assuming that shunted blood flow equals 50% of the cardiac output, mixture of shunted blood with an oxygen content equal to 15 vol% with oxygenated pulmonary capillary blood equal to 20 vol% yields arterial blood with an oxygen content of 17.5 vol%. (b) If the shunt fraction remains 50% and the mixed venous oxygen content decreases 10 vol%, the arterial oxygen content decreases to 15 vol%.

ventilation becomes more uneven. In such situations, the gas composition at the end of an exhalation is composed not only of alveolar gas, but also of gas from poorly perfused or nonperfused areas of the lung (dead space). Therefore, in this situation the $p_{\text{et}}\text{CO}_2$ will be lower than the $p_{\text{a}}\text{CO}_2$. However, even in situations where lung disease has caused the $p_{\text{et}}\text{CO}_2$ to be lower than the $p_{\text{a}}\text{CO}_2$, observation of trends in $p_{\text{et}}\text{CO}_2$ may be useful for suggesting similar changes in $p_{\text{a}}\text{CO}_2$. For example, a slow increase in $p_{\text{et}}\text{CO}_2$ in a patient with an acute pulmonary illness who is breathing spontaneously would suggest that the patient was beginning to hypoventilate. In addition, an acute increase in the $p_{\text{a}}\text{CO}_2 - p_{\text{et}}\text{CO}_2$ difference is indicative of an increase in dead space ventilation such as what would be observed with a pulmonary embolism. However, the most common use of measuring the $p_{\text{et}}\text{CO}_2$ is to assist in the recognition of an esophageal intubation. In contrast to intubation of the tracheal, inadvertent esophageal intubations are associated with negligible amounts of CO_2 .

Estimates of the efficiency of CO_2 gas exchange can be obtained by measurements of expired minute ventilation (V_{E}) and dead space ventilation to tidal volume ratio ($V_{\text{D}}/V_{\text{T}}$). In individuals without lung disease, a normal $p_{\text{a}}\text{CO}_2$ can be maintained with a minute ventilation of $\sim 5\text{--}6$ L/min. Situations where a higher minute ventilation is required to maintain a normal $p_{\text{a}}\text{CO}_2$, suggests either that there is increased ventilation to nonperfused areas of the lung (dead space) or that the production of CO_2 (V_{CO_2}) by the body has increased. The latter can often be excluded on clinical grounds (i.e., absence of fever, sepsis, hyperthyroidism), but both the $V_{\text{D}}/V_{\text{T}}$ and the V_{CO_2} can be measured. The $V_{\text{D}}/V_{\text{T}}$ can be measured using the Enghoff modification of the Bohr equation:

$$V_{\text{D}}/V_{\text{T}} = p_{\text{a}}\text{CO}_2 - p_{\text{E}}\text{CO}_2 / p_{\text{a}}\text{CO}_2$$

where $p_{\text{E}}\text{CO}_2$ represents the $p\text{CO}_2$ in a sample of mixed expired gas. The V_{CO_2} is determined by measuring the fraction of CO_2 contained in a timed collection of expired gas. Monitoring of V_{E} , $V_{\text{D}}/V_{\text{T}}$, and V_{CO_2} occasionally are helpful in determining the efficiency of CO_2 gas exchange in patients where the etiology of their high minute ventilation requirements is unclear.

METHODS AND DEVICES USED IN VENTILATORY MONITORING

Measurement of Breathing Frequency, Airflow, and Lung Volumes

Aside from visual observation, there are several automated methods of continuously monitoring breathing frequency and effort (Table 2). Since patients in the intensive care unit are universally monitored for cardiac rhythm, one of the most conveniently employed techniques is impedance pneumography. This method is based on the principle that electrical impedance across the chest wall varies with inspiration and expiration (17). Chest wall impedance is measured by passing a constant, sinusoidal, low intensity current between two electrodes affixed to the chest wall. Usually, impedance devices are incorporated into electrocardiogram electrodes, thus permitting the simultaneous

Table 2. Measurement of Breathing Frequency, Airflow and Lung Volumes

Breathing Frequency or Effort
Impedance pneumography
Pleural pressure measurements
Strain gages
Magnetometers
Intercostal and diaphragmatic electromyography
Inductance pneumography
Airflow
Thermistor flowmeters
Nasal pressure measurements
Fleish pneumotachograph
Variable orifice pneumotachograph
Ultrasonic flowmeters
Gas ionization
Oscillating vane spirometer
Mechanical spirometers
Infrared CO_2 absorption
Mass spectrometry

monitoring of cardiac rhythm and respiratory effort. Since the relationship between changes in impedance and lung volume are linear, it is possible to obtain measurements of lung volume. However, because the relationship between impedance and volume varies between patients (2.8–4.6 Ω/L), a different calibration must be obtained for each patient. While accurate measurement of lung volumes has been obtained by some investigators (18), others have observed poor reproducibility (17), which may be a result of variability induced by the changing contribution of intercostal and diaphragmatic contraction with alterations in body position. Therefore, most impedance devices are used only qualitatively to assess the frequency and depth of respiratory effort. In addition to their use in an intensive care setting, impedance pneumograms are used as monitors for sleep apnea studies (polysomnograms), and sudden infant death syndrome diagnosis and monitoring (pneumocardiograms and apnea monitors). Since impedance pneumography only measures respiratory effort, however, a major drawback is its inability to distinguish between normal respiratory effort and respiratory effort associated with airway obstruction.

A quantitative indicator of respiratory effort can be obtained from measurements of pleural pressure. Usually, this is performed indirectly by measurement of esophageal pressure with an esophageal balloon and a pressure transducer. However, esophageal pressure measurements also have been obtained using fluid filled catheters (19). With the use of a properly sized and inflated esophageal balloon or fluid catheter placed in the lower third of the esophagus and with the patient in the sitting or lateral decubitus position, esophageal pressures are accurate reflections of pleural pressure. Measurements recorded in the supine position are inaccurate because of artifacts produced by the weight of the mediastinum. Although a commercial esophageal balloon incorporated into a standard nasogastric tube is available (20), recording of accurate pressures is technically difficult in a research environment (7), and even more so in a clinical setting. Nevertheless, esophageal pressure measurements are continuously recorded as an

index of respiratory effort in some sleep disorders laboratories. Otherwise, however, they are not frequently used clinically for continuous monitoring. Direct measurement of pleural pressure using a small catheter placed into the pleural space has also been reported (21), but is not commonly performed.

Respiratory effort also can be monitored using strain gauges (2) and magnetometers (22) placed around the chest and abdomen. Magnetometers measure the change in magnetic fields, produced by electrical coils placed on opposite sides of the body. If expansion of the lungs can be considered as a volume change with only 2° of freedom, the rib cage being 1° of freedom, and the abdomen the other, then lung volume changes are nearly linearly related to changes in the anteroposterior diameters of the chest wall and abdomen. Although the accuracy of magnetometers is limited for the same reasons mentioned for impedance devices, their reliability may be increased if additional magnetometers are used to measure lateral chest wall movement. However, for clinical purposes, both strain gauges and magnetometers are not usually used for quantification of tidal volume. Rather, these methods are used primarily in sleep disorders laboratories for a qualitative estimate of respiratory effort.

Intercostal EMG measurements can be obtained by placement of surface electrodes in the intercostal spaces in the anterior axillary line, and can be employed as an indicator of respiratory effort (2). However, when used in diagnostic studies for sleep disordered breathing, intercostal EMG activity may not be representative of respiratory effort during rapid eye movement sleep because of the inhibition of skeletal muscle tone during this stage of sleep. Frequency analysis of the intercostals and diaphragm EMG has been used by some investigators to study fatigue of the respiratory muscles (23). Although EMG evidence of respiratory muscle fatigue can be detected before clinical findings of acute respiratory failure, routine application of processed respiratory muscle EMG data to patient care has not occurred.

A more reliable method of quantifying breathing frequency and airflow is inductance pneumography. Similar to magnetometers, this technique considers lung volume to be a result of alterations in two compartments: the rib cage and the abdomen. Therefore, the volume of each tidal breath is theoretically attributable to the independent movement of both these compartments. Hence, an inspired volume will be equal to the sum of the volume changes from both compartments. To measure the volume changes in each compartment, bands of insulated wire coils are placed around the chest and abdomen and held in place with a mesh vest. With appropriate calibration, changes in inductance of the coils are accurate reflections of both alterations in lung volume, and the relative contributions of the chest and abdomen to the volume changes (24). Clinical usage of inductance pneumography has been limited primarily to ventilatory monitoring for polysomnography.

There are several categories of devices available that primarily measure airflow (Table 2). With many of these devices, flow rate can be integrated over time to derive a measurement of V_t or VC. Thermistor flow meters are commonly used in sleep disorders laboratories, and in

portable ventilation monitors. These devices are based on the principle that the resistance of a thermistor varies with temperature (25). Using a Wheatstone bridge, changes in resistance can be easily measured. The determinants of heat transfer between the thermistor and a gas are gas density and flow rate, and the temperature difference between the thermistor and the gas. In general, an increase in gas flow cools the thermistor and increases thermistor resistance, whereas a decrease in gas flow has the opposite effect. By calibrating the resistance changes to flow rate, quantitative measurements of flow and volume can be obtained. Unfortunately, the changes in resistance with airflow are nonlinear, and correlation with directly measured airflow is relatively poor. Nevertheless, less complex devices are frequently used in sleep disorders laboratories to qualitatively assess airflow during polysomnography.

Flow rate can be calculated by measuring the pressure difference between two sequential points. The most commonly used device employing this principle is the Fleish pneumotachograph (25). With this device, airflow is directed through a mesh screen having little resistance. The pressure drop across the screen is measured using a sensitive differential pressure transducer. Over the flow range for which the device has been designed, flow rate varies with the pressure changes. Inaccuracies result from condensation of water vapor in the device, but can be prevented by heating the pneumotachograph. The Fleish pneumotachograph is primarily used in research applications although some mobile pulmonary function units measure airflow using this device. The pressure difference across a variable orifice also can be used to measure airflow (25). With this device, flow is directed past an elastic flap that acts as a variable orifice. As flow rate increases, the flap opens larger. Similar to the Fleish pneumotachograph, the pressure drop across the orifice is proportional to airflow.

Ultrasonic principles can be employed to monitor ventilation in several ways (26). One method utilizes the principle of vortex shedding. Devices using this principle direct airflow past precisely sized struts creating vortices within the gas stream. An ultrasonic transducer directs sound waves through the air stream to a receiver. These sound waves are interrupted by the formation of vortices within the gas stream. The rate of vortex formation is proportional to airflow. Another method is based on the transit time of an ultrasonic burst between two crystals placed at an angle to the direction of gas flow. Frequencies between 3 and 10 MHz are transmitted, and the transit time can be related to airflow. Advantages of ultrasonic flowmeters are that no moving parts are present, and that there are no problems with condensation in the system.

Ionization of gases by a radioisotope (americium 241) has been proposed as a method of measuring airflow (27). With this technique, gases are ionized by alpha particles emitted by the radioisotope. The alpha particles displace electrons from some of the gas molecules, leaving them positively charged. A downstream electrode measures the amount of ionized particles reaching it. Since the ionized gas particles quickly neutralize themselves by recombination, the number of ionized gas particles reaching the

downstream collector is proportional to the gas flow rate. Similar to ultrasonic devices, potential advantages of this technique are the absence of moving parts, and the lack of interference from condensation in the system.

Interruption of a light beam by a small flapping vane in the gas stream has been used to measure airflow. With this device gas flow oscillates a flapping vane in the gas stream (25). Oscillation of the vane interrupts a light beam from a photoelectric cell. As air flow increases, the frequency of oscillation of the vane and, consequently, the frequency of interruption of the light beam increase. A disadvantage of this type of device is that high flow rates and condensation can cause malfunction.

Airflow can be measured by a variety of mechanical type devices (25,26). The most common example of this category used at the bedside is the Wright Respirometer. This device, and others that are quite similar, direct gas flow through rotating vanes. The vanes, which spin with the flow of gas, are connected through a series of gears to a calibrated dial. The dials on these devices usually are calibrated to measure V_t , VC, and V, but not flow rate. An exception is the Wright Peak Flow Meter, which also uses the principle of gas hitting a rotating vane. However, this device is calibrated to measure flow rate and has a brake that keeps the needle indicator on the highest flow rate obtained until the device is reset. The main disadvantage of these devices is that undermeasurement of airflow occurs at flow rates $<3\text{--}4$ L/min and that the units can be damaged if flow rates are >300 L/min. Nevertheless, their compact size makes them attractive to use.

An indirect estimate of airflow through the nose can be obtained by measuring changes in nasal pressure using a simple oxygen cannula and a pressure transducer. Changes in nasal pressure correlate well with airflow. In addition, flattening of the peak of the nasal pressure tracing during inspiration is indicative of elevations in upper airway resistance. However, it is susceptible to artifact due to mouth opening that results in signal loss. This technique is increasingly used during polysomnography.

A qualitative estimate of breathing frequency and airflow can be obtained by monitoring expired CO_2 . Since inspiratory gas has negligible amounts of CO_2 , the expiratory phase of breathing is marked by the appearance of measurable amounts of CO_2 in the exhalate. Periods during the ventilatory cycle where no CO_2 is detected indicate the presence of apnea. Carbon dioxide in respiratory gases is usually monitored with an infrared (IR) absorption system or a mass spectrometer (see below). The most common application of the qualitative monitoring of expired CO_2 is in sleep disorders laboratories where the absence of expired CO_2 is a marker for apneic episodes.

Measurement of Oxygen in Blood

Development of the polarographic $p\text{O}_2$ electrode by Dr. Leland Clark in 1953 revolutionized the diagnosis and treatment of patients with impairments in oxygen gas exchange. Today, measurement of arterial oxygen tension is nearly as common place as determination of the hematocrit and can be performed using a modern blood gas analyzer within a minute. The standard $p\text{O}_2$ electrode

consists of a platinum cathode sealed in a glass tip and a silver or silver-silver chloride anode. Electrical contact between the anode and the cathode is made by placing them both in a potassium chloride and hydrogen phosphate buffer solution. The blood sample to be analyzed is separated from the electrodes by a gas permeable membrane (polypropylene, Teflon, or Mylar). A constant polarizing voltage is applied between the anode and cathode. Oxygen in the blood sample diffuses across the membrane and combines at the cathode with water and electrons in the electrolyte to form hydroxide ions. The hydroxide ions then are attracted to the anode where electrons are transferred to form silver oxide and water. The current represented by the transfer of electrons at the anode is proportional to the oxygen tension in blood (28).

Blood samples for measurement of oxygen tension usually are obtained from indwelling intravascular catheters or percutaneous needle puncture of a vessel. However, an estimate of $p_a\text{O}_2$ sometimes can be obtained from capillary samples. Warming the skin to induce vasodilation in skin capillaries will cause the $p\text{O}_2$ in these capillaries to approach arterial levels, so-called arterialization. A small amount of this arterialized blood can be obtained by performing a skin prick in the area that was heated, and collecting the blood with a capillary tube. The sample then can be analyzed in a blood gas analyzer designed to accommodate small samples. In infants, such samples frequently are obtained from the heel ('heel stick blood gases'). However, blood gas tensions obtained in this manner are less accurate than direct arterial sampling, and may be inaccurate in situations where skin perfusion is poor (29).

Although the intermittent sampling of arterial and mixed venous blood for oxygen tension is now considered an essential element in the care of critically ill patients in intensive care units, continuous monitoring may be more desirable in very unstable patients. In the late 1970s, a device incorporating a sampling catheter attached to a portable gas chromatograph was introduced into clinical use (30). With this device, a 0.7 mm external diameter probe was inserted through a standard 18-gauge arterial pressure monitoring catheter. The probe consisted of a chamber made from heparin-bonded gas-permeable rubber that was connected through tubing to the gas chromatograph. Oxygen and carbon dioxide were allowed to diffuse from the blood into the probe, and then were carried by helium in the tubing to the chromatograph for analysis. Both $p\text{O}_2$ and $p\text{CO}_2$, which represented the average values of the previous 3.5 min, were displayed every 4 min. Although reasonably accurate, the device was not a commercial success in part because of the requirement that the probe be inserted into an 18-gauge catheter in comparison to the smaller and more commonly used 20-gauge catheter. Continuous $p\text{O}_2$ monitoring also can be performed using miniature Clark-type electrodes inserted intravascularly. One such device has an external diameter of 0.65 mm, making it suitable for insertion into a peripheral artery in adults (31). In addition, these devices have been used to continuously monitor $p\text{O}_2$ circulating externally in heart-lung oxygenators during cardiac bypass surgery. Although potentially attractive for monitoring patients with a very unstable respiratory status, continuous intravascular $p\text{O}_2$

monitoring has yet to find any substantial clinical acceptance.

A continuous approximation of oxygen tension in arterial blood also can be obtained by measuring transcutaneous oxygen tensions ($T_c pO_2$). Using a small polarographic oxygen electrode operating on a principle similar to that employed in a standard blood gas analyzer, the tension of oxygen diffusing through the skin can be measured. In infants, when the skin is heated to 40°C, the $T_c pO_2$ approximates the $p_a O_2$, and can be used as an indicator of arterial oxygenation. However, the response time of the electrode is relatively slow (95% response time 50–100 s). In adults, skin thickness is greater, and the amount of O_2 able to diffuse to the skin surface is less. Therefore, the $T_c pO_2$ in adults is significantly less than the $p_a O_2$ (32). In addition, $T_c pO_2$ measurements are affected by skin perfusion. When perfusion of the skin falls, as would occur with any condition producing a decrease in cardiac output, $T_c pO_2$ becomes perfusion dependent. In this situation, a decrease in cardiac output results in a decrease in $T_c pO_2$ that may not be related to a change in $p_a O_2$.

Oxygen tension measured from the conjunctiva using a small polarographic electrode has been shown to reflect the $p_a O_2$ in the absence of hemodynamic impairment. Conjunctival pO_2 values generally range from 50 to 75% of simultaneously measured $p_a O_2$ values. However, when blood flow to the eyelid is diminished, conjunctival pO_2 measurements are more accurate indicators of changes in peripheral perfusion than approximations of oxygen gas exchange (33). Although measurements of pO_2 in the conjunctiva and the skin appear to monitor similar physiological changes, there are several advantages of measuring the pO_2 of the conjunctiva instead of the skin. First, since the conjunctiva lacks a keratinized surface and has its capillary bed lying just below its surface, it is not necessary to supply extrinsic heating. Second, the capillaries supplying the conjunctiva are branches of the ophthalmic artery, which in turn is a branch of the internal carotid artery. Therefore, measurements of the conjunctival pO_2 may represent a method of indirectly monitoring cerebral perfusion and oxygenation. Last, variability in values resulting from using different skin sites is not a factor in conjunctival pO_2 measurements.

The oxygen saturation of hemoglobin in blood can be calculated from the pO_2 , pH, and pCO_2 of the blood sample and the body temperature using a standard nomogram (34), or directly measured using an oximeter. However, use of calculated values is strongly discouraged since they do not account for such factors as the level of carboxyhemoglobin or shifts in the oxyhemoglobin dissociation curve resulting from changes in 2,3-diphosphoglycerate. The latter is a phosphate compound that can shift the oxyhemoglobin dissociation curve independently of pH, pCO_2 , and temperature. Oximeters measure oxygen saturation by applying observations made by both Lavoisier and Priestly in 1774 that the color of blood and atmospheric oxygen were related. Subsequently, techniques were developed that allowed determination of the concentrations of different species of hemoglobin by measuring differences in their absorption or reflection of light to

different wavelengths. Current blood oximeters pass at least two wavelengths of light through the blood sample to a photodetector whose output is used to calculate absorbances. The oxygen saturation can then be determined by the relative absorbances of each wavelength. Usually, when measuring the relative amounts of saturated or oxyhemoglobin and unsaturated or deoxyhemoglobin, wavelengths of light are chosen so that, with at least one wavelength, the absorbance difference between the two is maximum, and with the other wavelength, the difference approaches zero.

In addition to performing oximetry directly on blood samples, indwelling fiberoptic probes are now available that permit continuous monitoring of the oxygen saturation in mixed venous and umbilical artery blood. In the case of continuous monitoring of mixed venous oxygen saturation, the fiberoptic probe is incorporated into a flow-directed pulmonary artery catheter. Therefore, pulmonary artery pressures and cardiac outputs can be obtained simultaneously with mixed venous oxygen saturation. When compared to oxygen saturation measurements obtained with intermittent blood sampling, these devices have been shown to measure oxygen saturation quite accurately (35). Although continuous monitoring of arterial saturation with such devices is possible, there has not been a large demand for this clinical application since noninvasive methods of measuring arterial oxygen saturation are available (see below), and because there does not appear to be notable clinical utility for measuring mixed venous oxygen saturations continuously. The principle involved with continuous monitoring of oxygen saturation using fiberoptic probes is reflection oximetry. A fiberoptic bundle transmits light of several different wavelengths down the bundle to the blood where light reflected by hemoglobin is transmitted back to a photodetector through a separate fiberoptic bundle. The amount of light reflected by the different wavelengths then can be converted into hemoglobin saturation.

Oxygen saturation can be continuously monitored noninvasively *in vivo* using principles similar to that employed with blood oximeters. The earliest of these devices (Hewlett Packard Ear Oximeter) measures the oxygen saturation of blood flowing through the ear lobe. The technique involves heating the ear to arterialize the blood flow and measuring the optical transmittance of light passed through the ear at 8 wavelengths in the 650–1050 nm range. Oxygen saturation can be computed using a model based on the Beer-Lambert law (36), which states that the absorbance for any wavelength of light is a function of the layers, concentration, and thickness of absorbers. Using the model and converting absorbance to transmittance allows for calculation of the oxygen saturation. The major clinical disadvantage of this device was its bulkiness, and the requirement that it be placed on an ear lobe. More recently, a variety of pulse oximeters have been introduced into clinical use (37). With pulse oximeters, any pulsating arterial vascular bed such as a fingertip, is placed between a light source transmitting different wavelengths and a detector. Expansion and relaxation of the vascular bed by the arterial pulsations alter the length of the light path. The amount of the light detected therefore varies, and this results in the

production of a plethysmographic waveform. The amplitude of the waveform is a function of the wavelengths of light used, the hemoglobin saturation, and the size of the pulse changes. Since the wavelengths of light used are known, beat-to-beat arterial oxygen saturation can be calculated using the magnitude of the pulse changes and a mathematical model based on the Beer–Lambert law. Current pulse oximeters are quite compact, and can be used for ambulatory and home monitoring. In comparison to ear oximeters, they also provide continuous monitoring of the arterial pulse rate. Their use in clinical medicine has proliferated exponentially. The major disadvantage to both ear and pulse oximeters is the inability to measure arterial saturation in low cardiac output states when tissue perfusion is impaired.

The oxygen content of blood can be calculated from the hemoglobin concentration, the saturation of hemoglobin with oxygen, and the pO_2 in the blood sample, as previously described for the arterial oxygen content. In most situations, calculated oxygen contents are used when determination of oxygen contents is required. However, devices and approaches have been developed to directly measure oxygen content. In one approach that is used commercially, the hemoglobin is lysed to release all of the oxygen in the sample into solution. The amount of oxygen is then analyzed using one of a variety of techniques (see below) More cumbersome and difficult is to measure oxygen content by carbon monoxide displacement, or by volume extraction and manometric measurement. The first technique involves displacement of oxygen in the sample to be analyzed by addition of carbon monoxide, and subsequently measuring the increase in pO_2 in the resultant mixture (38). The second technique, described by Van Slyke and Neill (39), involves extraction of the gas from solution with various reagents over mercury, creating a Torricellian vacuum, and measuring pressure changes with a manometer. Both of these techniques are seldomly used today and are primarily of historical interest.

Measurement of Oxygen in Gas

Several techniques are available to measure the concentration of oxygen in gas:

- Paramagnetic analyzers
- Electrical analyzers
- Galvanic fuel cell
- Polarographic gas analyzers
- Ionized oxygen electrode
- Scholander volumetric technique
- Gas chromatography
- Mass spectrometry

Paramagnetic analyzers use the principle that oxygen alters a magnetic field when introduced into it (40). With one such device, oxygen introduced into the magnetic field causes a glass dumbbell, filled with nitrogen and suspended in the field, to rotate. The amount of rotation of the dumbbell is proportional to the percentage of oxygen concentration. Since the alteration in the magnetic field is

a response to the amount of oxygen present in the sample gas, a paramagnetic analyzer more accurately reflects the partial pressure of oxygen rather than the oxygen percentage.

Electrical oxygen analyzers compare the resistance of a wire in a reference gas to that of a wire in contact with the sample gas using an electrical current that is proportional to the number of oxygen molecules in the Wheatstone bridge apparatus (40). As the oxygen concentration in the sample gas increases, the resistance of the wire decreases, allowing more current to flow through the wire in contact with the sample gas in the Wheatstone bridge. This phenomenon occurs because oxygen, having a greater mass than nitrogen, cools the wire and decreases its resistance. These devices are sensitive to the presence of gases, other than oxygen and nitrogen, that have a high mass. For example, significant concentrations of carbon dioxide will result in a falsely high oxygen percentage. Since it is the resistances between a sample gas and a reference gas that are being compared, electrical analyzers more accurately reflect oxygen percentage rather than oxygen partial pressure.

Another commonly used type of oxygen analyzer is the galvanic fuel cell (26,40). With this type of device, a semipermeable membrane separates the gas sample from an hydroxide bath. The bath consists of a lead anode and a gold cathode. Oxygen diffuses across the membrane where electrons from the gold electrode and water in the bath combine with the oxygen to form hydroxyl ions. The hydroxyl ions are then attracted to the anode where another chemical reaction occurs that yields lead oxide, water, and electrons. The released electrons produce an electric current that is proportional to the concentration of oxygen in the sample gas. Fuel cell devices sense the amount of oxygen molecules in a sample gas, and therefore most accurately sense oxygen partial pressure. Their readings are adversely affected by excessive moisture, gas under high pressure such as with PEEP, and altitude.

The polarographic principle used in the pO_2 electrode of a blood gas analyzer also can be applied to measurement of oxygen in a gas (40). In fact, many blood gas analyzers will allow processing of gas samples, thus providing an alternative method of measuring the pO_2 in a small gas sample. Factors adversely affecting polarographic units are the same as those for fuel cell devices. In addition, polarographic units most accurately reflect partial pressure. Although both fuel cell and polarographic-type devices utilize positive and negative electrodes and measure the current generated by an electrochemical reaction with oxygen, a faster response time is observed with polarographic units because their electrodes are polarized with a battery. However, the electrodes with galvanic units may have greater longevity.

Another method of measuring the concentration of oxygen in a gas is the use of an ionized oxygen electrode. With this technique, a thin coating of platinum is layered on both sides of a zirconium oxide ceramic tube that is permeable to oxygen. When the tube is heated to a high temperature, O_2 is ionized to O_2^- . The zirconium oxide tube becomes selectively permeable to these ions, and an electrical potential is

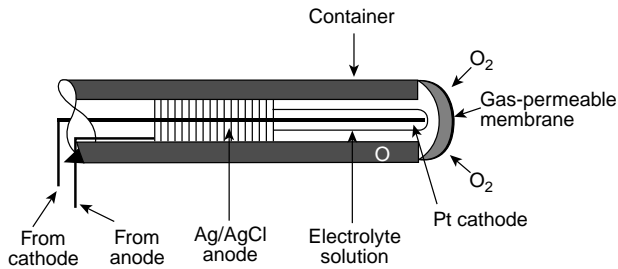


Figure 4. The pO_2 electrode. Sealed platinum cathode and silver-silver chloride anode are bathed in a potassium chloride and hydrogen phosphate electrolyte buffer solution.

generated between the two platinum surfaces that is proportional to the oxygen concentration outside the tube (4).

The most cumbersome method of measuring oxygen concentration in a gas is the Scholander volumetric technique (41). With this technique, volume changes in the sample gas are measured as the various gaseous components are absorbed from the sample. This method is of little clinical utility for continuous monitoring of oxygen concentrations, but is an important reference technique against which more portable units can be compared.

Another, uncommonly used, gas analytical technique in clinical medicine is gas chromatography. With this method, the gas sample to be analyzed is added to the flow of a carrier gas. The carrier gas (usually helium) transports the sample through a column that contains material having a differential affinity for the components of the sample. The column impedes the passage of components with higher affinities, and therefore separation of the components occurs. After passage through the column, the components of the gas sample can be measured using a nonspecific sensor. Usually, this is done by detecting changes in thermal conductivity induced by the components of the gas sample, but other types of sensors also can be employed (42). Gas chromatography is an accurate method for analyzing O_2 , CO_2 , and anesthetic gas concentrations. However, the slow response time (minutes) is a major drawback. Nevertheless, it has been used for *in vivo* measurements of arterial pO_2 (see the section Measurement of Oxygen in Blood).

The most expensive, but also most versatile technique to analyze oxygen and other gases is mass spectrometry (MS) (4,26). With MS, the sample gas is passed through an electron gun that ionizes the gas particles. The ionized gas particles are then drawn into a magnetic field, where the particles are deflected toward a collection plate. The larger the mass of the gas particle, the farther it will travel down the collection plate. The collection plates count the number of particles of each molecular mass impacting on the plate over time. From this information, the relative composition of the sample gas can be computed. Mass spectrometers are quite accurate, but very expensive. However, one unit can be used to sample gas from multiple sources in a similar way to the method a mainframe computer uses to perform multiple tasks. Such arrangements are used in some operating rooms and intensive care units.

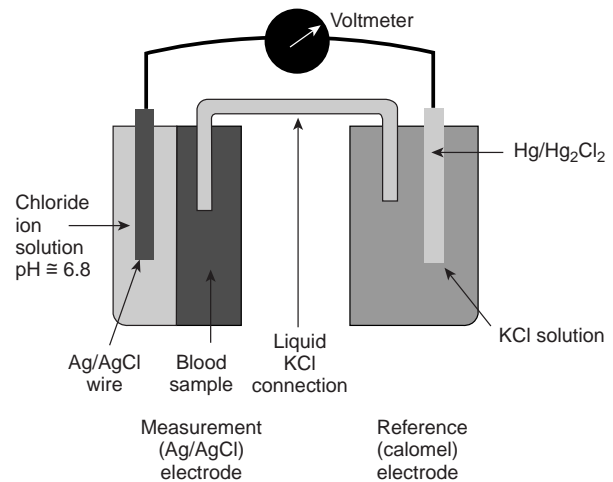


Figure 5. Schematic of the pH electrode. The pH electrode consists of a measurement (silver-silver chloride) electrode and a reference (calomel) electrode.

Measurement of pH and Carbon Dioxide in Blood

Measurement of the pH of blood or another body fluid such as pleural fluid is performed using a pH electrode (Fig. 5). With all pH electrodes, a sample of unknown pH, such as one containing arterial blood, is placed on one side of a pH sensitive glass membrane and a reference pH solution is in contact with the other side. A potential difference develops that is proportional to the pH difference of the two solutions. Since the pH of the reference solution is known, the pH of the sample can be calculated. The actual pH electrode consists of two separate electrodes—a pH-sensitive glass electrode and a reference electrode. Components of the pH sensitive electrode are a silver-silver chloride wire connected to a voltmeter surrounded by a chloride ion buffer with a pH of ~ 6.8 , and a pH-sensitive glass membrane. The reference electrode is constructed so that a platinum wire attached to a voltmeter is in contact with a mercury chloride paste. A 20% potassium chloride solution is used as a wetting agent to insure that the platinum wire maintains contact with the paste. The pH and the reference electrodes are placed in proximity so that the sample is introduced on one side of the pH-sensitive membrane with the buffer of known pH on the other side.

The pCO_2 electrode is actually a modification of the pH electrode. With the pCO_2 electrode, a pH-sensitive glass electrode and a reference pCO_2 electrode are in contact with an electrolyte solution behind a gas-permeable membrane. When a sample is placed on the opposite side of this membrane, CO_2 diffuses across the membrane in both directions and equilibrates with the electrolyte solution. Hydration of the CO_2 results in the production of carbonic acid and a consequent increase in hydrogen ion activity. The pH electrode detects the change in hydrogen ion concentration and develops a voltage change that is proportional to the pCO_2 in the sample.

Whereas reliable methods are available to continuously monitor oxygenation in blood, there are few techniques that are currently used to measure pCO_2 .

Intravascular $p\text{CO}_2$ has been measured using a portable gas chromatograph as previously described (30). However, an intravascular sensor using optode technology (see below) to measure $p\text{CO}_2$ and pH, and a miniaturized intravascular Clark $p\text{O}_2$ has been successfully used in humans (43). Devices to measure transcutaneous $p\text{CO}_2$ ($T_c p\text{CO}_2$) are commercially available, and employ a small $p\text{CO}_2$ electrode similar to that used in a standard blood gas analyzer (44). In general, $T_c p\text{CO}_2$ values are higher than corresponding $p_a\text{CO}_2$ values. Similar to $T_c p\text{O}_2$ measurements, the $T_c p\text{CO}_2$ is adversely affected by changes in skin perfusion, and the response time of the electrode is slow (63% response time: 2 min) (31).

Measurement of Carbon Dioxide in Gas

For monitoring purposes, determination of the concentration of CO_2 in a gas is necessary for calculations of V_D/V_T , $V\text{CO}_2$, and measurement of $p_{\text{et}}\text{CO}_2$. In addition, it is required in pulmonary function laboratories for performance of CO_2 and O_2 response testing. Mass spectrometry (see the section Measurement of Oxygen in Gas) is a rapid and accurate method of measuring the concentration of CO_2 in a gas, and also is adaptable to measuring gas samples from multiple sources at a central location. However, mass spectrometers are very expensive and not generally portable. The major alternative method of measuring the concentration of CO_2 in a gas is by IR absorption. This technique is based on the principle that all gases whose molecules contain either more than two atoms (e.g., CO_2 or SO_2) or two atoms of different elements (e.g., CO) will absorb IR radiation. With an IR CO_2 analyzer, IR radiation is directed through the gas sample to be analyzed where the sample absorbs a small amount of energy. After passing through the sample, the remainder of the IR radiation is directed into another chamber containing pure CO_2 where the rest of the energy is absorbed. Absorption of IR energy generates heat that can be easily quantified. As the concentration of CO_2 in the gas sample increases, less IR radiation will reach the second chamber that contains pure CO_2 . Therefore, the heating effect measured in this chamber will be inversely proportional to the concentration of CO_2 in the gas sample (45). Unfortunately, if a gas sample contains two different gases that can absorb IR energy, the analyzer will be unable to distinguish between the concentrations of the two gases. This problem occasionally arises during monitoring of $p_{\text{et}}\text{CO}_2$ during nitrous oxide anesthesia. However, placing a chamber filled with nitrous oxide between the IR radiation source and the gas sample to be analyzed will filter out any IR absorption secondary to nitrous oxide.

Practical and Technical Considerations in Blood Gas Analysis

Currently manufactured blood gas analyzers incorporate $p\text{O}_2$, $p\text{CO}_2$, and pH electrodes into a single blood gas analyzer unit so that a blood sample can be analyzed simultaneously for all three parameters. There are several practical considerations in the collection of blood samples for blood gas analysis. First, care should be exercised in the collection of the sample to eliminate air bubbles and not use excessive amounts of anticoagulant. Heparin, a com-

monly used anticoagulant, is acidic. Excessive amounts therefore will decrease the pH and increase the $p\text{CO}_2$ artifactually. Large air bubbles will produce a decrease in $p\text{CO}_2$ as CO_2 diffuses into the bubble. Oxygen tension will either increase or decrease depending on whether the $p\text{O}_2$ of the sample is higher or lower than the $p\text{O}_2$ of the air bubble. Second, plastic syringes have been observed to perform equally as well as glass syringes provided that the analysis is undertaken within an hour of obtaining the sample. Third, samples should be placed in ice for transport immediately after being obtained to minimize metabolic consumption of oxygen by white blood cells. This may be a cause of spurious hypoxemia in patients with leukemia, who may have pathologically elevated white blood cell counts. Fourth, the patient's temperature should be noted so that correction for the effect of temperature on the relative solubility of oxygen in blood can be made. Several technical factors may introduce error into blood gas analysis. Since oxygen electrodes generally are calibrated using a gas with a $p\text{O}_2$ between 130 and 150 mmHg, measurement of a sample with a higher $p\text{O}_2$ may be in error. In such cases, the electrode should be recalibrated with tonometered blood containing a high $p\text{O}_2$ prior to analysis. In addition, the sample chamber should be flushed with 100% oxygen to leach out any residual wash fluid containing a low oxygen tension. Most blood gas analyzers perform their measurements at a temperature of 37°C . Variation of the temperature in the sample chamber will alter pH by 0.015 units, $p\text{O}_2$ by 7%, and $p\text{CO}_2$ by 4.5% for each degree of temperature change.

Alarms. Alarms are used during ventilatory monitoring to either signal the occurrence of life-threatening conditions, or indicate the presence of conditions that may not be immediately life-threatening, but are abnormal. Apnea and ventilator disconnection are the most common clinical situations where life-threatening alarms are used. Apnea is detected by monitoring parameters that indicate either absence of ventilatory effort or airflow. However, devices that only detect ventilatory effort such as impedance pneumographs may not be able to distinguish normal respiratory effort from respiratory effort during an obstructive apneic event. Therefore, detection of apnea by monitoring airflow is more sensitive. Ventilator disconnection can be monitored by either a low inspiratory-airway-pressure alarm or a low exhaled-volume alarm. On many current model ventilators, both parameters are monitored. In situations where alarms are used to signal life-threatening events, alarms should be designed such that it is not possible to silence them except for very short periods of time. Otherwise, alarms may be inadvertently left off, and the patient placed at risk.

Inspired oxygen concentration and high airway pressure are two other important parameters that are frequently monitored on a continuous basis. With respect to inspired oxygen concentration, it is important to differentiate between alarms that monitor the gas pressure delivered to a device and alarms that actually sense oxygen concentration. In the first situation, true oxygen concentration is not being measured and in the unlikely situation where an oxygen source becomes contaminated with

another gas, such an alarm would not sound. High pressure alarms on mechanical ventilators not only detect when airway pressure exceeds a certain level, but they also provide a feedback signal to terminate inspiration so that the alarm pressure is not exceeded. Alarms can be used on almost any device used for continuous ventilatory monitoring. However, too many alarms can be ill-advised. If alarms are constantly sounding, all alarms, including life-threatening ones, tend to be ignored.

Evolving Technology in Ventilatory Monitoring. *In vivo* continuous measurement of O₂, CO₂, pH, and other substances in the future may be clinically useful using optode technology (46,47). An optode is a fiberoptic light bundle coupled to a microcuvette where a chemical reaction can occur. For example, the substance to be analyzed diffuses through a semipermeable membrane into the cuvette whereupon a chemical reaction would be initiated. The chemical reaction would produce a change in the optical properties of the reagent in the cuvette that would be detected by light directed along the fiberoptic light bundle. Wavelength specific absorbance can be used to quantify p_aCO₂ and pH, and fluorescence is useful for p_aO₂, as well as p_aCO₂ and pH. Imprecision of the p_aO₂ sensor has led to a hybrid probe, with a miniaturized Clark pO₂ electrode combined with absorbance p_aCO₂ and pH sensors, as well as a thermocouple for temperature measurement (43). Clinical challenges for continuous on line measurement, utilizing an *in vivo* placement, primarily deal with inaccuracies of p_aO₂ measurement secondary to arterial wall contamination or limited blood flow because of thrombus formation and vasospasm. In addition, probe longevity and cost, as well as pulse pressure dampening over time, have been issues. Because of these problems, an on demand *ex vivo* placement in the arterial line has been tested and has sufficient accuracy, but does not have the advantage of continuous measurement (48).

The ion selective electrode (ISE) or the field effect transistor-based chemical sensor (ISFET) are other new methods of continuously monitoring pH and pCO₂ that may become feasible in the future (47,49). With both these devices, the ion to be analyzed is isolated from the blood using an ion selective membrane. An ISE consists of an electrode surrounded by a buffer solution that is separated from a biological fluid or tissue by an ion-selective membrane. The electrical potential at the measurement electrode is compared to that of a reference electrode, and the potential difference is proportional to the concentration of the ion being measured. The ISFET is an insulated gate field-effect transistor without a metal electrode at the gate. Measurement of the concentration of a specific ion is a function of modulation of current flow within the semiconductor induced by the ion in the surrounding biological tissue or fluid. However, similar to optodes, both the ISE and the ISFET are still in the early developmental stages. Current problems include long-term stability, ambient light and temperature sensitivity, and interference from extraneous ions. With both the ISE and ISFET, control thrombus formation on the probe remains a problem. In addition, an adequate O₂ sensor has not been developed.

Although novel methods of ventilatory monitoring are being developed or are currently under investigation, perhaps the greater challenge will not be to simply prove efficacy. Rather, given the economics of healthcare delivery, the cost-effectiveness of any new device or technique on patient care will need to be unequivocally demonstrated for it to be clinically adopted.

BIBLIOGRAPHY

- Burrows B, Knudson RJ, Quan SF, Kettel LJ. Respiratory Disorders—A Pathophysiologic Approach. Chicago, IL: Year Book Medical Publication; 1983. p 69–93.
- Kryger MH. Monitoring Respiratory and Cardiac Function. In: Kryger MH, Roth T, Dement WC, editors. Principles and Practice of Sleep Medicine. Philadelphia: W.B. Saunders; 1989. p 984–993.
- Suter PM, Fairley HB, Isenberg MD. Optimum end-expiratory airway pressure in patients with acute pulmonary failure. *New Engl J Med* 1975;242:284–289.
- Severinghaus JW. Monitoring anesthetic and respiratory gases. In: Blitt CD, editor. Monitoring in Anesthesia and Critical Care Medicine. New York: Churchill-Livingstone; 1985. p 265–290.
- Fallat RJ. Respiratory monitoring. *Clin Chest Med* 1982;3:181–194.
- Hyatt RE, Scanlon PD, Nakamura M. Interpretation of Pulmonary Function Tests. Philadelphia: Lippincott Williams and Wilkins; 1997. p 103–109.
- Agostoni E. Mechanics of the pleural space. *Physiol Rev* 1972;52:57–125.
- Comroe JH, Forster RE, Dubois AB, Briscoe WA, Carlsen E. The Lung. Chicago, IL: Year-Book Medical Publishers; 1962. p 178–190.
- Murray JF. The Normal Lung. Philadelphia, PA: Saunders; 1976. p 104–105.
- Bone RC. Thoracic pressure–volume curves in respiratory failure. *Crit Care Med* 1976;4:148–150.
- Matamis D, Lemaire F, Rieuf P. Augmentation de la capacité résiduelle fonctionnelle induite par la ventilation en pression positive expiratoire. Prediction par la courbe pression-volume thoracopulmonaire. *Ann Fr Anesth Reanim* 1984;3:199–204.
- Comroe JH, Forster RE, Dubois AB, Briscoe WA, Carlsen E. The Lung. Chicago, IL: Year Book Medical Publications; 1962. p 191–195.
- Henning RJ, Shubin H, Weil MH. The measurement of the work of breathing for clinical assessment of ventilator dependence. *Crit Care Med* 1977;5:264–268.
- Yang KL, Tobin MJ. A prospective study of indexes predicting the outcome of trials of weaning from mechanical ventilation. *New Engl J Med* 1991;324:1445–1450.
- Johnson DC, Kazemi H. Central control of ventilation in neuromuscular disease. *Clin Chest Med* 1994;15:607–617.
- Burrows B, Knudson RJ, Quan SF, Kettel LJ. Respiratory Disorders—A pathophysiologic Approach. Chicago, IL: Year Book Medical Publishers; 1983. p 61–65.
- Grenvik A, Ballou S, McGinley E, Millen JE, Cooley WL, Safar P. Impedance pneumography. *Chest* 1972;62:439–443.
- Baker LE, Hill DW. The use of electrical impedance techniques for the monitoring of respiratory pattern during anaesthesia. *Br J Anaesth* 1969;41:2–17.
- Karason S, Karlsen KL, Lundin S, Stenqvist O. A simplified method for separate measurements of lung and chest wall

- mechanics in ventilator-treated patients. *Acta Anaesthesiol Scand* 1999;43:308–315.
20. Leatherman NE. An improved balloon system for monitoring intraesophageal pressure in acutely ill patients. *Crit Care Med* 1978;6:189–192.
 21. Downs JB. A technique for direct measurement of intrapleural pressure. *Crit Care Med* 1976;4:207–210.
 22. Mead J, Peterson N, Brinby G, Mead J. Pulmonary ventilation measured from body surface movements. *Science* 1967;156:1383–1384.
 23. Cohen CA, Zaghelbaum G, Gross D, Roussos CH, Macklem PT. Clinical manifestations of inspiratory muscle fatigue. *Am J Med* 1982;73:308–316.
 24. Cohn M. Respiratory monitoring during sleep: Respiratory inductive plethysmography. In: Guilleminault C, editor. *Sleeping and Waking Disorders—Indications and Techniques*. Menlo Park, CA: Addison-Wesley; 1982. p 213–224.
 25. McPherson SP. *Respiratory Therapy Equipment*. St. Louis, MO: Mosby; 1981. p 206–214.
 26. Sergejev IP. Monitoring of respiratory function during anesthesia. *Int Anesthesiol Clin* 1981;19:31–59.
 27. Jeretin S, Martinez LR, Tang IP, Ito Y. pneumotachography by the ionization principle—a new approach. *Anesthesiology* 1971;35:218–223.
 28. Laver MB, Seifen A. Measurement of blood oxygen tension in anesthesia. *Anesthesiology* 1965;26:73–101.
 29. Siggard-Andersen O. Acid-base and blood gas parameters: Arterial or capillary blood? *Scand J Lab Invest* 1968;21:289–292.
 30. Richman KA, Jobs DR, Schwab AJ. Continuous in-vivo blood gas determination in man. *Anesthesiology* 1980;52:313–317.
 31. Ledingham IM, MacDonald AM, Douglas IHS. Monitoring of ventilation. In: Shoemaker WC, Thompson WL, editors. *Critical Care Medicine—State of the Art*. 1981. p E1–E52.
 32. Tremper KK, Shoemaker WC. Transcutaneous oxygen monitoring of critically ill adults, with and without low flow shock. *Crit Care Med* 1981;9:706–709.
 33. Isenberg S, Fink S, Shoemaker W. The eye as a peripheral sensor. *Ann Ophthalmol* 1984;16:1105–1108.
 34. Severinghaus JW. Blood gas calculator. *J Appl Physiol* 1966;21:1108–1116.
 35. Baele PL, McMichan JC, Marsh HM, Sill JC, Southorn PA. Continuous monitoring of mixed venous oxygen saturation in critically ill patients. *Anesth Analg (Cleveland)* 1982;61:513–517.
 36. Hill DW. *Physics Applied to Anaesthesia*. London: Butterworth; 1976. p 339–340.
 37. Yelderman M, New W. Evaluation of pulse oximetry. *Anesthesiology* 1983;59:349–352.
 38. Duke GS, Newhouse YMC. Micromethod for measuring blood oxygen content by determining oxygen tension after saturation with carbon monoxide. *Am Rev Respir Dis* 1974;110:814–816.
 39. Van Slyke DD, Neill JM. The determination of bases in blood and other solutions by vacuum extraction and manometric measurement. I. *J Biol Chem* 1924;61:521–573.
 40. McPherson SP. *Respiratory Therapy Equipment*. St. Louis, MO: Mosby; 1981. p 153–154.
 41. Scholander PF. Analyzer for accurate estimation of respiratory gases in one-half cubit centimeter samples. *J Biol Chem* 1947;167:235–250.
 42. Banner N, Olsen RJ. Basic knowledge of gas chromatography and gas chromatographic instruments. In: Ettre LS, Zlatkis A, editors. *The Practice of Gas Chromatography*. New York: Wiley-Interscience; 1967. p 6–13.
 43. Menzel M, Henze D, Soukup J, Engelbrecht K, Senderreck M, Clausen T, Radke J. Experiences with continuous intra-arterial blood gas monitoring. *Minerva Anesthesiol* 2001;67:325–331.
 44. Tremper KK, Shoemaker WC, Shippy CR, Nolan LS. Transcutaneous PCO₂ monitoring on adult patients in the ICU and the operating room. *Crit Care Med* 1981;9:752–755.
 45. Mapleson WW. Physical methods of gas analysis. *Br J Anaesth* 1962;34:631–636.
 46. Peterson JI, Vurek GG. Fiber-optic sensors for biomedical applications. *Science* 1984;224:123–127.
 47. Eberhart RC. Indwelling blood compatible chemical sensors. *Surg Clin North Am* 1985;65:1025–1040.
 48. Mahutte CK. On-line arterial blood gas analysis with optodes: Current status. *Clin Biochem* 1998;31:119–130.
 49. Cheung PW. Chemical sensors. In: Gravenstein JS, Newbower RS, Ream AI, Smith NT, editors. *Essential Noninvasive Monitoring in Anesthesia*. New York: Grune & Stratton; 1980. p 183–216.

See also MONITORING IN ANESTHESIA; PULMONARY PHYSIOLOGY; RESPIRATORY MECHANICS AND GAS EXCHANGE.

VESSELS, PROPERTIES OF. See ARTERIES, ELASTIC PROPERTIES OF.

VISION CORRECTION DEVICES. See CONTACT LENSES.

VISUAL FIELD TESTING

AMOL D. KULKARNI
University of Wisconsin Madison
Madison, Wisconsin

INTRODUCTION

The visual field is the total area where objects can be seen in the peripheral vision while the eye is focused on a central point. Visual field testing is a critical part of the eye examination and is mandatory for the detailed evaluation of unexplained visual loss. Of the various reasons for conducting a visual field examination, the most common disorder is glaucoma. In glaucoma, field testing is essential not only to establish the diagnosis, but also to follow-up and determine the effectiveness of treatment. There are various modalities available for visual field testing customized to particular eye disorders. This article provides a concise review of the various methods of visual field testing and their applications for clinical and research purposes.

METHODS OF VISUAL FIELDS TESTING

There are various ways of testing the visual field (1). It can be done in a simple manner in the clinic or may involve sophisticated equipments. The commonly used modalities are described below.

1. **Confrontation visual field exam:** It is a quick evaluation of the visual field done by a physician sitting directly in front of you. With one eye covered,

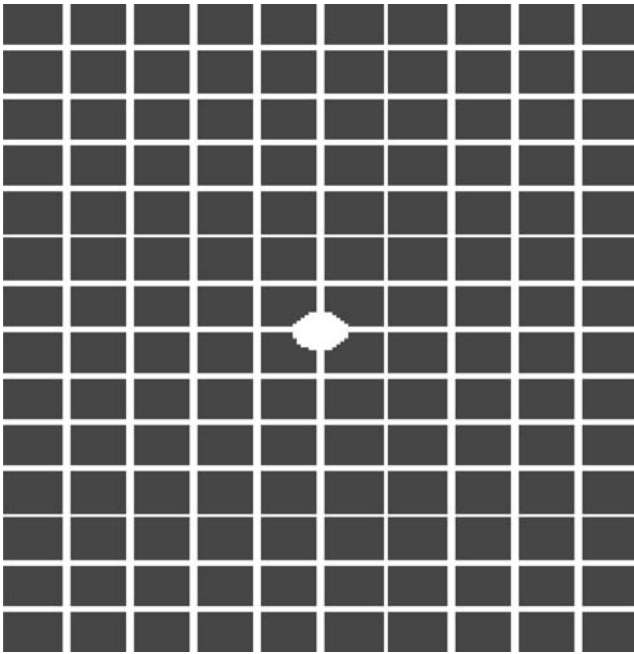


Figure 1. Amsler grid.

the patient is asked to look at the examiner's eye. The patient has to alert the examiner when they can see the examiner's hand in the various zones on the nasal, temporal, superior, and inferior aspect while still looking at the examiner's eye. This is a simple method of testing in the clinic and can detect gross visual field loss, such as hemifield loss.

2. **Amsler grid:** It is a quick method of self-assessment of the central 10° of the visual field. It consists of small squares of 1×1 mm with a central dot as shown in Fig. 1. After wearing the necessary correction, the Amsler grid is held at 14 in. (35.56 cm) from the eye and the central dot is focused. If any changes in the patterns of adjacent lines or squares are disrupted, a visual defect is suspected. It can detect small central or para-central visual defects called as scotoma especially in disorders of macula.
3. **Tangent screen exam:** The patient is asked to sit 3 ft. (91.44 cm) from a screen with a target in the center. While staring at the target, the patient alerts the examiner on visualizing an object brought into their peripheral vision. This helps in mapping the extent of peripheral vision.
4. **Perimetry:** It is a systematic measurement of the total area where objects can be seen in the peripheral vision while the eye is focused on a central point. The two most commonly used modalities of perimetry are Goldmann kinetic perimetry and threshold static automated perimetry.

Perimetry

There are two types of perimetry based on the stimulus characteristics (1). In kinetic perimetry, a stimulus is moved from a nonseeing area of the visual field to a seeing area along a set meridian. The procedure is repeated with

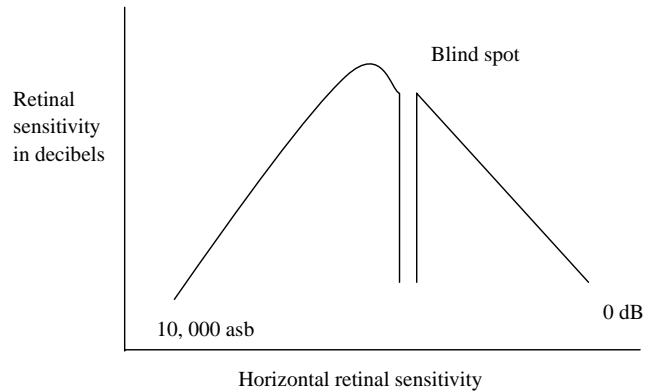


Figure 2. Normal hill of vision.

the use of the same stimulus along other meridians spaced at every 15° . In static perimetry, the size and location of the stimulus remain constant. The retinal sensitivity at a specific location is determined by varying the brightness of the stimulus. Based on the measurements from perimetry the normal visual field is defined as the area perceived by a fixating eye extending $\sim 60^\circ$ into the superior field, 75° into the inferior field, 110° temporally, and 60° nasally. This creates an island of vision with a steep central peak that corresponds to the fovea, the area of greatest retinal sensitivity. The shape of the island is defined by conducting measurement at various locations in the field of vision (Fig. 2).

Goldmann Perimetry

The Goldmann perimeter (1) is the most commonly used instrument for manual perimetry. It is a calibrated bowl projection instrument with a background intensity of 3 1.5 apostilbs (asb). The size and intensity of targets can be varied to plot the visual field kinetically and to determine local static thresholds. The stimuli used to plot a visual field are identified by a Roman numeral, a number, and a letter. The Roman numeral represents the size of the object, from Goldmann size 0 ($1/16$ mm²) to Goldmann size V (64 mm²). Each size increment equals a twofold increase in diameter and a fourfold increase in area. The visual field corresponding to each stimulus size is called as isopter. The Goldmann visual field testing is primarily used for determining field defects caused by disorders of the brain (neuroophthalmology). However, the reliability of field mapping depends on the expertise of the operator (perimetrist).

Automated Perimetry

The introduction of computers and automation lead to the development of a new generation of perimeters (1). These enabled static testing in an objective and standardized fashion with minimal perimetrist bias. A quantitative representation of the visual field can be obtained more easily than with manual testing. The computer allows stimuli to be presented in a pseudorandom, unpredictable fashion.

Patients do not know where the next stimulus will appear, so fixation is improved, thereby increasing the

reliability of the test. Random presentations also increase the speed with which perimetry can be performed, and thereby bypass the problem of local retinal adaptation. Static computerized perimetry measures retinal sensitivity at predetermined locations in the visual field. These perimeters measure the ability of the eye to detect a difference in contrast between a test target and the background luminance. The differential light threshold is designated as the dimmest target seen 50% of the time. Suprathreshold stimuli are brighter than threshold stimuli, and they will be seen > 50% of the time. Infrathreshold stimuli are dimmer than threshold stimuli, and they will be seen < 50% of the time. The various perimeters used in clinical practice include Humphrey's visual field analyzer, and Octopus perimeter.

Comparison of static and kinetic perimetry (2,3) reveals that kinetic evaluation can clearly outline the normal visual field. However, kinetic perimetry may miss shallow scotomas and poorly define the flat slope seen nasally. Static perimetry readily detects shallow scotomas and can define the slope of both shallow and steep scotomas.

GLAUCOMATOUS FIELD DEFECTS

Perimetry is of paramount significance in the management of glaucoma. In glaucoma, there is damage of the nerve fiber layer that causes loss of visual field. Automated perimetry is widely used and the commonly used programs for glaucoma are the Octopus program 32 and the Humphrey program 30-2. These programs are tests of the central 30° with 6° of separation between locations. The various field defects in glaucoma include diffuse depression, localized nerve fibre bundle defects, paracentral defects, arcuate scotomas, nasal step defects, temporal wedge-shaped defects, early visual field defects, and blind spot changes.

NEWER MODALITIES OF FIELD TESTING

Blue–yellow perimetry, also known as short wavelength automated perimetry (SWAP), represents a recent advance in the early identification of glaucomatous visual field loss. It differs from standard automated perimetry only in that blue light is used as the stimulus, and yellow light is used for the background illumination. Moreover new algorithm, such as SITA, which stands for Swedish Interactive Thresholding Algorithm (4) have been introduced to reduce the length of a visual field test while enhancing sensitivity and specificity.

Frequency doubling technology (FDT) is used to isolate a particular pathway of visual stimulus. It involves use of alternate bars of white and black and helps in early detection of glaucoma.

FUTURE DIRECTIONS

There have been considerable developments in visual field testing in the past decade, and it has opened up areas for the development of new testing and analysis programs. A

larger database for patients with glaucoma will improve the accuracy and detection of glaucomatous visual field defects. Additional clinical studies are needed to determine the role of these and other modalities in the future.

BIBLIOGRAPHY

Cited References

1. Johnson CA, Keltner JL. Principles and techniques of the examination of the visual sensory system. In: Newman NJ, Miller NR, editors. Walsh and Hoyt's Clinical neuro-ophthalmology. 5th ed. Baltimore (MD): Williams and Wilkins; 1997; p 194.
2. Trope GE, Britton R. A comparison of Goldmann and Humphrey automated perimetry in patients with glaucoma. *Br J Ophthalmol* 1987;71:489–493.
3. Tschopp C, et al. Automated static perimetry in the child: methodologic and practical problems. *Klin Monatsbl Augenheilkd* 1995;206:416–419.
4. Sekhar GC, et al. Sensitivity of Swedish interactive threshold algorithm compared with standard full threshold algorithm in Humphrey visual field testing. *Ophthalmology* 2000;107:1303–1308.

VISUAL PROSTHESES

JEAN DELBEKE
 CLAUDE VERAART
 Catholic University of Louvain
 Brussels, Belgium

INTRODUCTION

Minute electrical stimuli delivered to the retina, the optic nerve, or the occipital cortex can induce light perceptions called phosphenes. The visual prosthesis aims at exploiting these phosphenes to restore a form of vision in some cases of blindness. Very schematically, a camera or a picture capturing device transforms images into electrical signals that are then adapted and passed on to some still functional part of the visual pathways, thus bridging the defective structures. The system has at least some parts implanted, including electrodes and their stimulator circuits. A photo-sensitive array in the eye could provide the necessary image input, but most approaches use an external miniature camera. Typically, the visual data handling requires an external processor and the power supply as well as the data are provided to the implant by a transcutaneous transmission system.

Despite a first pioneering attempt by Brindley and Lewin as early as 1968 (1) only very few experimental visual prostheses have been implanted in humans so far. The limited accessibility of the involved anatomical structures, the poorly understood neural encoding, and the huge amount of information handled by the visual nervous system have clearly hampered a development that can not yet be compared with the far more advanced evolution of cochlear implants (see article on Cochlear Implants in this encyclopedia). The visual prosthesis is still at a very

early stage, exploring different methodological directions, and seeking minimal performances that would justify clinical applications in the most severe cases of blindness. The first fully fledged clinical study has yet to begin and the experimental character of existing systems limits all trials to a restricted number of well-informed adult volunteers.

Vision Basics and Retinotopy

A basic knowledge of the anatomy and physiology of the visual system is necessary for a proper understanding of the visual prosthesis in its various forms.

The eye can be compared to a camera with an adjustable optics including the cornea and the lens, focusing inverted images of the external world onto the retina. The retina is in fact a thin slice of brain that has expanded into the eye during the embryo development. It is made up of several cell layers. Among these, the photosensitive cells called cones and rods form the external layer with, as a result that light entering the eye has to cross the complete retina before to reach them (see Fig. 1c).

The 120 million or so photosensitive cells are not uniformly distributed. Their density is highest at the fovea, a region that corresponds to the fixation point, near the center of the visual field. Cones outnumber the rods in the central retina and are the only kind of photosensitive cells at the fovea. They discriminate colors. Rods predomi-

nate at the periphery. These very sensitive sensors play a major role in night vision, but do not contribute to high resolution nor color perception. The cellular hyperpolarization generated by light impinging on the photoreceptors is carried over to other the neuronal cell layers ultimately connected to output layer represented by the ganglion cells. With the exception of the fovea (most central part) and especially in the periphery of the retina (i.e., of the visual field), there is a great deal of convergence and encoding of the visual signal that must be squeezed from the analog modulation of ~ 130 million photoreceptor potentials into the discharge bursts of little more than 1 million ganglion cells. Roughly, the bipolar cells represent the main converging vertical link between the photoreceptors and the ganglion cells. The horizontal cells at the junction between photoreceptors and bipolar cells and the amacrine cells at the junction between bipolar cells and the ganglion cells provide sideways connections. Functionally, the retinal network can be subdivided in a number of parallel channels each focusing on the transmission of one aspect of images. These include a color coding system and movement detecting circuits. Bipolar cells and ganglion cells also subdivide in ON OFF types. The ON cells increase their activity on exposure of the center of their receptive field to light while their ongoing firing slows down when light strikes in the periphery of that region. The OFF cells react in the opposite way. Ganglion cells also respond more or less

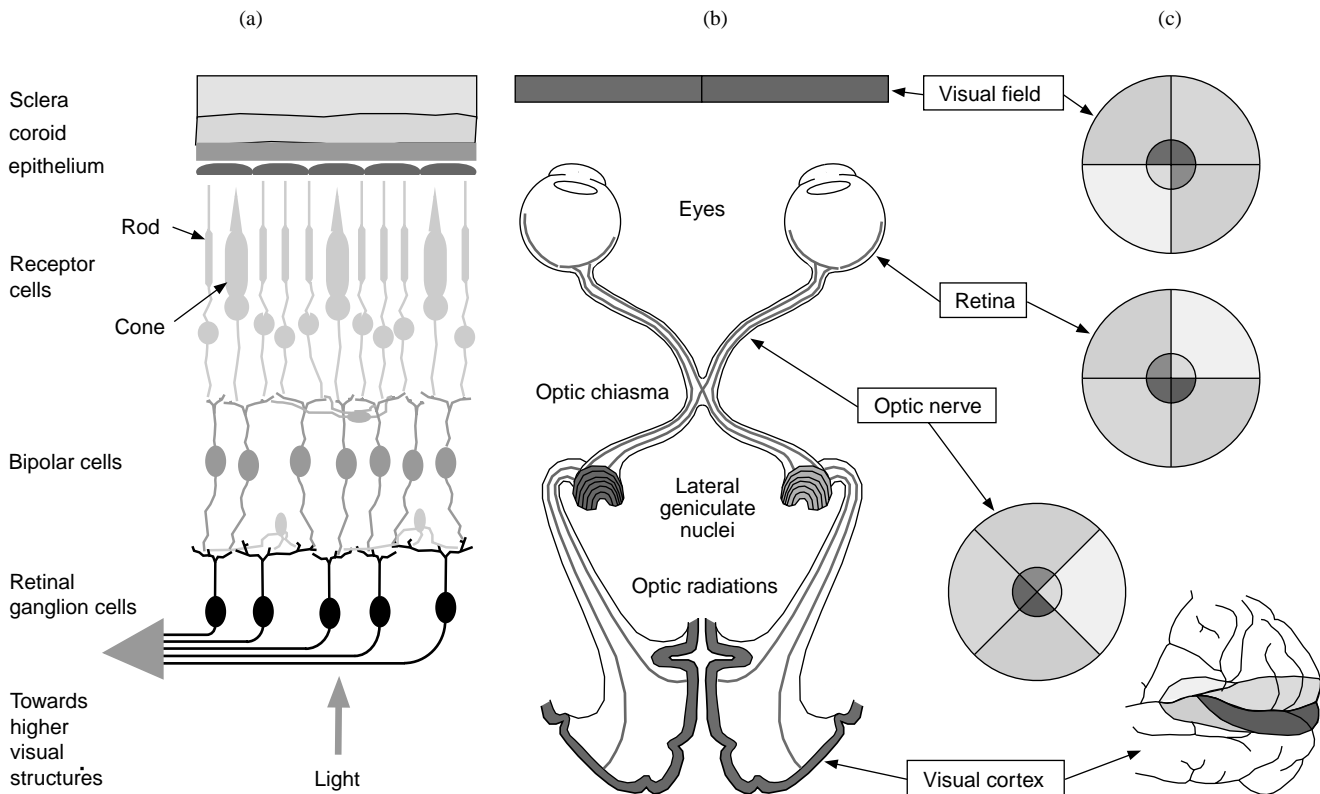


Figure 1. (a) Schematic representation of the visual pathways from the eye to the visual cortex with an indication of the hemifield segregation at the level of the chiasma. (b) Retinotopy mapping all along the visual pathways. Note also the cortical amplification, that is, disproportional cortical representation of the central visual field. (c) Enlarged view of a cut through the retina showing the various layers and their position in relation to light entering the eye.

transiently. The result is a spatiotemporal retinal filter providing the optic nerve with a complex and still not well understood neural code. Ganglion cell axons first run over the inner surface of the retina to join the optic nerve head located slightly nasally from the center of the retina. There, all ganglion cell axons join before to leave the eye and form the optic nerve.

In the orbit, the optic nerve is relatively slack, laying between extraocular muscles, fat tissue and various blood vessels and nerves. It is protected by a strong sleeve of dura mater as well as a very thin pia mater with cerebrospinal fluid in between. When the optic nerve enters the skull, it loses its dura mater sleeve, which now lines the inner surface of the skull. After a little >1 cm, the two intracranial optic nerves meet and exchange fibers at the level of the chiasma (see Fig. 1a). Ganglion cell axons of the nasal hemiretina or temporal visual field cross the midline and join the temporal axons of the other eye to form the optic tract. Foveal axons split into one branch toward each side. Some fibers (not represented) corresponding to accessory functions leave the mainstream visual pathway to reach the hypothalamus, pretectum and superior colliculus. The axons directly involved in vision end in the lateral geniculate nucleus. This structure performs further signal processing and receives control signals from various parts of the brain. From the lateral geniculate nucleus, the visual information reaches the occipital lobe of the brain through the optic radiation. Interestingly, corresponding inputs from both eyes are arranged in close proximity, but remain segregated up to the level of the primary visual cortex. From there, signals corresponding to various aspects of the visual stimulus are dispatched to different brain locations for further processing.

Consequently, of the fiber exchanges at the level of the chiasma, except for the representation of the fovea, one visual cortex receives only information about the contralateral visual field. In addition, albeit with much distortion, cells of the visual cortex tend to retain the topological relationship of the retinal location from which they receive their input. The resulting point-to-point correspondence with the retinal locations, and hence the visual field is called retinotopy. Some form of retinotopy is found at most levels of the visual pathways up to the cortex (see Fig. 1b). Despite an obvious need for corrective remapping, retinotopy of the structure to be stimulated will be an essential consideration for the development of a visual prosthesis. Right from the level of the retinal ganglion cells on, retinotopy nevertheless remains an approximation that does not take into account other significant aspects of the neural signal encoding.

Blindness

Under ideal conditions, the human minimum angle of resolution (MAR) is ~ 0.5 arc min (20/10 vision). However, the standard definition of normal visual acuity (20/20 vision) is the ability to resolve a spatial pattern separated by a visual angle of one minute of arc ($\sim 4 \mu\text{m}$ on the retina). The Snellen visual acuity measures the pattern recognition acuity as the ratio d/D , where D is the distance at which a sign subtends 5 min of arc, and d is the distance at which

they can be recognized. Reference patterns subtend 5 min of arc at a distance of 60 m and have features of 1 min of arc, corresponding to the standard normal acuity. A visual acuity of 3/60 means that such a sign can only be recognized at a distance of 3 m. For practical purposes, the Snellen chart is made up of different sized letters such that the examination can be performed at a single distance. Because the normal acuity corresponds to 1 min of arc, the MAR value in arc minutes has the same numeric value as the reciprocal of the Snellen fraction.

However, sight is a multidimensional ability that cannot be measured by acuity alone. The effect of a visual field defect or a reduced sensitivity to light cannot be compared directly to acuity.

The International Statistical Classification of Diseases and Related Health Problems of the World Health Organization (ICD-10) uses codes 1–5 to describe moderate, severe, profound, near total, and total visual impairments, respectively. Within that range, the label low vision (categories 1 and 2) designates a visual acuity $>6/18$ (0.3), but $<3/60$ (0.05) in the better eye with optimal correction or a visual field between 10 and 30° . The label blindness encompasses categories 3–5. Code 3 corresponds to a visual acuity <0.05 on the Snellen scale for the best eye using appropriate correction, or a central visual field diameter of $<10^\circ$ in its largest diameter. An acuity of <0.02 or a visual field $<5^\circ$ is coded 4 (near total) while total visual impairment means deprived of light perception.

To measure a visual acuity in near total visual impairment, alternative methods are used including close range on the Snellen chart reading, finger counting, the detection of hand motion, or the perception of light.

Blindness can result from any cause potentially affecting the visual pathways: genetic abnormalities, infections, metabolic diseases, trauma, vascular deficiency, or cancer for example. In one subgroup, mainly represented by retinitis pigmentosa (RP), age related macular degeneration, and stargardt's disease, it has been shown that blindness can result from a total destruction of the photosensitive cell layer while a proportion of other cells of the retina and the remaining of the visual pathways survive. The hereditary disease retinitis pigmentosa in particular, is a relatively frequent cause of severe and incurable blindness in developed countries with a prevalence of 1 in 5000 (2).

An essential distinction must also be made between early and late blindness. Indeed, several years of normal behavioral experience are required after birth for the human visual system to fully organize and suitably weight its synaptic connections (3). Those years are referred to as the critical period. The visual brain of people who are born blind is functionally different from that of those who lost sight later in life (4). Early blindness indicates a loss of vision before the end of the critical period. There is no sharp separation between early and late blindness and different visual functions each having their own critical period. Motion detection would probably mature before spatial localization, followed by object recognition, color vision, and stereopsis. According to ophthalmologic observations in diplopia, the critical development period for binocular vision ends around the age of 6. In addition, the visual impairment might result from a progressive

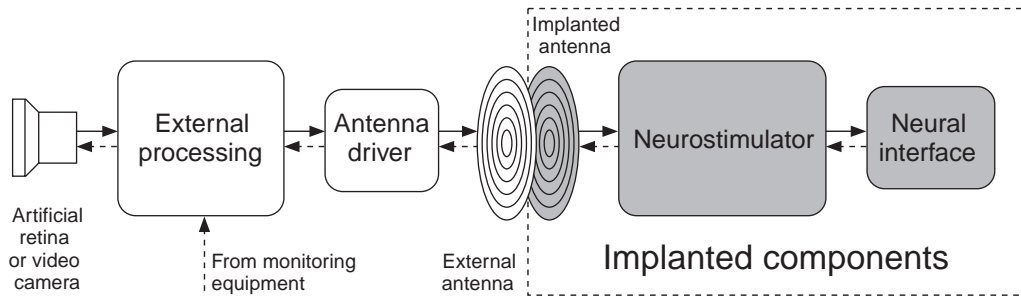


Figure 2. General layout of a visual prosthesis, with the external components on the left and the implanted components (shaded) on the right.

disease. Early blindness type increased metabolic activity has been reported in cases where progressive total blindness occurred around the age of 10 (4). By contrast, late blindness features are observed when the first visual defect appears before the age of 8–12, but only evolve toward total blindness thereafter, or when vision is accidentally lost at the age of 12 or later.

A prosthesis implanted after the critical period is doomed to failure. Age of blindness onset, however, is not the only factor resulting in a deviant visual system. In diseases such as retinitis pigmentosa, the loss of photoreceptors itself results in important remodeling of the remaining retinal network (5), resulting in aberrant neural connections and shielding by scar tissue, all potentially limiting more or less severely the efficiency of a visual prosthesis.

History

The idea of an electrical treatment for blindness is perhaps as old as the discovery of electricity itself. The first real attempt to implant a visual prosthesis, however, dates back from 1968 (1). A set of 80 electrodes were implanted over the occipital pole of the cortex of a blind person. Each electrode could be activated transcutaneously by an equal number of implanted radio receivers. Small precisely located phosphenes were obtained, suggesting that a useful prosthesis could indeed become possible, but the subdural cortex surface electrodes had very high thresholds and did

not provide an adequate resolution. While intracortical electrode arrays were being developed using technologies derived from the semiconductor industry (6), it became clear that in terminal retinitis pigmentosa, a significant fraction of the ganglion cell population remains functional despite total blindness (7,8). In such cases, a visual prosthesis interfacing with the surviving layers of the retina or with the optic nerve, all referred to as the anterior type could be as useful as the more complex cortical implant. As a result, starting around 1990, perhaps partially dragged by the success of the cochlear prosthesis, a renewed and still growing interest in visual prosthesis rapidly expanded worldwide.

Basic Theory of Operation of the Visual Prosthesis

In all published visual prosthesis approaches, the visual system is very schematically seen as a transmission chain in which retinal image pixels are encoded into series of electrical pulses ultimately activating the visual cortex. The visual prosthesis is meant to replace or by-pass the defective link in the neural chain. At the cost of not using still functional body parts (eye optics, e.g.), most visual prosthesis designs replace the whole front end of the sensory chain, requiring only the prosthesis output to be connected to the nervous system (see Fig. 2). A picture of the external and implanted components of a prototype of the optic nerve visual prosthesis is given in Fig. 3 as an example.

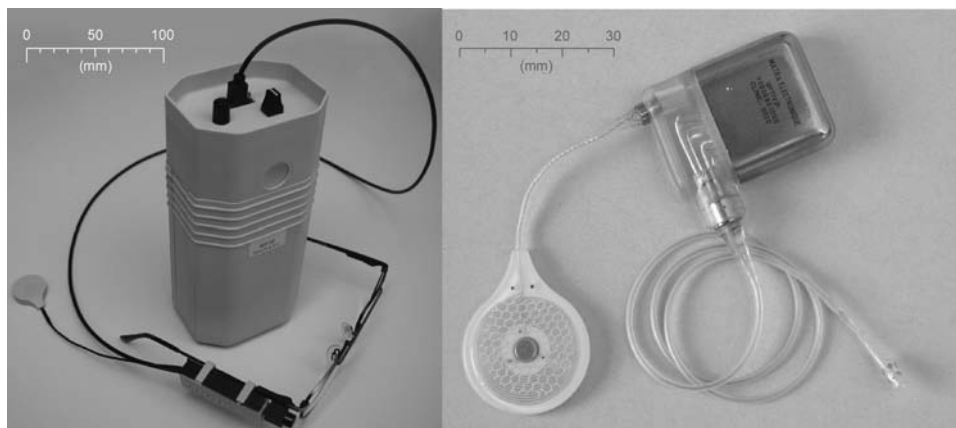


Figure 3. Implementation example: picture of the optic nerve visual prosthesis prototype. On the left, the external processor linked to glasses on which a camera and the external antenna are attached. To the right, the implanted components with the internal antenna, the neurostimulator and the spiral cuff electrode with their leads and connectors.

Depending on the approach, the prosthesis sensors can be any imaging device from an implanted photoarrays (9) to a miniature cameras worn on a pair of glasses. In this last case, a small-size camera is also an asset for esthetical reasons.

A processor is necessary to transform the video images in properly encoded and modulated commands to the implanted stimulator (10). The prosthesis must allow real time object localization and identification. Visual information quickly represents a huge amount of data imposing high demands on the computational performances while the processor including its power supply, must remain easy to carry around. This portability often represents a technological challenge by itself. Adaptability to new and rapidly emerging developments is also an essential requirement in this quite young technology. Transcutaneous electromagnetic transmission of data and power using magnet fixated antennas is now standard for cochlear implants. This principle was already applied in a more primitive way by Brindley back in 1968 (1). He then solved the problems of parallel transmission through the implantation of an array of transmitters. Recent projects use combined power and data electromagnetic telemetry systems similar to those of the cochlear implant systems but with improved power and data performances.

Different Approaches

The neural interface typically represents the most critical and challenging component involved. In keeping with this, the different approaches to develop a visual prosthesis are often classified according to their connection point along the visual pathways: subretinal, epiretinal, transretinal, hybrid, optic nerve, cortex surface, or intracortical. The interface can be chemical or electrical. Until now, the chemical approach has been limited to some work on retinal devices. Many more projects use electrical stimulation through specifically designed electrodes.

Chemical Stimulation. The basic idea of this approach is to use neurotransmitters (L-glutamate, e.g.) or other specific chemicals that are known to activate neural or retinal cells. Directly over the retina, photochemical reactions (typically requiring ultraviolet, UV light) liberate active components from an inactive parent molecule or from 'caging' molecules (C_{60} , fullerene) (11). The result is a direct translation of retinal images into a corresponding neural activation pattern. Alternatively, an electronically activated multichannel microfluidic device could deliver the needed chemicals locally and be used in a visual prosthesis as a substitute to electrodes (12). At present, most research efforts are still devoted to basic problems such as to reduce the required light energy level, biocompatibility, transport of the chemicals and reservoir refill (13). In the future, however, chemical stimulation could have several advantages. There is no electrode corrosion. Stimulation selectivity can bear on the subgroups of retinal cells, and so mimic the physiological activation achieved by synaptic transmission. In addition, the proposed stimulating electrodes can be made on soft materials supposed to be less damaging for the retina than electrode arrays (12).

Subretinal Implants. The most appealing aspect of the subretinal implants is that it exploits the supposedly healthy eye optics and interfaces to the visual pathways before any neural processing has blurred the neural code. The aim is indeed to replace the damaged photosensors by an array of passive miniature photosensitive devices transforming the retinal light image into local electrical stimulating currents. These currents would in turn activate the surviving neuronal circuits of the retina in keeping with the light intensity they receive, much in the same way as photosensitive cells do. The idea is straightforward and logical. Natural accommodation and physiological eye movements would remain functional. Very thin ($100\ \mu\text{m}$) flexible electrode construction and perforations allowing nutrient and other metabolic exchanges between the retina and the choroid could insure biocompatibility (14). The use of glycoprotein coating has been suggested to improve the biocompatibility of the components. Small implants can be quickly and securely trapped between the neural elements and the pigmentary epithelium, which seems to pump out this space (15). The light wavelength sensitivity of the microphotodiodes is in the 500–1100 nm range, which corresponds reasonably well with the visible spectrum of 400–700 nm.

Chow's group developed a $25\ \mu\text{m}$ thick subretinal implant of ~ 5000 subunits on a 2 mm diameter chip, enough to provide a tunnel vision of a little $> 8^\circ$. Such devices have been implanted in a number of retinitis pigmentosa patients. After 6–18 months, no significant side effect has been noticed and some patients reported a transient improvement not related to the implant position in the visual field. All the implants were electrically functional, but no visual response of the implant themselves was demonstrated (9). These devices are only powered by incident light but, as shown by Zrenner's team, currents generated by microphotodiodes are by far too low to activate bipolar cells. Available devices, would require 12 klx (16) to do so while normal ambient light typically reaches $\sim 8\ \text{lx}$. An active amplification is therefore necessary, finally sharing with other approaches the need for an external power supply raising again the problems of bulkiness, heat dissipation, energy and data transmission.

Epiretinal Implants. In the epiretinal approach, micro-electrode arrays are attached on the vitreal side of the retina, just over the inner layer that contains the ganglion cells and their axons. The stimulation contacts are supposed to activate local ganglion and/or the bipolar cells. Acute stimulation of the retina in human volunteers has gone very far in demonstrating the feasibility of a prosthesis with a resolution of $\sim 1.75^\circ$ (17). Concerns have been raised about the possibility to activate passing-by optic nerve fibers as well, which would result in aberrant perceptions (18). Also, because of the retinal preprocessing, activity in the ganglion cells can no longer be seen as a point-to-point representation of the retinal image and the encoding at that level already exploits the time dimension as well. Humayun's team developed a prosthetic device with an array of 16 electrodes connected to an implanted stimulator located outside the eye. An external system for

image acquisition and processing sends data and power to the implanted electronics by telemetry. This device was implanted in several blind RP patient. Initial results seem encouraging (19).

The basic principle in most visual prosthesis approaches is that stimulation through a small electrode will result in the perception of a point-like phosphene of corresponding retinotopic localization. An array of such electrode contacts would produce a number of phosphenes that can be distinguished by their location in the visual field. After correction for any nonconformal localization, an image perception could then be obtained by activation of the corresponding pixel electrodes. It will be seen that this pixel phosphene method to selectively activate a subset of fibers or ganglion cells does not hold in the case of the optic nerve stimulation.

The Transretinal Approach. A transretinal approach (20) has also been suggested whereby a needle placed in the vitreous is used as a single cathode facing a miniature array of anodes slit under (or in the sclera), thus yielding a transretinal stimulation. Evoked potentials have been obtained in animals using eight contact on a $2 \times 4 \times 0.18$ mm electrode with polyimide substrate. There is no indication yet as to which cells form the primary target.

The Hybrid Approach. The team of Yagi and Tano has started research to grow transplanted neural cells from a subretinal implant to the central nervous system using axon-guiding substrates. This approach could also be applied when ganglion cells are destroyed. This work remains very much preliminary and no results are available yet.

The Optic Nerve Approach. A direct stimulation of the ganglion cell axons with an optic nerve electrode can be seen as an alternative to the epiretinal stimulation. The basic idea here is that the simultaneous activation of a number of contacts can focus the stimulation on a chosen subset of the axon bundle by controlling the applied electric field. As a result, electrode contacts around the nerve can yield a selective activation (21). The number of different fiber subsets that can be stimulated independently and thus the number of phosphene perceptions that can be obtained is much larger than the number of electrode contacts available. This principle could be applied to all electrodes carrying contacts at some distance from the target cells. Focal stimuli are thus generated serially by each multicontact activation instead of in parallel through individual contacts.

This concept has been validated in a human implantation (22). The results confirm a retinotopic organization of the intracranial optic nerve and phosphenes are obtained with safe electric charges. Interestingly, due to the signal encoding in the optic nerve, phosphenes do not reproduce the distribution of the fiber activation and are much smaller than expected. Their position in the visual field can be controlled, which is of course essential in the prospect of the visual prosthesis development, making it possible to convey image information, even without resorting to more

complex selective stimulation schemes (e.g., superficial fiber blocking).

It has been suggested (23) that a penetrating electrode could increase the number of available independent responses, but the damage inflicted to the nerve has not yet allowed to validate such an alternative.

The first optic nerve electrode was implanted behind the orbit just in front of the chiasma. A new surgical technique has been developed to implant an eight contact electrode in the orbit. Avoiding intracranial surgery is certainly reassuring for the patient, but the intraorbital approach is technically difficult and has several drawbacks. At that location, the optic nerve is indeed covered by the dura mater, which shields off the fibers from the stimulation and therefore results in higher thresholds. Also, somatic sensory nerve fibers as well as blood vessels are present in the dura mater, and eye movements could limit the stability of the electrode contacts. Nevertheless, the feasibility of this approach has been demonstrated recently.

The Cortical Approach. The very first implanted human visual prosthesis prototype (1) included an array of 80 electrode contacts placed over the occipital cortex of a blind person and linked to an equal number of miniature transmitters placed under the scalp. High thresholds and poor selectivity have led to the conclusion that intracortical rather than cortex surface electrodes were necessary (24). Subsequently, a two-dimensional (2D) device known as the Huntington electrode (25) and a three-dimensionally (3D) structured, single plane, Utah electrode (26) were proposed for intracortical implantation. Resolutions of $\sim 400 \mu\text{m}$ can be obtained (27) where the surface electrodes of Brindley could only resolve minimal distances of 2–3 mm. In an acute experiment, 38 intracortical microelectrodes have been implanted, for a period of 4 months in the right visual cortex of a human volunteer (6). Two-point resolution was about five times better than had typically been achieved with surface stimulation. All phosphenes were located in the left hemi-field with the majority above the horizontal meridian. There was a clustering of most of the phosphenes within a relatively small area of the visual space (6). As opposed to subdural electrodes, intracortical devices have the potential advantage to reach the hidden parts of the cortex in the depth of the calcarine fold corresponding to the big gap in the region of the horizontal meridian of the visual field as observed by Brindley.

Brindley himself stressed the important variability of the cortical maps among individuals. Individual mapping of each cortical implant is thus expected to be necessary. Biocompatibility is still a major point of concern, owing to the large number of electrode contacts and stimulator connections, especially for chronic human implantation. However, compared with the prechiasmatic approaches, the intracortical alternative would in the long run have the advantage of being applicable to many more conditions, not just diseases of the retinal photosensors.

A consortium led by Troyk has undertaken the development of an intracortical prosthesis based on a 256-channel stimulator module and a 1024-contact array. Recently, 152 intracortical microelectrodes have been chronically implanted in area V1 of a male macaque.

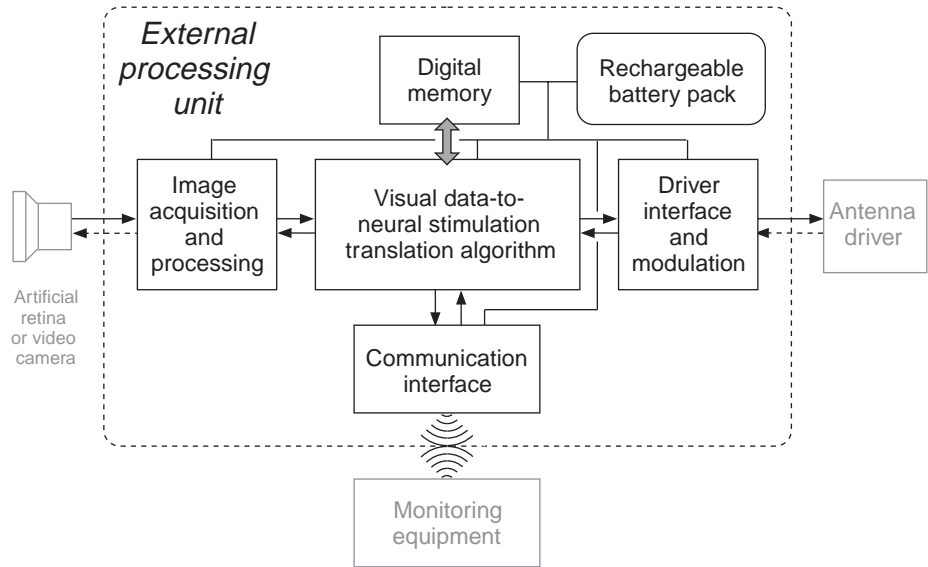


Figure 4. Detailed representation of the external components of Figs. 2 and 3.

Receptive field mapping was done using a memory saccade task (28). It is expected that such new animal psychophysical tests will compensate for the lack of a linguistic report and allow further developments in animals before finally turning to human trials.

ENGINEERING ASPECTS

The Hardware

Typically, the hardware of a visual prosthesis is composed of an external system and an implanted part. The external system (see Fig. 4) includes some image capturing device, an external processor and the external part of a transmission unit. Implanted (see Fig. 5) are the other one-half of the transmission system, stimulators, and an electrode.

Depending on the approach, there are important variations on this basic scheme. In subretinal projects, a photosensor array implanted directly in the eye could replace the external camera. Some specifications of the example of the optic nerve prosthesis are given as an example in Table 1.

External Components

The Image Grabber. A miniature video camera typically mounted on glasses provides the image capturing device of the visual prosthesis. A low weight camera with good esthetical appearance is of importance to the blind person and can be improved by miniaturization. It could further be stated that the most trivial imaging devices largely outperforms the needs of all present day visual prostheses (29). Nevertheless, a good image quality including some

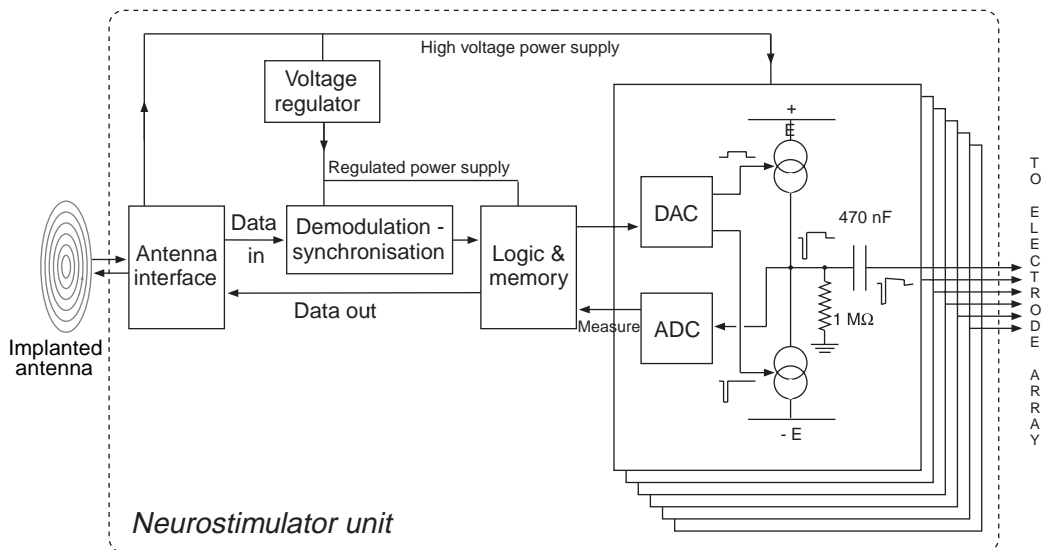


Figure 5. Detailed representation of the implanted components of Figs. 2 and 3.

Table 1. Example: Main Specifications of the Optic Nerve Prosthesis

Image processing	
Pixel size	$1 \times 1^\circ$
Processed visual field	$32 \times 64^\circ$
Telemetry	
Forward carrier frequency	12 MHz
Forward data rate (Frequency Shift Keying modulation)	$3 \text{ Mbit} \cdot \text{s}^{-1}$
Data frames length (including 6 bit CRC)	64 bit
Return carrier frequency	24 MHz
Return data rate	$1.5 \text{ Mbit} \cdot \text{s}^{-1}$
Maximal power transmission (class E)	120 mW
Maximal antennae separation distance	5.5 mm
Implant	
Power consumption	80 mW
No. of independent current sources	8
Current sources	
Time resolution	21 μs
Output voltage range	-9 to +9 V
Maximum output current	3.2 mA
Current resolution (exponential)	8 bit + sign
Electrodes	
Number of contacts	8
Contact recess depth	80 μm
Contact area (platinum)	0.2 mm^2

standard correction features for luminosity compensation, autofocus, and avoidance of glare (e.g., could greatly simplify the later required image analysis). Later, when the usefulness of some image analysis procedures, such as edge enhancement or nonuniform resolution will have been demonstrated, it might become worthwhile to consider implementing such features in the front end hardware. The video camera would then be replaced by a specific imaging device that could evolve into a real artificial retina.

In conditions, such as blindness, due to retinitis pigmentosa, the main target group for all prechiasmatic types of visual prosthesis, the optics of the eye can still be functional. In the subretinal approach, a photosensor array attached to the retina would transform the eye in an artificial video camera, preserving an essential functionality of the natural eye, namely, gaze orientation. However, in addition to biocompatibility requirements, mounting such a device as well as the necessary control electronics will not be easy (15). Provision must also be made to power the device and send its output signal to an external processor, all with an acceptable level of power dissipation (30).

The external processor. The bottleneck of all visual prosthetic systems is the rather limited quantity of information that can be handled by the neural interface. The amount of data to be transmitted must therefore be reduced by all possible means including image analysis. This topic is likely to become very important in the near future. Limiting images to black and white, thresholding, and edge detection are just preliminary steps. More sophisticated image processing techniques will have to be implemented. Therefore, powerful belt wearable processors will

be necessary to translate visual data to neural stimulation (see Fig. 4). Quite unlike the situation with cochlear implants, however, little is known about the precise encoding of visual information in the visual pathways and the first human implants will contribute to such knowledge (31) allowing more efficient algorithms to be developed.

Image processing can be subdivided in several steps including: analysis, selection, mapping, and encoding. The purpose of the image analysis is essentially to reduce the amount of visual data to be transmitted through the prosthesis. After an image data reduction step, a selection procedure should allow only the most important features to be sent through. Mapping refers to the method used to establish a correspondence between selected image pixels or features and the phosphenes that can be generated. Finally, a control signals must be generated such that the implanted current sources will issue the intended stimulus. This last encoding step is entirely defined by the characteristics of the implanted device. Mapping on the other hand is linked to the neural code in the neural interface. Much of it is unknown and still requires experimental work with implanted volunteers. Later on, because of the anatomical variability, at least some individual remapping will be required before revalidation itself can be started with a visual prosthesis. Finally, except for the most basic and empirical aspects, image analysis and item selection will have to resort to further psychological studies on perception. Some of these studies could be done with healthy volunteers using virtual reality simulations.

A detailed description of one example of stimulation algorithm is given below. It should be stressed that a communication interface with the processor (see Fig. 4) is an important tool in the development and adaptation of the translation algorithm. Using monitoring equipment, the perceptions of volunteers can be explored and the system can be customized or adapted according to the findings.

Typically attached to the external processor, and probably the major weight to carry along, are the power supply batteries. Note that all published visual prosthesis designs use externally powered implants. Therefore, current is also drawn from this main battery to provide power to the implant. The main specifications to be taken into account are the user friendliness of wearing and reloading the rechargeable batteries. Their size will be defined by a trade-off between weight and autonomy.

The Transmission Unit. Electromagnetic telemetry is based on the classical cochlear implant design. It typically uses two similar antennas holding a small biocompatible (stainless steel encapsulated) magnet in their center. One of these is implanted under the scalp just above the mastoid, behind the ear. The magnets maintain the external antenna over its internal counterpart. This turns out to be the most popular transcutaneous transmission system. The normal skin thickness separating the antenna coils is from 3.5 to 5 mm. However, just after surgery, swelling up to ~ 7 mm can be observed that can take >3 months to recede. During this period, the increased distance might cause malfunction of the antennas.

A telemetry return channel from the implant to the external system is an important feature, providing acknowledgment signals as well as technical diagnostic and monitoring information. For example the output voltage of the implanted controlled current sources gives an indication about the proper operation of the system as well as an estimation of the electrode contact impedance. Also, a measurement of the supply voltage indirectly proves that the power transmission is working adequately.

Alternative antenna arrangements have been considered. For example, as proposed by the Boston group (see Fig. 6), an external primary antenna or coil could be mounted on spectacles and the secondary coil could be implanted on the eye surface or in the anterior segment of the eye. Other transmission techniques exist. A transcutaneous socket on the head (32) has been used but it exposes the patient to severe infectious complications (33) and it is not really acceptable on esthetical grounds.

An all-in-the-eye alternative has been suggested whereby an infrared (IR) (820 nm) laser would transmit power and signal to an intraocular prosthesis through the transparent media of the eye. Data can be transmitted efficiently but heat dissipation in the implanted components is still not compatible with the power requirements of practical devices. Eye movements would also represent a tremendous challenge.

Implanted Components

The Stimulator Case. The following description pertains to the intra-orbital optic nerve visual implant. The purpose of choosing one example is to provide a set of realistic numbers, but the principles apply to most visual prosthesis approaches. The antenna is connected to an 8 mm thick titanium case half engraved in the parietal skull. This neurostimulator hybrid circuit (see Fig. 5) contains the transmission electronics as well as control logic circuits and the

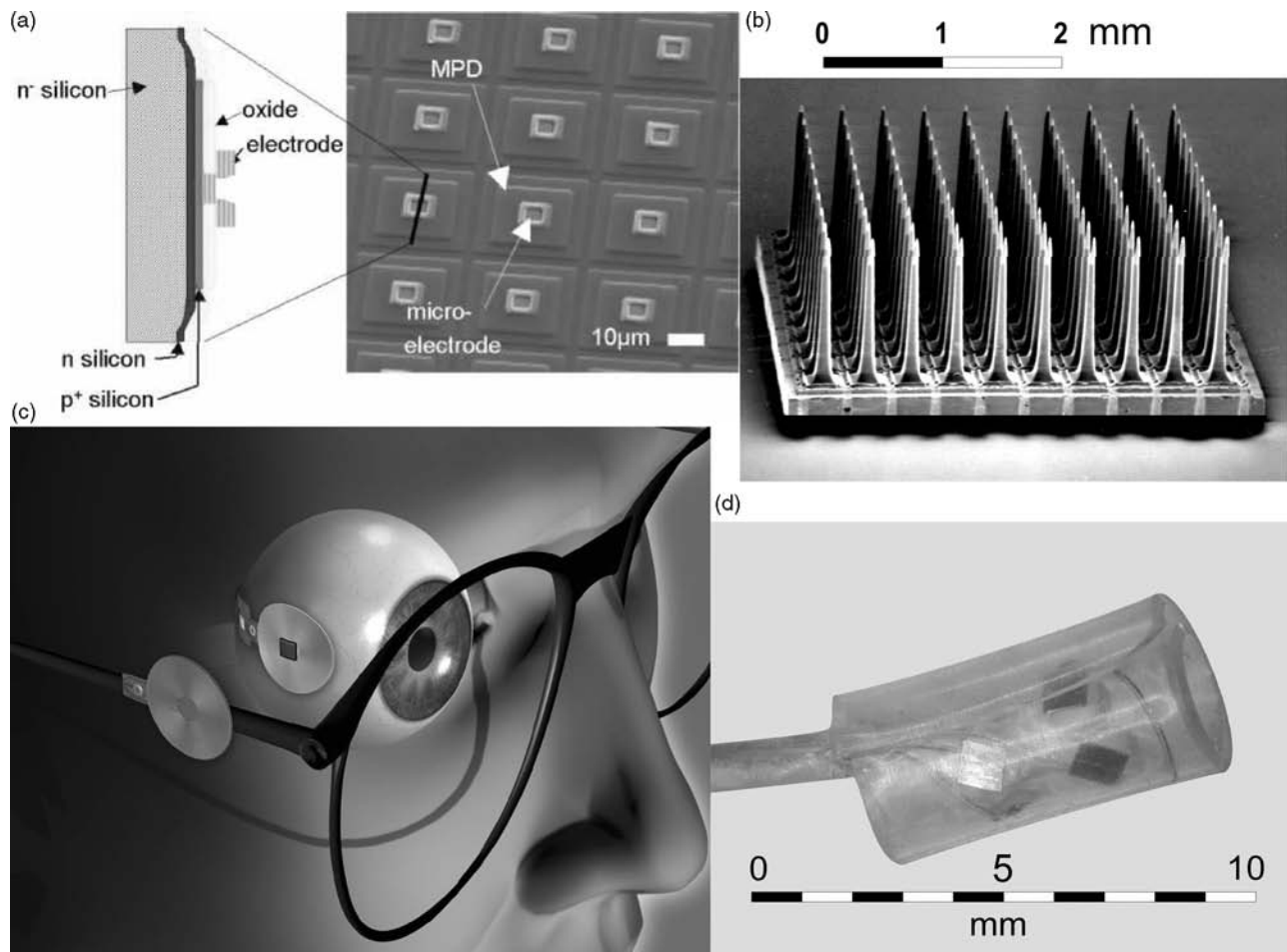


Figure 6. Picture of a number of different electrodes characterising the various approaches to the visual prosthesis. (a) Details of a MicroPhotoDiode (MPD) chip intended to be implanted subretinally in the space normally filled up by cones and rods. (Courtesy of Eberhart Zrenner.) (b) The Utah Cortical Electrode Array. (Courtesy of Richard A. Normann.) (c) An overview of the epiretinal prosthesis with wireless radio frequency transmission being built by researchers at the Boston VA hospital, Harvard Medical School and Massachusetts Institute of Technology. Only a small square end on the left will enter the eye. (Courtesy of Joseph Rizzo.) (d) The optic nerve spiral cuff electrode.

current sources. This circuit occupies the major part of a titanium encasing is closed by a laser welded plate carrying 10 feedthrough connections. This number of feedthroughs is a main factor in determining the minimal size of the stimulator case. The metal encasing is internally connected as the common output reference electrode. A biocompatible polymer cap carries two connectors (a two-contact one for the antenna and an eight-contact connector to the electrode) and protects their junction to the feedthrough wires.

Each stimulator connection receive has two current sources, one for each polarity, as required for the application of biphasic charge balanced pulses. As a rule, implanted current sources are coupled to the stimulation electrode through large output capacitors for safety reasons. These capacitors (470 nF), significantly contribute to the size of the implant.

In an attempt to reduce the chip size while providing a large number of channels, alternative designs (34) do multiplex the output of a single stimulator to a number of electrode contacts instead of having one dedicated stimulator circuit for each contact. Also, the telemetry data rate could be reduced if less degrees of freedom were given in the stimulus definition. The possibility to modify the shape of each pulse individually does indeed require a high rate data transfer.

The Neural Interface or Electrode Assembly. Electrodes (see Fig. 6) are the most characterizing component of the different visual prosthesis approaches. They represent a rather critical and important component in electrophysiology applications (see the chapter on Bioelectrodes). In a nutshell, electrodes form the interface between the electric current carried by electrons along metallic conductors and the ionic conduction found in aqueous solutions, such as the body tissues. This potentially harmful transformation is chemical in nature, called a reduction reaction at the cathode and an oxidation process at the anode. A local pH change is induced, becoming more acid at the anode and more alkaline at the cathode. Of course, the exact nature of the chemical reactions taking place will be influenced by the electrode metal as well as the chemicals present in the solution, their concentrations and the applied potential among other factors. Irreversible chemical reactions cause corrosion and might severely limit the electrode lifetime. However, electric charge limits can be defined within which the chemical changes remain mostly reversible. Maximal values of $0.4 \text{ mC} \cdot \text{cm}^{-2}$ (35) are given in the literature for platinum and $4 \text{ mC} \cdot \text{cm}^{-2}$ for iridium oxide (36). These metals are most often used in implantable electrode contacts. The reversible charge limits will preserve the electrode contacts, but they have no direct bearing on the potential damage to the living tissues being stimulated.

As a rule, biphasic stimulation pulses are used in implants in order to reduce the risk of deleterious effect. The cathodic phase is meant to activate the neural tissue while the anodic phase, often of smaller amplitude, but with a correspondingly longer duration, is supposed to compensate for the injected electric charge and thereby revert the local chemical changes. Because the currents are not uniformly distributed over a contact area, the electrode

Table 2. End Pulse Voltage–Current Ratio^a

Pulse Duration	42 μs	342 μs
	<i>Cathodic Current</i>	
0.4 mA	2.4 k Ω	5.0 k Ω
1.6 mA	2.3 k Ω	3.2 k Ω

^aPlatinum contacts of 0.2 mm^2 referred to titanium case as anode

geometry should also be taken into account in estimating the allowable charge limit.

The contact impedance or voltage–currents is an important characteristic of the electrodes. Dependent on a large number of factors, it is nonlinear and decreases with the current strength while increasing with the pulse duration. Table 2 gives average values as measured in the example of the optic nerve prosthesis. With epiretinal stimulation (37), using 1 kHz $10 \mu\text{A}$ sine waves, impedance values are found to be distributed from $<1 \text{ k}\Omega$ to $40 \text{ k}\Omega$ for the same contact area 0.2 mm^2 and larger values for 0.05 mm^2 contacts. The higher values correspond to lower thresholds and are found when the contacts are closer to the retina.

The Prosthesis at Work

Phosphene Generation

Phosphene Description. A first step in the rehabilitation process with a visual prosthesis is to explore the available tools, that is, the identifiable phosphenes that can be generated by the system. A large number of stimulus variable (selected contact, pulse amplitude, pulse duration, train duration and frequency) vectors are systematically tested. The resulting perceptions must be carefully recorded for later use. A more straightforward relationship between the stimulation parameters and especially the contact being activated and the perceived phosphene localization can be expected with the retinal approaches. Still, however, the thresholds will have to be determined and data will be required to correct for the expected (18) mapping distortions.

With the optic nerve prosthesis, the phosphene diameters are variable from a point-like perception to $>10^\circ$. The distribution of these values has modal peaks $\sim 2, 4,$ and 7° depending on the stimulus parameters. The luminosity is most often reported as weak. Some phosphenes, especially at the periphery of the visual field have the appearance of solid patches while others have variable textures, such as orderly arranged rows or columns of points or small lines. Phosphenes are colored sometimes with contrasting colors between texture elements and the local background. The global background of the visual field is usually described as black, but can sometimes be perceived as gray or slightly colored. Fluctuations of the diffuse background or spontaneous phosphenes occasionally hinder the perception of induced phosphenes.

In the optic nerve example, phosphenes remain located in a quadrant of the visual field that reflects the retinotopical position of the contact used. The center of the phosphenes is located peripherally if near threshold, high frequency and prolonged train stimuli are used and their position in the visual field takes on a more central position for stronger

stimuli, the strength being expressed as a stimulation to threshold current ratio. The most eccentric position that can be accessed by a given train stimulus is located nearer the center for short and/or low frequency trains (31). Phosphenes generated by single pulses are systematically located in the center of the visual field. It is thus possible to model and control the average phosphene center position in the visual field. Unfortunately, in the first implanted volunteer, phosphenes could only be induced in a small region of the visual field (from 8.5° on the left, to 11.5° on the right, and from 6° up to 47° down). This limitation is likely to be linked to the retinitis pigmentosa itself as it is known that only a fraction of the ganglion cells survive this disease (8) and the periphery is more affected than the central retina (38).

With retinal stimulation as well, it has been found that the form of the percept did not always match the stimulation pattern (18) despite the fact that small phosphenes could be obtained at visual field positions corresponding to the point of the retina being stimulated (18).

For all approaches, the phosphene position is clearly referred to retinal coordinates and therefore to the eye orientation at the time of stimulation. Any change in gaze direction or head movement preceding the stimulation will accordingly shift the perceived phosphene location. Thus, for every visual prosthesis with an image capturing device that is not attached to the eye, the users must learn to keep their eyes fixed in the orbit.

Phosphene perceptions are described as short-lived, flash-like. With repeated stimuli, flicker fusion has been found to occur ~ 8 – 10 Hz and is associated with a perception threshold decrease. At very low repetition frequencies (1–2 Hz) flashing phosphenes remain of stable brightness as long as the stimulation is maintained. By contrast, at higher frequencies, successive phosphenes are perceived with decreasing brightness and disappear after 1–3 s.

Brindley (cortex, subdural stimulation) did not obtain a flicker fusion of cortical stimuli. Flicker fusion frequency was found to be similar for normal vision and for electrically generated perceptions using epiretinal electrodes (39). In this last approach, the perceived stimulus brightness increased with increasing stimulus rate as well as with increasing the stimulus current, which is not the case with optic nerve stimulation whereby the perceived phosphenes become larger with stronger stimuli. This difference might be related to the fact that the epiretinal stimulus can activate bipolar cells in addition to ganglion cells.

Simulation on Mathematical Models. Computational models play an important role in this kind of work. Numerical evaluation of the electric potential distribution in inhomogeneous and anisotropic media (volume conduction model), and of the resulting neural activation (neural membrane model), can provide a better understanding of the observed responses and provide a predictive design tool. As later confirmed experimentally (40), modeling the epiretinal stimulation has shown that cells or the initial segments have lower threshold than the passing axons (41). For pulses longer than 500 μ s, however, the bipolar cells become activated first. Confronting modeling results with experimental data has been useful in the interpreta-

tion of the results of optic nerve stimulation studies as well, showing that the retinotopic extension of the perceived phosphenes is quite unlike the corresponding topologic distribution of the activated axons in the optic nerve.

Other techniques including an hybrid association between an adaptive neural network model and analytical expressions of physiological laws can describe the expected localization of phosphenes generated by optic nerve stimulation. These tools will in the future efficiently help to develop individual lookup tables derived from preliminary open loop testing sessions so that these can be kept as short as possible. Such tables or mapping systems are of primary importance for the development of a visual prosthesis, no matter the approach.

Stimulation Thresholds. Despite the functional importance of the activation threshold, values applicable to a given situation are hard to find in the literature. A first reason for this is the lack of standardization. Current-controlled stimulators being used in most approaches, threshold current intensity could seem to be the logical variable to describe. However, as shown by the strength-duration relationship, the electric charge required for activation approaches a minimum asymptotically for short pulses, but increases very fast for pulses longer than the above the chronaxy value (see the article on Functional Electrical Stimulation). A threshold electric charge for realistically (in terms of the electronic circuits and tissue or stray capacitance) short stimuli can thus be seen as an appropriate expression, with the advantage that it directly relates to potential electrochemical damage to the electrode and surrounding tissue. The charge per unit area of electrode contact would be an even better characteristic value including the size factor as a minimal geometric parameter.

In an homogeneous medium, and at a distance much larger than the electrode contact diameter, the thresholds current is proportional to the square of the distance to target ($I_{th} = k r^a$), where $a = 2$. This power law can be seriously distorted in real tissues. For example in the retina (ganglion cell stimulation) exponent a can vary between 0.84 and 3.19 depending on electrode geometry and anatomical factors (29). Threshold values can further be expected to depend on factors such as the pulse shape, the nature of the primary target and the animal species.

The nerve membrane activation is not necessarily linked to the production of a cortical evoked potential neither the perception of a phosphene. For example, with optic nerve stimulation, long high frequency train stimuli have been found to yield much lower perception threshold than identical single pulses while in both cases expected to activate the axons at the same level (30).

Another aspect to be considered is that a disease, such as retinitis pigmentosa itself, can be responsible for an important increase in the stimulation threshold (42). The design of a prosthesis must therefore take this fact into account as well.

With a subretinal array, 1 nC or $10 \mu\text{C} \cdot \text{cm}^{-2}$ is reported to be sufficient to excite surviving retinal neurons (43). Phosphenes are generated by epiretinal stimuli of $350 \mu\text{C} \cdot \text{cm}^{-2}$ through an electrode contact with a diameter

of 520 μm (19). However, this threshold can vary between 24 and 702 μA depending on the proximity to the retina (37). The same authors indicate that there is no simple relationship between the threshold and the electrode contact area. For the transretinal stimulation, a threshold value of 56 $\mu\text{C}\cdot\text{cm}^{-2}$ is given, corresponding to a total charge of 28 nC (20). In the case of the optic nerve prosthesis, an estimation for trains of 25 pulse (200 μs duration) at 200 Hz would yield a threshold of 60 μA with the intracranial cuff (without dura mater) and 700 μA with the intracortical cuff. The respective threshold charges are 6 and 70 $\mu\text{C}\cdot\text{cm}^{-2}$. The corresponding chronaxy values are 130 and 192 μs . Brindley (1) gives a value of 13 V on an impedance of 3 k Ω , yielding 4.3 mA to reach threshold with a good electrode driven at 30 Hz with 200 μs pulses. This represents a charge of 860 nC. Thresholds down to 0.4 nC were obtained with 200 μm^2 intracortical electrodes, which corresponds to 1.9 μA for 200 μs pulses at 200 Hz (24).

Stimulation Upper Limits. The upper limit of the stimulation range (44) is even more important to the designer than the threshold. The maximal current will indeed define the required minimal supply voltage as calculated from the expected load impedance. Any stimulus strength above a full activation of the target would represent a waste of the stimulator range and lead to unwanted physiologic response. However, despite the fact that they require higher current values, short stimulation pulses in fact activate the structures with lower charges, and therefore with a lower risk for local tissue and electrode damage. Unfortunately, the maximal current available is directly related to the size of the stimulator ASIC, which should be minimized, especially in a multichannel implant.

Increasing the electrode area to reduce the current density might seem to be an alternative. However, much of the gain in safe stimulus strength could easily be lost in the higher threshold characteristics of larger contacts that cannot be placed close to the target. Every approach will thus lead to a trade off between the stimulus strength required for proper activation, the safety limits and the size of the implant. In the example of the intracranial stimulation of the optic nerve, stable threshold levels have been observed for >6 years of compliance with an upper limit of 150 $\mu\text{C}\cdot(\text{cm}^2\cdot\text{phase})^{-1}$ for charge compensated biphasic pulses up to 50 Hz and <50 $\mu\text{C}\cdot(\text{cm}^2\cdot\text{phase})^{-1}$ at higher frequencies.

From Phosphene to Visual Perception (Stimulation Algorithm). Once phosphene perceptions can be elicited in a controlled way, one is left with the question as how to use them in order to convey visual information to the visual system of the blind person. In the optic nerve visual prosthesis example, a look-up table is first established using a phosphene position model as a mean to average and interpolate limited experimental data. From the collection of theoretically elicitable phosphenes, only those obtained with a charge density <300 nC/phase (0.2 mm² contact area) are considered. Train stimuli longer than 40 ms total duration are excluded as well. The phosphene center has to be at least 1° of visual angle apart and, in

cases of choice, the phosphene produced with the lowest stimulus strength is selected. These criteria resulted in a set of 109 individually addressable phosphene with defined position. The portion of the visual field covered is limited to 14° horizontally and 41° vertically. Even within that region, there are holes where no phosphene can be obtained. In addition, the phosphene area is usually >1° and there is thus a clear overlap in the patches of visual field covered by neighbors. The usable look-up table is thus far from representing a complete set of nicely tiled point sized light perceptions covering the entire visual field.

Therefore, black and white images from a 108° head-mounted camera are cropped and digitized to an array of 32° × 64° field of view with one square degree pixels and 8 bit resolution. Next, thresholding is applied in order to further reduce the amount of visual information. In some tasks, image processing also includes edge detection. In real time, the processed image is then superimposed on the position coordinates of the phosphenes of the look-up table. If there is a coincidence between a phosphene position and any part of the image, the corresponding stimulation variable values are selected and send to the optic nerve stimulator. A list of the last 10 occurrences is continuously updated in order to avoid repeatedly inducing the same visual sensation when there are several coincident phosphenes. The least frequently used coincident phosphene is always chosen as the next stimulus. When a single phosphene is generated for each frame captured by the camera, 25 phosphenes can be induced per second.

The processor software and data can be accessed through a communication port (see Fig. 4) allowing to modify parameters, data tables, or the applied algorithms. For example, the random selection above could be replaced by a nearest neighbor selection or reference tables and parameters could be adapted to the perceptions reported by the volunteer. Some authors (45) have proposed a fitting optimization algorithm comparing the input images with the generated perception. This could work as an automatic processor training method. A major problem, however, is to make the subjectively perceived image available for comparison with the input counterpart. It is likely that much *a priori* knowledge will always have to be included in any system. A large share of that knowledge is still not available and will come, among others, from the contribution of first blind volunteers to the preliminary experiments.

From Image to Phosphene Production (Image Processing).

No matter what encoding algorithm is used, it is obvious that the amount of information that can be transferred on by present day visual prosthesis prototypes remains extremely poor compared to real world images. Although some improvement can be expected from further interface developments, severe image reductions will be unavoidable for quite some time.

Using some form of virtual reality in healthy volunteers, a few teams have explored the minimum requirements to obtain a useful pixelized vision according to the behavioral task involved. An array of ~600 dots is found to be a minimum for useful reading performance (46). Face recognition could be obtained with 10 × 10–32 × 32 grids (47). By contrast, some subjects can recognize simple objects and

symbols using a 4×4 pixel simulated prosthetic vision (48). Subject's performances clearly increase with training.

It should be stressed, however, that quite unlike the image pixels used in simulations phosphenes are not identical point-like ordered spots that neatly tile a surface. Also, these laboratory experiments only deal with experimental objects presented in an otherwise empty environment. In the real world, subjects will first have to localize and segregate target objects by some preprocessing. However, little has been achieved along these lines hitherto.

Vision Rehabilitation. The ultimate goal of a visual prosthesis is to rehabilitate a visually handicapped person. Results must therefore be evaluated from the blind person's performances point of view and not in terms of device features. Issues, such as the number of available phosphenes or their density, although contributing to the overall performance, cannot be considered as representative of the value of a prosthetic system. Furthermore, isolated analytic characteristics, such as the visual acuity, often mean very little because they can be adjusted by accessories, such as a straightforward optical compensation.

A measurement of the performances will most reliably be obtained in laboratory conditions, but it is the usefulness in real-life that will decide of the success of a prosthesis. For mobility, the prosthesis will be judged against available alternative obstacle detectors and rehabilitation means, such as the long cane or the guide dog in mobility tasks. Stationary visual tasks, such as object localization, identification, and grasping as well as character reading, face recognition, and scene identification are less likely to be available through alternative means. In such tasks, the error rate and the time to task completion will probably be major criteria for acceptance.

Some evaluations of visual prosthesis implant prototypes have been published (19), showing that light and movement can be detected and simple shapes recognized. With the optic nerve implant, basic patterns formed by bars of 22×320 mm and backprojected on a screen as black shapes against a white background can be recognized with the optic nerve prosthesis. The volunteer, sitting at a distance of ~ 0.5 m from the screen, uses scanning head (and hence camera) movements to explore the pattern, then draws the perceived figure using aluminum rods. A learning effect can be demonstrated as well as an improvement of the results with the number of elicitable phosphenes used in this test (49). A score of 87% of correct recognition is obtained with 109 phosphenes after training. Simultaneously, the task time decreased from >2 min to ~ 53 s (49). With the same system and after substantial training, the volunteer was able to localize, discriminate, and grasp objects on a table in front of them. Three among six familiar objects lay each in one of the nine subdivisions defined on a table surface. Grasping a specified object among the three was systematically successful in ~ 60 s. In both experiments, the scanning strategy probably explains the prolonged task completion times.

The emergence of multiple alternative designs and improvements for the prosthesis now call for evaluation standards. As suggested by these early results, these should include stationary tests (pattern identification,

figure orientation, object localization, object discrimination, and grasping) as well as mobility trials (obstacle localization and avoidance, landmark localization, and identification). Evaluation of the usefulness of the visual prosthesis in a natural environment will be essential. Further down the road to improvement, face recognition, scene identification, and finally reading tests will perhaps also be considered but much better resolutions are still required (29).

HUMAN AND MEDICAL ASPECTS

Candidates for a Visual Prosthesis

No matter what approach is chosen, all visual prosthesis system require a functional visual brain to ultimately interpret their output. This means that, with today's limited performances, only people losing sight after the critical period of development can be considered as suitable candidates.

Another selection criterion is the severity of blindness. As long as the performances of the systems are questionable, only totally blind persons should be considered as candidates. The risk of interfering with residual vision can indeed only be taken if the expected results are significantly better than the remaining visual abilities. In addition, the evaluation of the rather limited performances of an implant could be obscured by any surviving visual functionality.

Other selection criteria are dependent on the chosen approach. All the prechiasmatic approaches (subretinal, epiretinal, optic nerve) require the survival of retinal ganglion cells and their axons in the optic nerve. Terminal retinitis pigmentosa emerges as the condition most typically fulfilling all the selection criteria. Except in cases of brain lesions, the cortical approach would be more generally applicable, including in many cases of acute blindness where the psychological distress is usually more important. Finally, the individual's general health should also be considered because satisfactory candidates are usually terminal cases of RP and therefore rather old while the implantation surgery requires a prolonged anesthesia.

A complete assessment procedure must precede implantation.

A standard ophthalmologic examination is a good starting point. Because of the chronic nature of the condition, the diagnosis should be checked according to up-to-date knowledge. Some candidates have not had an ophthalmologic investigation for many years and only know that they have an incurable eye disease. The blind patient could be unaware of some additional eye problem or other interfering condition. A proper evaluation of the total degree of blindness is necessary and objective tests, such as absent VEP and ERG, are very useful for comparison with the postimplantation evaluation.

Taking into account the heterogeneous nature of retinitis pigmentosa, this diagnostic label is not enough to warrant the survival of ganglion cells in a given patient. Eyelid surface stimulation is the technique of choice here. A cathode is placed on the closed eyelid while an anode is stuck on the heterolateral mastoid (50). Small current

pulses allow to generate a phosphene perception. In healthy subjects, for pulses >2 ms duration, phosphene perception occurs well before the stimulus can be felt. A threshold strength–duration curve (rheobase of 0.28 mA, chronaxy of 3.07 ms in sighted subjects) can show the perceptions to be genuine. Electrically evoked potentials (51) is an alternative technique that would not have to rely on the patient's subjective perceptions. However, much of this added confidence is lost in important stimulation artifacts and the possible confusion with somatosensory potentials, especially in RP patients in whom thresholds are much higher. Some patients describe relatively abundant spontaneous phosphenes and these can be enhanced by the surface electrical stimulation. The same phenomenon could completely jeopardize the working of a visual prosthesis and perhaps induce permanent unpleasant symptoms.

A psychological evaluation is essential as long as the procedure remains experimental. People do accept the idea of pioneering research, but quite rightly want to make sure they will not be misused as guinea pigs for the sake of science alone. The motivation put forward is to help in the development of treatments or simply for the satisfaction of an active contribution or to give their grandchildren a better chance in the frame of their hereditary disease. Esthetic aspects are questioned right from the first contacts. Visibility of a camera is a point of concern.

When the visual prosthesis will have become a clinical treatment, expectations will still have to be confronted with the systems limitations. Also, much attention should be paid to human factors, such as the impact of an implant on a person's social integration.

Magnetic resonance imaging (MRI) is necessary for some visual prosthesis approaches. For example, the optic nerve size and diameter must be estimated on MRI images for an appropriate nerve cuff electrode to be selected for the optic nerve stimulation. The cortical approach might also benefit from a detailed anatomical image before surgery. An MRI examination could be dangerous and will yield very distorted images after implantation of a prosthesis. If for any reason, it is felt that such images will later be useful, then they should be acquired before implantation.

Classical presurgical investigations including thorax X rays and an electrocardiogram (ECG) are standard presurgical procedures.

Surgical Methods

The surgical method is very specific to each of the approaches. The cortical approach can obviously start with a standard craniotomy. Implantation of intracortical electrodes, however, can require a more specific method including specifically designed instruments such as the pneumatic insertion device (52).

The optic nerve approach has resorted to two kinds of surgical methods, one to place the electrode intracranially just in front of the chiasma and the other to implant the electrode in the orbit. Basically, the intracranial method uses a standard pterional transsylvian approach. That is, after right temporo-fronto-parietal cutaneous incision and preparation of the temporal muscle, a craniotomy is

performed at the meeting line between the great wing of the sphenoid with the frontal, parietal and temporal bone (pterion). Opening the dura gives access to the sylvian fissure. From there, surgery further proceeds under a microscope to carefully dissect the arachnoid, opening the sylvian fissure until, in the depth, the optic nerve can be separated from its surroundings. Only minimal retraction of the basal posterior aspect of the frontal lobe is required. The electrode is then wrapped around the optic nerve and the lead suture to the dura.

The second surgical method involves the implantation of the spiral cuff electrode around the intracortical optic nerve. A medial orbital approach is used. After detaching the internal rectus to allow careful retraction of the eye, a thread can be inserted behind the optic nerve and be used to pull the cuff in place.

The electrode leads exit temporally from the skull (intracranial implant) or from the orbit (intraorbital implant) and run backward under the scalp toward the neurostimulator half buried in a recessed well made in the parietal bone. The neurostimulator is also connected to the antenna inserted under the scalp above the mastoid.

The epiretinal system involves the implantation of a similar neurostimulator and wireless link unit as described above. The placement of the electrode is of course completely different and works (19) as follows (19): The periorbital space is reached through a lateral canthotomy. The cable and electrode are passed subconjunctivally all around the eye behind each of the four recti muscle insertions and then introduced into the eye through a 5 mm circumferential scleral incision placed 3 mm posterior to the limbus. Prior to the introduction of the implant, the majority of the vitreous gel is removed. The electrode array is the positioned just temporal to the fovea and a single retinal tack is inserted through the electrode array and into the sclera. The attachment of the electrode to the retina is a main issue here. Several solutions have been proposed including biocompatible glues, but recently developed types of miniature nail-like devices called retinal tacks appear to work well.

At least two different subretinal electrode implantation methods have been developed (15). The *ab interno* technique, follows established vitreoretinal and submacular surgery procedures. Surgical instruments are inserted into the eye. The vitreous body is then removed while the intraocular pressure is maintained with infusion. Finally, the retina is locally incised and the electrode array is inserted in the subretinal space using a special forceps. The *ab externo* implantation is designed to avoid damage to the internal structures of the eye. In this procedure, a flexible foil is inserted into the subretinal space through an incision in the sclera and choroid. The implant is then slit into a macular position along the guiding support foil. The path opened by this implantation can also be used for any required external energy supply leads.

Risks of Active Implants

A complete risk evaluation can only be performed on the basis of a review of a significant number of cases after a long postoperative period. Unfortunately, only a limited number

of human trials have been described so far and most of them using methods too different to allow any globalization. A risk analysis will thus have to consider the various aspects of the implantation in the light of similar procedures. The required anesthesia is in itself a well known, low but finite life threatening factor linked to the duration of the procedure as well as the age and general condition of the volunteer. The surgery as such is often mentioned as the most feared aspect by patients. For most of the prechiasmatic approaches, only structures of a nonfunctional organ exposed and the worst failure would thus result in the loss of the possibility to implant a new prosthesis. Risks linked to the visual prosthesis itself can probably be considered to be similar to those for cochlear implants where they are reported to be negligible (53). In some approaches (intracranial optic nerve and cortical), a craniotomy is performed and electrodes are placed in the direct vicinity or in the brain. In such cases, the possibility of an infection or of an abnormal inflammatory reaction or even direct damage to the brain are potential hazards of major consequences.

The possibility of an infectious metastasis around the foreign material must be borne in mind and is well known from passive implant applications. Similarly, experience with other active implants can help to evaluate the burden represented by possible electromagnetic interference, including airport or other safety systems, mobile telephone, interference between multiple implants and the fact that magnetic resonance imaging is no longer available to these patients.

The possibility of a total or partial failure of the implanted system must be considered. Again, however, useful figures, such as the typical lifetime of an implanted system and the percentage of failure in the initial period, cannot be estimated from the present heterogeneous and limited trials.

Heat production by implanted components and electrochemical changes at the level of the electrode contacts are potential hazards insofar that they are not easy to predict because many factors are involved. Safety limits for functional electrical stimulation are difficult to establish. Electrode failure or tissue damage are real risk factors. Also, unwanted stimulation of neighboring structures can lead, for example, to pain or abnormal muscular contractions.

Living tissues can suffer from the activation itself. Axonal potentials are propagated at the cost of cellular metabolic energy. Too strong a functional demand on some structures could create a state of imbalance between the energy demand and supply, leading to cell death. This type of limit will vary very much in different tissues. For example, peripheral motor axons typically discharge at frequencies ~ 20 Hz and could be damaged by chronic stimulation at 50 Hz while 100 Hz is a typical frequency for afferent activities in the optic and cochlear nerves.

The electrical stimulation could also induce more subtle changes. For example, with cortical stimulation, there is a possibility to trigger repetitive firing and even epileptic fits. For high stimulation currents, Brindley has indeed described phosphenes lasting minutes after the stimulation has ended (1). Such after-discharges could be minimized by reducing the stimulus charge and avoiding prolonged regular stimuli (24).

Psychological complications represent another possible issue that should be monitored. Blindness is indeed a severe disability to which most of the implantation candidates as well as their surroundings have adapted over time. An effective prosthesis will shake this equilibrium as well as the person's social insertion with consequences that could look paradoxical if only the technical success of the prosthesis was to be considered.

Ethical Aspects

With the project to fight one of the most basic human fears using high tech methods supposed to carry out miracles, the visual prosthesis is likely to enjoy a high profile to the layman and to trigger suspiciousness to scientists. This is thus a very emotional and sensitive subject that could be driven by many political and psychological forces alien to the interest of patients. That is why ethical questions should be dealt with most cautiously, especially in the early stages of development.

Basic ethical rules still derive from the Nuremberg code of 1947. As a consequence of the Second World War, it appeared that compliance with national laws could sometimes lead to unacceptable human behaviors. Some more fundamental ethical principles were raised above the law. This gave ethics a very special status. It is not a set of rules dictated by any form of power but pertains to every human being alike, above national or cultural differences (54). Ethics has neither organized a hierarchical structure nor an absolute reference. Progressively, from conferences to declarations, sets of principles gain universal acceptance.

The World Medical Association Declaration of Helsinki, now at its fifth revision since 1964 (55), is most often considered as the main reference. Laws in democratic countries as well as many organizations including scientific publishers enforce these basic principles. The European governments have extended these rules in the Convention on Human Rights and Biomedicine (56).

It is generally admitted that the implementation of ethical principles is very dependent on cultural factors. As a result, compliance of research projects with ethical principles is considered to be ideally insured by submission to an independent local ethical committee. National laws and institutional rules tend to organize the working of such committees. Typical questions investigated by these committees are the quality of the information to the volunteer, signature of an informed consent, the risk/benefit ratio to the volunteer, the evaluation of motivations and free decision as well as a specific insurance coverage including for removal of the implanted material if requested. An absolute preservation of the volunteers' private life is a must usually requiring anonymity.

The initial development of a visual prosthesis requires a prolonged collaboration with volunteers. In that frame, it was found that the organization in collaboration with the Ethics Committee in charge, of within project meetings between members of the experimental team and representatives of the volunteers could often be a very useful place to solve communication issues and take some consensus decisions with a volunteer.

A more general ethical question in human research is at what stage to move experiments from animals to humans

(57). Because human applications are the final goal, there is always a point where one will have to decide on the involvement of human volunteers. The prosthesis implantation is not trivial. Although blindness is a severe burden, the expected benefits are not life saving and for the time being, at best limited. The balance between risks and benefits is therefore not established. As a result, the informed consent and the volunteer's motivation become central issues. The adequate information to be provided to a lay candidate pertains to a very technical field and the many scientific uncertainties are even more difficult to explain. As long as the proposed visual implants are rather experimental devices, candidates for an implantation must be explicitly informed and accept that preliminary status. They should also be aware of the fact that the postimplantation fine tuning and rehabilitation might take much more time than what would be expected from a well-established treatment.

The volunteer's motivation on the other hand can be biased by difficulties to cope with a heavy handicap and the possible hereditary character of the disease involved. Time spent with a candidate volunteer can certainly solve many of the information and decision criteria issues, avoiding above all to misuse volunteers as mere study objects.

MAIN ISSUES

Evaluation

The systems presently implanted in human volunteers have the capacity of producing either 16 differently located phosphenes simultaneously (19) either >100 phosphenes send to the optic nerve at a pace of 25 phosphenes per second (49). At first sight, it might seem that the serial phosphene generation used in the optic nerve approach will end up in less information per unit of time than in the parallel scheme. The difference might perhaps not be as important as initially thought considering the power required to drive each stimulator output and the size of each stimulation circuit. In a stimulator with a large number of outputs, some multiplexing scheme is required that makes the system serial as well. The tradeoffs between the distance to target, the electrode contact area, the required stimulation charge, the power dissipation, the size of the current sources and the biologic tolerance might represent the real bottleneck. In 2005, despite measurable results, the performances of the visual prosthesis prototypes remain limited and achieved through an unpractical systematic scanning movement. No global image perception has been obtained so far. Learning effects have been demonstrated however and much can still be done on the side of image analysis to obtain an appropriately reduced amount of visual data. A more optimistic view can perhaps be derived from the past experiences with the now well established and accepted cochlear implants. Based on the number of papers published, the visual prosthesis seems to lag ~20 years behind its auditory counterpart.

Hurdles and Limitations

The number of pixels is not the only issue in the evaluation of the quantity of visual data that can be transferred. In an

image, each pixel also carries luminance and color information. Although the phosphene perceptions generated can be colored, no approach has hitherto developed means to control it and this dimension can therefore not be exploited. The brightness of phosphene perceptions is modulated by the stimulus intensity, but intensity also modulates the spatial recruitment and therefore, localization and brightness are not independent. In the optic nerve approach, strong stimuli will yield only large phosphenes centrally located in the visual field. With the retinal approaches a stronger stimulus increases the perceived luminosity but is likely to stimulate different structures and interfere with the neural code. In the case of stronger stimuli on the visual cortex surface, Brindley reported the appearance of additional phosphenes and their persistence up to 2 min after the stimulus ceased. Until now, epiretinal or optic nerve approaches do not explicitly take into account matters, such as the existence of ON as well as OFF ganglion cells. Stimulating more than a single cell or axon perhaps leads to some cancellation of their perceptive effect. In addition, the electrical stimulation may interfere with ongoing spontaneous firing of the ganglion cells. The link between stimulus and light perception can thus be expected to be rather complex.

The number of phosphenes that can be generated is still by far not large enough to come to a realistic pixelization. On the other hand, the central visual network does not build an internal projection of the outside world images as implied by such principles. The photosensors in the eye capture images as pixels, but that is also where the pixel structure stops. All other layers of the retina analyze and encode the image according to many characteristics other than a mere pixel position. Beyond a rough retinotopy, edge enhancement, movement detection, and a few other known features, the spatiotemporal encoding appears to be very complex. Except for the subretinal approach, recreating a natural image perception will require to interface a prosthesis with this complex and largely unknown code. This hurdle is likely to be even more challenging for cortical implants than for epiretinal or optic nerve approaches.

Another difficulty is that as long as a near perfect perception will not be achievable, bilateral implantation will not be able to convey distance information. Other distance clues are also linked to a relatively high quality vision. This means that before visual prosthesis reach a very high degree of performance, indirect systems will be required to provide the important distance information.

Another direction for future developments is set by the need for image stabilization. The subretinal approach uses the eye optics with the result that eye movements are included in the normal physiological network of vision. In all other approaches, an external camera is used. This camera is attached to spectacles and follows head movements instead of eye movements. For a correct localization of the surroundings, the prosthesis user must learn to inhibit eye movements as far as possible and only use slow head movement to scan the environment. This is of course a severe limitation. Eye and head movement tracking methods will at some point be necessary to stabilize images. The most advanced solution would be to implant or integrate a miniature camera in the eye as well. A feedback system taking into account eye movements in the image analysis

might be an intermediary step in the near future, before the intraocular camera becomes a reality.

Further progress in electrode design, stimulation algorithms, and electronic miniaturization will perhaps create the prospect of a system capable to generate a high density of phosphenes issued at cinematic rate. However, only that part of the visual field that is still matched by surviving ganglion cells will be available for rehabilitation, at least in all prechiasmatic approaches. As a result, although it is supposed to access the entire visual field, the optic nerve approach appears to reactivate only a restricted field of view. On the other hand, retinal approaches will have to match the electrode location with the remaining retinal cells in order to avoid an useless stimulation of death tissue. Clearly, terminal cases of retinitis pigmentosa will never regain a complete field using these methods. Better results can of course be expected in less affected patients where the prosthesis could be combined to some method to prevent further degradation.

Future Perspectives, Advanced Applications

The visual prosthesis in no way represents a treatment of the cause of blindness. The degenerative processes of retinitis pigmentosa can thus evolve further. The effect of chronic stimulation on sick ganglion cells is not known. The stimulation could either speed up the degeneration, have no influence on the natural course of the disease or, to the contrary, prevent further cell losses (58). The cause of retinitis pigmentosa lays in the biochemistry of the photosensor cells and the degeneration of the other cell layers of the retina could be considered as the result of disuse as well as the consequence of abnormal chemical intercellular exchanges. Keeping ganglion cells active might help them to survive. If this is true, as has been demonstrated in animals equipped with a cochlear implant (59), then the implantation should be done as early as possible, when a minimal number of bipolar and ganglion cells have degenerated. Recovery of a broader visual field can surely be expected in incomplete blindness stages but interference with existing remnants of vision might create new problems to be solved.

Again in keeping with the evolution of the clinical use of cochlear implants, there might be a good reason to foresee a visual prosthesis implantation in blind children. An efficient prosthesis implanted before the end of the critical period might perhaps induce the development of a functional visual cortex and preserve the individual's chances to later benefit from similar devices in adulthood. The limited efficacy of present day visual prosthetic devices does, however, clearly not yet justify such an undertaking.

BIBLIOGRAPHY

Cited References

1. Brindley GS, Lewin WS. The sensations produced by electrical stimulation of the visual cortex. *J Physiol* 1968;196(2): 479–493.
2. Easty LE, Sparrow JM. *Oxford Textbook of Ophthalmology*. Oxford (UK): Oxford University Press; 1999.
3. Hubel DH, Wiesel TN. The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *J Physiol* 1970;206(2):419–436.
4. Wanet-Defalque MC, et al. High metabolic activity in the visual cortex of early blind human subjects. *Brain Res* 1988;446(2): 369–373.
5. Marc RE, Jones BW, Watt CB, Strettoi E. Neural remodeling in retinal degeneration. *Prog Retin Eye Res* 2003;22(5):607–655.
6. Schmidt EM et al. Feasibility of a visual prosthesis for the blind based on intracortical microstimulation of the visual cortex. *Brain* 1996;119(Pt 2):507–522.
7. Santos A et al. Preservation of the inner retina in retinitis pigmentosa. A morphometric analysis. *Arch Ophthalmol* 1997;115(4):511–515.
8. Stone JL et al. Morphometric analysis of macular photoreceptors and ganglion cells in retinas with retinitis pigmentosa. *Arch Ophthalmol* 1992;110(11):1634–1639.
9. Chow AY et al. The artificial silicon retina microchip for the treatment of vision loss from retinitis pigmentosa. *Arch Ophthalmol* 2004;122(4):460–469.
10. Merabet LB et al. What blindness can tell us about seeing again: merging neuroplasticity and neuroprostheses. *Nat Rev Neurosci* 2005;6(1):71–77.
11. Iezzi R et al. Biocompatibility of Caging Chromophores for Use in Retinal and Cortical Visual Prostheses, ARVO Annual Meeting, Fort Lauderdale, FL, 5-5-2002. E-Abstract. Available at 4478; <http://www.iovs.org/search.dtl>.
12. Peterman MC et al. The Artificial Synapse Chip: a flexible retinal interface based on directed retinal cell growth and neurotransmitter stimulation. *Artif Organs* 2003; 27(11): 975–985.
13. Peterman MC et al. Localized neurotransmitter release for use in a prototype retinal interface. *Invest Ophthalmol Vis Sci* 2003;44(7):3144–3149.
14. Kohler K, Hartmann JA, Werts D, Zrenner E. Histological studies of retinal degeneration and biocompatibility of subretinal implants. *Ophthalmologie* 2001;98(4):364–368.
15. Zrenner E. Will retinal implants restore vision?. *Science* 2002;295(5557):1022–1025.
16. Peyman G, et al. Subretinal semiconductor microphotodiode array. *Ophthalmic Surg Lasers* 1998;29(3):234–241.
17. Humayun MS, de Juan Jr. E. Artificial vision. *Eye* 1998; 12(Pt 3b):605–607.
18. Rizzo JF et al. Perceptual efficacy of electrical stimulation of human retina with a microelectrode array during short-term surgical trials. *Invest Ophthalmol Vis Sci* 2003;44(12):5362–5369.
19. Humayun MS et al. Visual perception in a blind subject with a chronic microelectronic retinal prosthesis. *Vision Res* 2003;43(24):2573–2581.
20. Nakauchi K et al. Transretinal electrical stimulation by an intrascleral multichannel electrode array in rabbit eyes. *Graefes Arch Clin Exp Ophthalmol* 2005;243(2):eFIRST-6 Dec 2004.
21. Veraart C, Grill WM, Mortimer JT. Selective control of muscle activation with a multipolar nerve cuff electrode. *IEEE Trans Biomed Eng* 1993;40(7):640–653.
22. Veraart C et al. Visual sensations produced by optic nerve stimulation using an implanted self-sizing spiral cuff electrode. *Brain Res* 1998;813:181–186.
23. Branner A, Normann RA. A multielectrode array for intrafascicular recording and stimulation in sciatic nerve of cats. *Brain Res Bull* 2000;51(4):293–306.
24. Hambrecht FT. Visual prostheses based on direct interfaces with the visual system. *Baillieres Clin Neurol* 1995;4(1):147–165.

25. McCreery DB, Agnew WF, Bullara LA. The effects of prolonged intracortical microstimulation on the excitability of pyramidal tract neurons in the cat. *Ann Biomed Eng* 2002;30(1):107–119.
26. Jones KE, Campbell PK, Normann RA. A glass/silicon composite intracortical electrode array. *Ann Biomed Eng* 1992;20(4):423–437.
27. Normann RA et al. High-resolution spatio-temporal mapping of visual pathways using multi-electrode arrays. *Vision Res* 2001;41(10–11):1261–1275.
28. Bradley DC et al. Visuotopic mapping through a multi-channel stimulating implant in primate V1. *J Neurophysiol* 2005;93(3):1659–1670.
29. Weiland JD, Liu W, Humayun MS. Retinal prosthesis. *Annu Rev Biomed Eng* 2005;7:361–401.
30. Gosalia K, Weiland J, Humayun M, Lazzi G. Thermal elevation in the human eye and head due to the operation of a retinal prosthesis. *IEEE Trans Biomed Eng* 2004;51(8):1469–1477.
31. Delbeke J, Oozeer M, Veraart C. Position, size and luminosity of phosphenes generated by direct optic nerve stimulation. *Vision Res* 2003;43(9):1091–1102.
32. Dobelle WH. Artificial vision for the blind. The summit may be closer than you think. *ASAIO J* 1994;40(4):919–922.
33. Normann RA, Maynard EM, Rousche PJ, Warren DJ. A neural interface for a cortical vision prosthesis. *Vision Res* 1999; 39(15):2577–2587.
34. Jones KE, Normann RA. An advanced demultiplexing system for physiological stimulation. *IEEE Trans Biomed Eng* 1997; 44(12):1210–1220.
35. Brummer SB, Robblee LS, Hambrecht FT. Criteria for selecting electrodes for electrical stimulation: theoretical and practical considerations. *Ann N Y Acad Sci* 1983;405:159–171.
36. Weiland JD, Anderson DJ, Humayun MS. *In vitro* electrical properties for iridium oxide versus titanium nitride stimulating electrodes. *IEEE Trans Biomed Eng* 2002;49(12 Pt 2): 1574–1579.
37. Mahadevappa M, et al. Perceptual thresholds and electrode impedance in three retinal prosthesis subjects *IEEE Trans Neural Syst Rehabil Eng* 2005;13(2):201–206.
38. Humayun MS, Prince M, de Juan Jr. E, Barron Y, Moskowitz M, Klock IB, Milam AH. Morphometric analysis of the extramacular retina from postmortem eyes with retinitis pigmentosa. *Invest Ophthalmol Vis Sci* 1999;40(1):143–148.
39. Humayun MS, et al. Visual perception elicited by electrical stimulation of retina in blind humans. *Arch Ophthalmol* 1996;114(1):40–46.
40. Jensen RJ et al. Thresholds for activation of rabbit retinal ganglion cells with an ultrafine, extracellular microelectrode. *Invest Ophthalmol Vis Sci* 2003;44(8):3533–3543.
41. Greenberg RJ et al. A computational model of electrical stimulation of the retinal ganglion cell. *IEEE Trans Biomed Eng* 1999;46(5):505–514.
42. Rizzo III JF et al. Methods and perceptual thresholds for short-term electrical stimulation of human retina with micro-electrode arrays. *Invest Ophthalmol Vis Sci* 2003;44(12): 5355–5361.
43. Gekeler F et al. Subretinal electrical stimulation of the rabbit retina with acutely implanted electrode arrays. *Graefes Arch Clin Exp Ophthalmol* 2004;42(7):587–596.
44. Delbeke J et al. The microsystems based visual prosthesis for optic nerve stimulation. *Artif Organs* 2002;26(3):232–234.
45. Eckmiller R. Learning retina implants with epi-retinal contacts. *Ophthalmic Res* 1997;29(5):281–289.
46. Sommerhalder J et al. Simulation of artificial vision: II. Eccentric reading of full-page text and the learning of this task. *Vision Res* 2004;44(14):1693–1706.
47. Thompson Jr. RW, Barnett GD, Humayun MS, Dagnelie G. Facial recognition using simulated prosthetic pixelized vision. *Invest Ophthalmol Vis Sci* 2003;44(11):5035–5042.
48. Hayes JS et al. Visually guided performance of simple tasks using simulated prosthetic vision. *Artif Organs* 2003;27(11): 1016–1028.
49. Veraart C et al. Pattern recognition with the optic nerve visual prosthesis. *Artif Organs* 2003;27:996–1004.
50. Delbeke J et al. Electrical stimulation of anterior visual pathways in retinitis pigmentosa. *Invest Ophthalmol Vis Sci* 2001;42(1):291–297.
51. Potts AM, Inoue J, Buffum D. The electrically evoked response of the visual system (EER). *Invest Ophthalmol* 1968;7(3):269–278.
52. Rousche PJ, Normann RA. A method for pneumatically inserting an array of penetrating electrodes into cortical tissue. *Ann Biomed Eng* 1992;20(4):413–422.
53. Arnoldner C, Baumgartner WD, Gstoettner W, Hamzavi J. Surgical considerations in cochlear implantation in children and adults: a review of 342 cases in Vienna. *Acta Otolaryngol* 2005;125(3):228–234.
54. Pellegrino ED. Intersections of Western biomedical ethics and world culture: problematic and possibility. *Camb Q Health Ethics* 1992;1(3):191–196.
55. T.W.M.A.I WMA. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2000;284(23):3043–3045.
56. S.C.o.B CDBI. (1996). The convention for the protection of human rights and dignity of the human being with regard to the application of biology and medicine: Convention on human rights and biomedicine. Secretary General of the Council of Europe, Strasbourg. [Online.] Available at <http://conventions.coe.int/Treaty/EN/cadreprincipal.htm>.
57. Troyk P et al. A model for intracortical visual prosthesis research. *Artif Organs* 2003;27(11):1005–1015.
58. Leake PA, Hradek GT, Snyder RL. Chronic electrical stimulation by a cochlear implant promotes survival of spiral ganglion neurons after neonatal deafness. *J Comp Neurol* 1999;412(4):543–562.
59. Snyder RL et al. Chronic intracochlear electrical stimulation in the neonatally deafened cat. i: expansion of central representation. *Hear Res* 1990;50(1–2):7–33.

Reading List

- Alteheld N, Roessler G, Vobig M, Walter P. The retina implant—new approach to a visual prosthesis. *Biomed Tech (Berlin)* 2004;49(4):99–103.
- Archambeau C, Delbeke J, Veraart C, Verleysen M. Prediction of Visual Perceptions with Artificial Neural Networks in a Visual Prosthesis for the Blind. *Artif Intell Med* 2004;32(3):183–194.
- Bagnoud M, Sommerhalder J, Pelizzone M, Safran AB. Necessary visual information for restoring reading with a retinal implant in a blind patients with massive retinal degeneration of photo-receptors. *Klin Monatsbl Augenheilkd* 2001;218(5):360–362.
- Bak M et al. Visual sensations produced by intracortical microstimulation of the human occipital cortex. *Med Biol Eng Comput* 1990;28(3):257–259.
- Beebe X, Rose TL. Charge injection limits of activated iridium oxide electrodes with 0.2 ms pulses in bicarbonate buffered saline. *IEEE Trans Biomed Eng* 1988. 35(6):494–495.
- Berardi N, Pizzorusso T, Maffei L. Critical periods during sensory development. *Curr Opin Neurobiol* 2000;10(1):138–145.

- Besch D, Zrenner E. Prevention and therapy in hereditary retinal degenerations. *Doc Ophthalmol* 2003;106(1):31–35.
- Blakemore C, Van Sluyters RC. Reversal of the physiological effects of monocular deprivation in kittens: further evidence for a sensitive period. *J Physiol* 1974;237(1):195–216.
- Chow AY, Peachey NS. The subretinal microphotodiode array retinal prosthesis. *Ophthalmic Res* 1998;30:195–196.
- Chow AY, Peachey N. The subretinal microphotodiode array retinal prosthesis II. *Ophthalmic Res* 1999;31(3):246.
- Chow AY et al. Subretinal implantation of semiconductor-based photodiodes: durability of novel implant designs. *J Rehabil Res Dev* 2002;39(3):313–321.
- Chowdhury V, Morley JW, Coroneo MT. An *in-vivo* paradigm for the evaluation of stimulating electrodes for use with a visual prosthesis. *ANZ J Surg* 2004;74(5):372–378.
- Chowdhury V, Morley JW, Coroneo MT. Surface stimulation of the brain with a prototype array for a visual cortex prosthesis. *J Clin Neurosci* 2004;11(7):750–755.
- DeMarco SC et al. Computed SAR and thermal elevation in a 0.25 mm 2D model of the human eye and head in response to an implanted retinal stimulator. Part I: models and method. *IEEE Trans Antennas Propagat* 2002;(May 30):1–10.
- Elfar SD, Cottaris NP, Iezzi R, Abrams GW. Rapid Mapping of Cortical Multi-Electrode Arrays and Its Application for the Evaluation of Retinal Prostheses, Annual meeting of the Association for Research in Vision and Ophthalmology, Ft. Lauderdale, Florida, E-Abstract 3403. Available at <http://www.iovs.org/search.dtl> 2004.
- Fang X et al. Direct stimulation of optic nerve by electrodes implanted in optic disc of rabbit eyes. *Graefes Arch Clin Exp Ophthalmol* 2005;243(1):49–56.
- Fernandez E, Ferrandez J, Ammermuller J, Normann RA. Population coding in spike trains of simultaneously recorded retinal ganglion cells. *Brain Res* 2000;887(1):222–229.
- Finkelstein D, et al. Visual prostheses and visual rehabilitation in low vision research, assessment, and management. *Curr Opin Ophthalmol* 1991;2(6):729–732.
- Greenberg RJ et al. Electrical stimulation of the human retina: an update. *Invest Ophthalmol Vis Sci* 1995;36(4):S234.
- Grumet AE. Electric stimulation parameters for an epi-retinal prosthesis, Ph.D. dissertation. Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1999.
- Grumet AE, Wyatt Jr. JL, Rizzo III JF. Multi-electrode stimulation and recording in the isolated retina. *J Neurosci Methods* 2000;101(1):31–42.
- Guyen D et al. Long-term stimulation by active epiretinal implants in normal and RCD1 dogs. *J Neural Eng* 2005;2(1): S65–S73.
- Hallum LE, Suaning GJ, Lovell NH. Contribution to the theory of prosthetic vision. *ASAIO J* 2004;50(4):392–396.
- Hubel DH, Wiesel TN. Ferrier lecture. Functional architecture of macaque monkey visual cortex. *Proc R Soc London B Biol Sci* 1977;198(1130):1–59.
- Humayun MS et al. Pattern electrical stimulation of the human retina. *Vision Res* 1999;39(15):2569–2576.
- Humayun MS. Intraocular retinal prosthesis. *Trans Am Ophthalmol Soc* 2001;99:271–300.
- Jensen RJ, Ziv OR, Rizzo III JF. Thresholds for activation of rabbit retinal ganglion cells with relatively large, extracellular microelectrodes. *Invest Ophthalmol Vis Sci* 2005;46(4):1486–1496.
- Johnson L et al. Electrical stimulation of isolated retina with microwire glass electrodes. *J Neurosci Methods* 2004;137(2):265–273.
- Kanda H et al. Electrophysiological studies of the feasibility of suprachoroidal-transretinal stimulation for artificial vision in normal and RCS rats. *Invest Ophthalmol Vis Sci* 2004;45(2):560–566.
- Kato S, Saito M, Tanino T. Response of the visual system evoked by an alternating current. *Med Biol Eng Comput* 1983;21(1):47–50.
- Kerdran YA et al. Development and surgical implantation of a vision prosthesis model into the ovine eye. *Clin Experiment Ophthalmol* 2002;30(1):36–40.
- Lakhanpal RR et al. Advances in the development of visual prostheses. *Curr Opin Ophthalmol* 2003;14(3):122–127.
- Laube T et al. Chronically implanted epidural electrodes in Gottinger minipigs allow function tests of epiretinal implants. *Graefes Arch Clin Exp Ophthalmol* 2003;241(12):1013–1019.
- Laube T et al. Optical energy transfer for intraocular microsystems studied in rabbits. *Graefes Arch Clin Exp Ophthalmol* 2004;242(8):661–667.
- Lazzi G et al. Computed SAR and thermal elevation in a 0.25 mm 2D model of the human eye and head in response to an implanted retinal stimulator. Part II: results. *IEEE Trans Antennas Propagat* 2002;(May 30):1–8.
- Liu W et al. Electronic visual prosthesis. *Artif Organs* 2003;27(11):986–995.
- Loewenstein JI, Montezuma SR, Rizzo III JF. Outer retinal degeneration: an electronic retinal prosthesis as a treatment strategy. *Arch Ophthalmol* 2004;122(4):587–596.
- Margalit E et al. Retinal prosthesis for the blind. *Surv Ophthalmol* 2002;47(4):335–356.
- Marr D, Poggio T. Cooperative computation of stereo disparity. *Science* 1976;194(4262):283–287.
- Maynard EM, Nordhausen CT, Normann RA. The Utah intracortical Electrode Array: a recording structure for potential brain-computer interfaces. *Electroencephalogr Clin Neurophysiol* 1997;102(3):228–239.
- Maynard EM. Visual prostheses. *Annu Rev Biomed Eng* 2001;3:145–168.
- Montezuma SR, Rizzo III JF, Ziv OR. Differential recovery of the electroretinogram, visually evoked cortical potential, and electrically evoked cortical potential following vitrectomy: Implications for acute testing of an implanted retinal prosthesis. *J Rehabil Res Dev* 2004;41(2):113–120.
- Normann RA et al. The Utah 100 microelectrode array: an experimental platform for cortically based vision prosthesis, ARVO abstract # 192-B103. *Invest Ophthalmol Vis Sci* 1997;38(4):S41.
- Palanker D et al. Migration of retinal cells through a perforated membrane: implications for a high-resolution prosthesis. *Invest Ophthalmol Vis Sci* 2004;45(9):3266–3270.
- Pardue MT et al. Neuroprotective effect of subretinal implants in the RCS rat. *Invest Ophthalmol Vis Sci* 2005;46(2):674–682.
- Pardue MT et al. Possible sources of neuroprotection following subretinal silicon chip implantation in RCS rats. *J Neural Eng* 2005;2(1):S39–S47.
- Parrini S, Delbeke J, Legat V, Veraart C. Modelling analysis of human optic nerve fibre excitation based on experimental data. *Med Biol Eng Comput* 2000;38(4):454–464.
- Peterman MC et al. Fluid flow past an aperture in a microfluidic channel. *Anal Chem* 2004;76(7):1850–1856.
- Piyathaisere DV et al. Heat effects on the retina. *Ophthalmic Surg Lasers Imaging* 2003;34(2):114–120.
- Potts AM, Inoue J, Buffum D. The electrically evoked response of the visual system (EER). *Invest Ophthalmol* 1968;7(3):269–278.
- Radner W et al. Increased spontaneous retinal ganglion cell activity in rd mice after neural retinal transplantation. *Invest Ophthalmol Vis Sci* 2002;43(9):3053–3058.
- Rizzo III JF et al. Retinal prosthesis: an encouraging first decade with major challenges ahead. *Ophthalmology* 2001;108(1):13–14.
- Rizzo III JF et al. *In vivo* electrical stimulation of rabbit retina with a microfabricated array: strategies to maximize responses for prospective assessment of stimulus efficacy and biocompatibility. *Restor Neurol Neurosci* 2004;22(6):429–443.

- Sachs HG, Gabel VP. Retinal replacement—the development of microelectronic retinal prostheses—experience with subretinal implants and new aspects. *Graefes Arch Clin Exp Ophthalmol* 2004;42(8):717–723.
- Sakaguchi H et al. Electrical Stimulation with a Needle-type Electrode Inserted into the Optic Nerve in Rabbit Eyes. *Jpn J Ophthalmol* 2004;48(6):552–557.
- Sakaguchi H et al. Transretinal electrical stimulation with a suprachoroidal multichannel electrode in rabbit eyes. *Jpn J Ophthalmol* 2004;48(3):256–261.
- Schwahn HN et al. Studies on the feasibility of a subretinal visual prosthesis: data from Yucatan micropig and rabbit. *Graefes Arch Clin Exp Ophthalmol* 2001;239(12):961–967.
- Sommerhalder J et al. Pelizzone, Simulation of artificial vision: I. Eccentric reading of isolated words, and perceptual learning. *Vision Res* 2003;43(3):269–283.
- Suanning GJ, Lovell NH, Schindhelm K, Coroneo MT. The bionic eye (electronic visual prosthesis): a review. *Aust N Z J Ophthalmol* 1998;26(3):195–202.
- Sugiyama T et al. Optic cup enlargement followed by reduced optic nerve head circulation after optic nerve stimulation. *Invest Ophthalmol Vis Sci* 2001;42(12):2843–2848.
- Suzuki S et al. Comparison of electrical stimulation thresholds in normal and retinal degenerated mouse retina. *Jpn J Ophthalmol* 2004;48(4):345–349.
- Thylefors B, Negrel AD, Pararajasegaram R, Dadzie KY. Global data on blindness. *Bull World Health Organ* 1995;73(1):115–121.
- Troyk PR, Schwan MA. Closed-loop class E transcutaneous power and data link for microimplants. *IEEE Trans Biomed Eng* 1992;39(6):589–599.
- Troyk PR. (2003). Multichannel transcutaneous cortical stimulation system. [Online.] Available at <http://npp.ninds.nih.gov/npp/sow/stm256.htm?format=printable>.
- Uhlig CE, Taneri S, Benner FP, Gerding H. Electrical stimulation of the visual system. From empirical approach to visual prostheses. *Ophthalmologie* 2001;98(11):1089–1096.
- Veraart C et al. Vision rehabilitation in the case of blindness. *Expert Rev Med Devices* 2004;1(1):139–153.
- Walter P et al. Successful long-term implantation of electrically inactive epiretinal microelectrode arrays in rabbits. *Retina* 1999;19(6):546–552.
- Warren DJ, Fernandez E, Normann RA. High-resolution two-dimensional spatial mapping of cat striate cortex using a 100-microelectrode array. *Neuroscience* 2001;105(1):19–31.
- Weiland JD et al. Understanding the origin of visual percepts elicited by electrical stimulation of the human retina. *Graefes Arch Clin Exp Ophthalmol* 1999;237(12):1007–1013.
- Weiland JD, Humayun MS. Past, present, and future of artificial vision. *Artif Organs* 2003;27(11):961–962.
- Wilms M, Eger M, Schanze T, Eckhorn R. Visual resolution with epi-retinal electrical stimulation estimated from activation profiles in cat visual cortex. *Vis Neurosci* 2003;20(5):543–555.
- Yamauchi Y et al. Comparison of electrically evoked cortical potential thresholds generated with subretinal or suprachoroidal placement of a microelectrode array in the rabbit. *J Neural Eng* 2005;2(1):S48–S56.
- Yanai D et al. The value of preoperative tests in the selection of blind patients for a permanent microelectronic implant. *Trans Am Ophthalmol Soc* 2003;101:223–228.
- Zhou DD, Greenberg RJ. Microsensors and microbiosensors for retinal implants. *Front Biosci* 2004;10:166–179.
- Ziv OR, Rizzo JF, Jensen RJ. *In vitro* activation of retinal cells: estimating location of stimulated cell by using a mathematical model. *J Neural Eng* 2005;2(1):S5–S15.
- Zrenner E et al. The development of subretinal microphotodiodes for replacement of degenerated photoreceptors. *Ophthalmic Res* 1997;29(5):269–280.
- Zrenner E et al. Subretinal implants. *Ophthalmic Res* 1998;30:197–198.
- Zrenner E et al. Reply to the letter of drs. Chow and peachey: the subretinal microphotodiode array retinal prosthesis II. *Ophthalmic Res* 1999;31(3):247.
- Zrenner E et al. Can subretinal microphotodiodes successfully replace degenerated photoreceptors?. *Vision Res* 1999;39(15):2555–2567.

VOCAL REHABILITATION. See LARYNGEAL PROSTHETIC DEVICES.

X-RAY EQUIPMENT DESIGN

MARTIN TORNAL
Duke University
Durham, North Carolina

INTRODUCTION

Radiography is a form of physical diagnosis in which X rays are used to obtain medically useful information about a patient. X rays are generated in a controlled way, and a beam of X rays is passed through the patient. The differential absorption of ionizing X-ray radiation by different tissues modulates this beam. The transmitted beam is detected, and its information content is recorded. In addition, radiological imaging methods are used to control, guide, and monitor both diagnostic and therapeutic manipulations such as needle biopsy, percutaneous transluminal angioplasty (see CORONARY ANGIOPLASTY), and radiation therapy.

X-ray equipment is composed of many subsystems and components, each of which presents unique choices and tradeoffs for the equipment designer. These components include means for X-ray production, spatial and spectral shaping of the X-ray beam, patient handling, as well as means for image detection and capture, intermediate image handling, image processing, image display, and data storage. Auxiliary equipment and logic are customarily included in the design and construction of X-ray apparatus for the control, synchronization, and automation of the examination process. The X-ray equipment designer must balance different and often conflicting requirements to obtain an optimum solution for each examination category or application. For example, fine spatial detail requires minimization of both geometric blur (implying a fine focal spot along with low power and low magnification) and motion blur (implying short exposure time with high power and a large focal spot). This leads to specialization of the equipment for different medical applications. In addition to clinical requirements, the apparatus must also be designed to meet a wide variety of industrial and professional standards as well as comply with governmental regulations.

The functional organization of an X-ray imaging system is shown in Fig. 1. In one form or another, these elements are present in all X-ray imaging systems. This article will follow the logical flow shown in Fig. 1, with emphasis given to the material contained in the double-sided boxes. Other articles in this Encyclopedia discuss the remaining topics.

Power Components

The primary purpose of the power components in the generator is to control the X-ray tube voltage, current, and exposure time. The contrast and signal-to-noise ratio (SNR) of a radiographic image, as well as dose delivered to the patient, are partially controlled by adjustments of the

X-ray tube voltage. Other influencing factors include the X-ray beam filtration, the X-ray attenuation characteristics of the object (e.g., patient), and the nature of the image receptor. The voltage applied to the tube has a major influence on the shape of the X-ray spectrum emitted from the tube (Fig. 2). This change in spectral shape as voltage is varied produces changes in image quality, including the common image metrics of contrast and SNR (see X RAYS, PRODUCTION OF). An increase in voltage leads to decreased image contrast (see X-RAY ATTENUATION IN MATTER) and increased SNR. In addition, the total quantity of radiation emitted from the tube is a power function of the voltage applied across the tube (approximately a cubic function, depending on the voltage range and beam filtration), resulting in increasing dose to the patient. In the United States, it is common to use voltage as the primary control element for regulating the amount of X rays exiting from the patient (termed penetration).

The amount of radiation emitted from a tube is also controlled by the product of the current flowing through the tube and the time during which the current flows (expressed in units of milli-ampere-seconds, or mAs). Use of exposure time is somewhat limited in an attempt to minimize patient motion. Applied current can greatly increase tube heating and reduce tube life, causing tube cooling techniques to be a major design consideration, especially in applications such as tomography, where high rates of tube exposures are experienced due to the requirement for multiple sequential images to be produced and detected. Tube current is controlled by adjustment of the X-ray tube's filament temperature. In most generators, time is controlled by switches in the primary circuits. Some generators use switching tubes in the high-tension circuit of the generator, or a control element in the X-ray tube. Figure 3 is a simplified schematic diagram of conventional single-phase, line-operated generator. Such a system draws its power directly from the main electrical distribution system. Conventional generators may be designed to operate using either single- or three-phase power. When one considers power-line impedance and generator self-loading, single-phase equipment is usually limited to lower peak power devices, whereas three-phase equipment may be found at any power level. The consequences of power-line loading are discussed briefly at the end of this section.

The need for high-power mobile X-ray generators in hospitals has led to the construction of equipment incorporating energy storage devices, either capacitors or batteries. These generators accumulate energy from a low-capacity power line at appropriately low-power levels, and then they discharge the stored energy through the X-ray tube at a much higher power level. The currently preferred design is the battery-operated mobile generator with a high-frequency inverter.

Line-operated X-ray generators place unusual loads on the hospital's electrical distribution system. Most generators draw X-ray tube power directly from the lines during

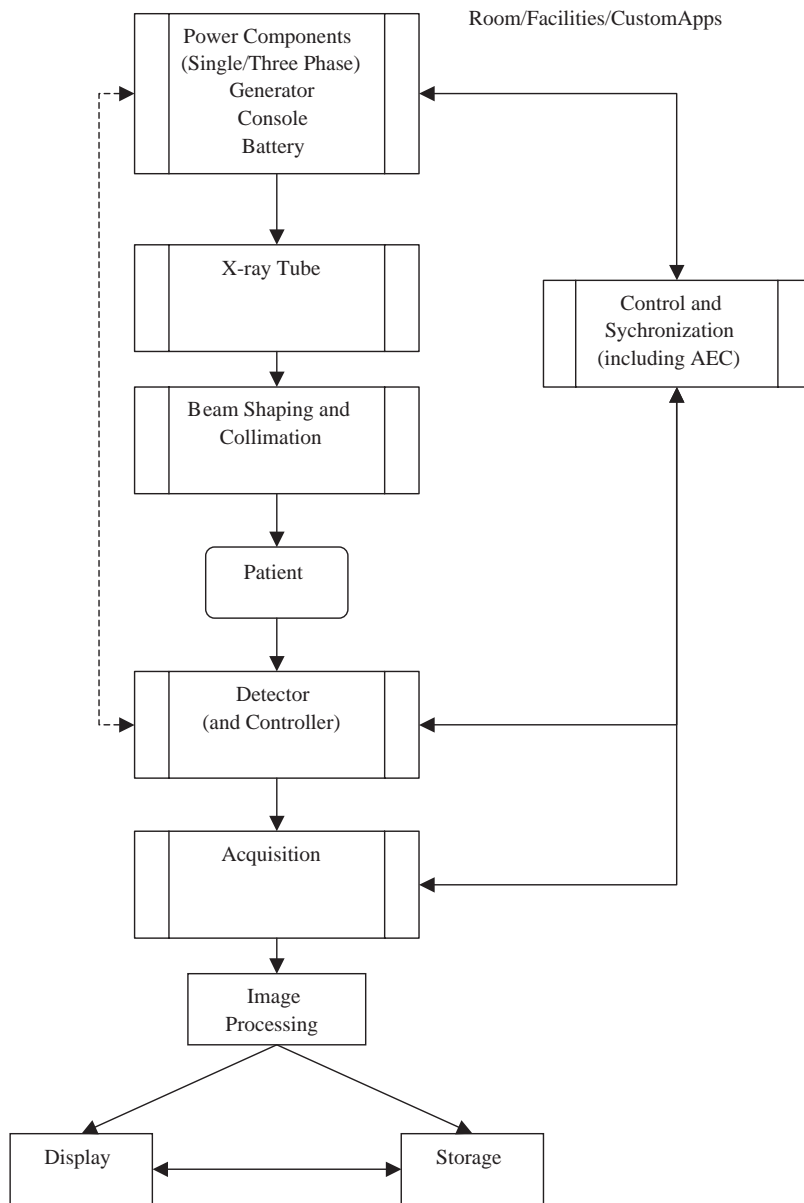


Figure 1. Functional diagram of an X-ray imaging system. Any system used for medical imaging must provide the minimum facilities for delivering a controlled X-ray beam to the patient, positioning the patient relative to the beam, detecting the modulated X-ray beam passing through the patient, and delivering the image to the radiologist.

the time of X-ray exposure. Classic X-ray circuits (Fig. 3) set the X-ray tube voltage by means of measurement of the primary voltage before exposure. The combination of these two effects requires that the power lines be well conditioned and protected from extraneous loads. In addition, other equipment in the hospital must be protected from the effects of the brief high-power demands placed on the line by the X-ray equipment. Also, in this era of ubiquitous digital control, two-way isolation is required to protect both the X-ray equipment and the remainder of the hospital from the broadcast of digital noise, or real digital control signals along the power lines. Power for X-ray generators should be obtained from separate low-impedance supplies, which include dedicated distribution transformers and appropriate size conductors. To minimize conductor size, high-power equipment is usually connected using a higher line voltage. Line-operated X-ray equipment, with the exception of low-power mobile units, is seldom operated

from 110 V supplies. Single-phase and low-power three-phase equipment typically operates using a 220 V line. Higher power equipment may require 440/480 V.

An undesirable characteristic of some X-ray generators is ripple, defined as the percent variation of voltage during an X-ray exposure. A high ripple factor can result in undesirable X-ray flux and spectral variations leading to unpredictability in image quality and patient dose. Single-phase, full-wave rectified generators suffered from very large ripple factors, and therefore they have been virtually replaced by the constant potential generator for the most demanding imaging applications, medical imaging being one of them. Although these types of generators are generally larger, more complex, and expensive, they provide extremely low (<5%) ripple factors, and they are essentially constant in their voltage output. These generators typically use single- or three-phase 50/60 Hz line voltages of 220 V or higher. One common type of constant potential

X-ray Spectra at Various Tube Potentials

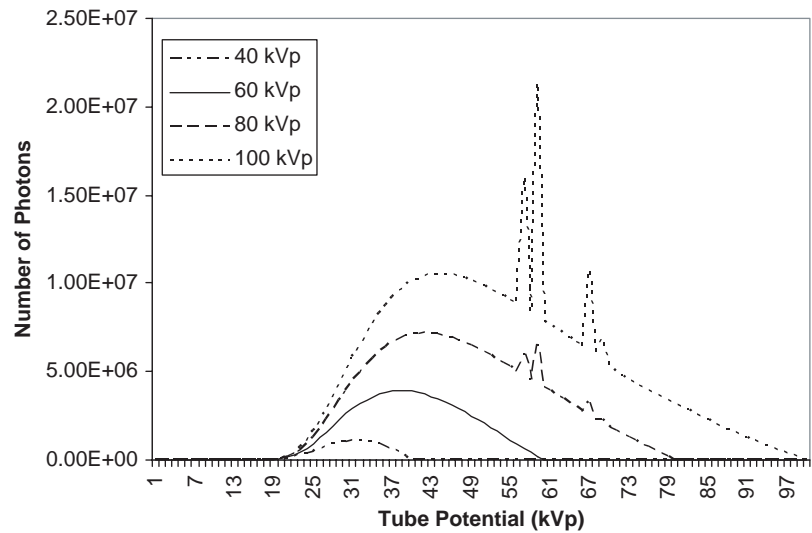


Figure 2. Sample spectra, generated via validated computer simulation, using various tube potentials, 1 mAs exposure, and with 0.5 cm thick Al extrinsic filtration, illustrate the effect on beam shape.

generator is the three-phase generator, using a three-phase transformer with three sets of primary and secondary windings. The newest type of constant potential generator is the medium or high-frequency generator (also called the inverter generator), where the constant potential voltage is controlled by the frequency of the current provided by the chopper/inverter. This generator type is popular both in battery-powered and in mobile units, as well as in the higher power applications of modern equipment. Starter speed is also a consideration available to designers. The starter controls the boost and operation of the rotating anode. Typical low-speed anodes boost in approximately 3 seconds and reach a rotation speed of 3000 rotations per minute (rpm). High-speed starters are often available as add-ons that allow faster boost as well as rotation speeds in excess of 10,000 rpm. Most generators also include a thermal switch that will automatically stop the X-ray exposure

when an overheat signal is received from the tube. Additional tube details are provided in the following section.

X-Ray Tubes

The normal X-ray tube is a vacuum diode. Figure 4 is a photograph of a typical X-ray tube housing as well as a cutaway diagram of the typical interior tube components. Electrons are generated by thermionic emission from the filament of the tube. The electron stream is electrostatically focused onto a small target on the anode by means of a carefully designed, shaped, and sometimes charged, highly polished nickel cathode focusing cup.

X rays are produced by interactions between the electrons and the target (see X RAYS, PRODUCTION OF). Most electrons emitted by the hot filament become current carriers across the tube. One can, therefore, set tube current

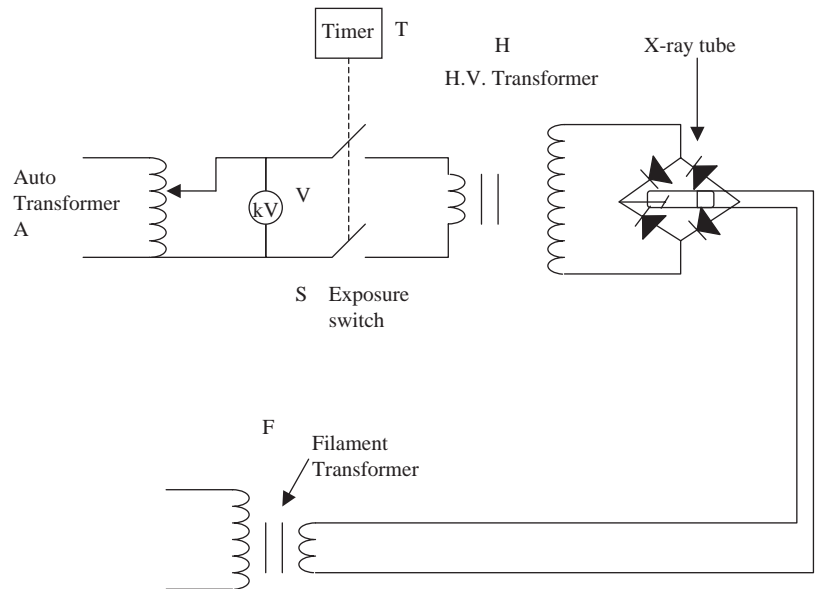


Figure 3. Schematic diagram of a single-phase, full-wave rectified generator. Feedback and stabilization elements are not shown. X-ray tube voltage is adjusted by setting the autotransformer before exposure. Tube current is adjusted by setting the filament voltage and hence its temperature. A, Autotransformer; H, high-voltage transformer; F, filament transformer; S, exposure switch; T, radiographic timer; V, prereading voltmeter used for kV indication.

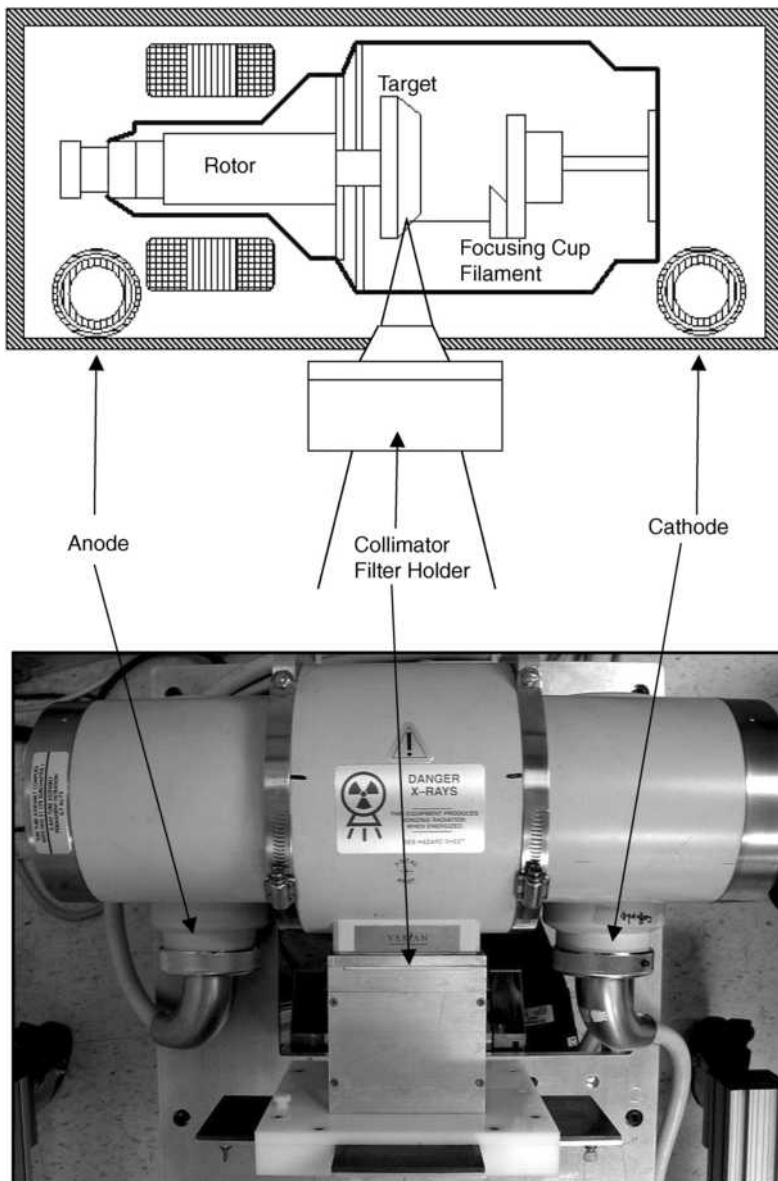


Figure 4. Diagram of a generic X-ray tube and photograph of tube enclosed in its protective housing. The useful X-ray beam emerges from the beam port at the bottom center. The anode is rotated by coupling magnetic fields through the glass envelope of the tube. The stator coils that are used for this purpose surround the tube on the left.

(by adjusting the filament temperature) and tube voltage (by adjusting primary voltage) as independent parameters. Some X-ray tubes function as a triode with a bias voltage applied between the filament and the cathode cup. At low voltage, this bias can be used to modify the size and shape of the focal spot by focusing the electron beam in the tube. A further increase of the bias voltage can serve as a switching device by gating the electron beam ON or OFF.

The X-ray tube converts a very small fraction of the electrical energy delivered to it into useful X rays. In the voltage region corresponding to diagnostic radiology (~20–150 kVp), the physical processes involved in X-ray production (primarily collisional deceleration of electrons in the target) result in an X-ray production efficiency of <1%. The remaining electrical energy is converted into heat. Tungsten is used as the target material for most general-purpose diagnostic X-ray tubes. Its high atomic number ($Z = 74$) maximizes X-ray production efficiency. Tungsten's high melting point (3400 °C) and reasonable specific heat

capacity ($C_p = 130 \text{ J/kg-K}$) help with the thermal problems associated with the waste heat. As special cases, molybdenum (Mo), rhodium (Rh), or Mo/Rh alloy targets are used to produce the specific X-ray spectra needed for mammography. The use of these lower atomic number metals as targets and filters isolates the K-characteristic X rays produced by the target, resulting in a narrow spectral beam at relatively low energy, useful for improving subject contrast in screen-film imaging of the compressed breast. The lower efficiency of X-ray production and lower melting point makes these metals unacceptable for general radiography. Other materials, including silver, cerium, and other exotic materials, continue to be proposed for target materials for specialized applications ranging from small animal imaging to computed tomography and additional industrial applications.

A major task in X-ray tube design is the provision of means for dissipating waste heat before tube structures are damaged. Different tube structures place different thermal

limitations on radiographic techniques and ultimately limit the speed of performing examinations. Tube rating charts that account for these different thermal restrictions are available from all X-ray tube manufacturers. The first consideration is the melting of the tungsten target at the focal spot, the target area bombarded by the electron beam. For sharp projection imaging, the geometry of the X-ray "optics" requires that the focal spot be as small as possible. Design elements such as high-speed rotating anodes (3,000–10,000 rpm) and shallow target angles (5°–15°) are used to limit the power density incident on the physical target location while creating a small effective focal spot, yielding improved resolution. Effective short-term loading in excess of 50 kW/mm² of effective focal spot is realized using these techniques. Some tubes even provide user-selectable dual-anode angles or dual-position focal spots for increased flexibility.

Once heat has entered the target region of the anode, it is conducted away from the focal track into the bulk material of the anode. This bulk material usually consists of molybdenum or prolytic graphite. Massive anodes (several kilograms) are used to temporarily store the thermal energy before radiating it to the tube housing and eventually into the environment. The large anode mass, needed for significant thermal energy storage (e.g., for applications in computed tomography and cinefluorography), places limitations on the starting time of the tube, due to the moment of inertia of the anode, and on the length of life of the anode's bearings. The technical requirement of high anode heat storage capacity is optimally met with a heavy smaller diameter disk. The clinical requirement for short radiographic exposure time demands high instantaneous power levels. This requirement is optimally met with a light large-diameter rotating anode disk. These conflicts result in the design and use of many types of X-ray tubes mechanically specialized to meet specific examination requirements.

The tube housing serves several technical purposes. Figure 4 (top) is a diagram of such a tube in its protective housing. The housing is part of the electrical isolation between the high-voltage circuits and the environment. It also provides radiation protection for both patient and operator. Tube housings are lead lined to keep the amount of leakage radiation below legal limits (this requirement assures that the major source of irradiation outside of the beam comes from scatter and the useful beam in the patient). The tube housing illustrated is a variety commonly used, whereby the low atomic number (usually Beryllium) exit window for the X rays is located in the center of the tube between the two poles (cathode and anode) of the power supply from the generator. It is also possible to obtain unipolar X-ray tubes for special applications. With only one single pole for the power supply, it becomes possible to locate the exit window nearer the edge of the tube to allow closer proximity imaging in special-purpose applications, such as in computed mammatomography. Also, these tubes are smaller and lighter than standard dual pole tubes. Drawbacks of such tubes include much lower maximum power (i.e., on the order of ~100 W to 6 kW, compared with the 100+ kW available on other tubes), static anode and, consequently, the need for a larger

focal spot to reduce concentration of heat on the anode. These tubes are usually liquid cooled.

Finally, the tube housing is a key portion in the waste-heat handling system. Housings for tubes used at low mean power levels (<100 W or so) can be adequately air cooled, with or without a fan. As the mean load increases (i.e., for applications such as angiography and computed tomography), air cooling becomes inadequate. Additional cooling may be obtained by circulating liquid through a heat exchanger contained in the tube housing or by circulating insulating oil through an external radiator.

Collimation and X-Ray Beam Definition

The geometry of projection X-ray imaging requires that the produced X-ray photons be directly projected from the focal spot onto a detector system. Radiation produced elsewhere in the X-ray tube (off-focal radiation) and scattered radiation from the vicinity of the tube contribute to geometric blur in images. Scattered radiation emanating from the vicinity of the patient, patient supports, and image receptor may not cause observable blur but will still degrade image SNR and contrast. Patient integral radiation dose is minimized, and image quality is maximized by confining the beam of radiation to the smallest possible area. Patient dose and radiological contrast are also influenced by the X-ray spectrum. This section will discuss means for controlling the spatial extent and spectral shape of the X-ray beam.

The production of scatter is reduced by limiting the spatial extent of the X-ray beam by means of a device called a collimator on the source side. The conventional radiograph is produced using a two-dimensional, four-sided cone of radiation. Minimizing the projected beam size by restricting it to the area of interest reduces both patient dose and image degradation due to scatter. Most systems are equipped with adjustable source collimators. Specialized equipment may be found that uses fixed collimation. Modern collimators, both fixed and adjustable, use several sets of apertures to minimize off-focal radiation. X-ray photons emerging from the tube whose real or virtual source is not the focal spot are termed off-focal radiation. Off-focal radiation is produced by electrons scattering from the focal spot with sufficient energy to produce X rays when they collide with some other portion of the anode structure. The use of metallic or partially metallic tubes, in place of all-glass tubes, reduces the production of off-focal radiation by providing alternative return conduction paths for scattered electrons. X-ray photons produced in the focal spot or elsewhere in the tube may be internally scattered and seem to be produced anywhere in the tube. Simple, single-plane diaphragms limit the size of the X-ray beam emerging from the beam port, but they are not very efficient in minimizing the emission of off-focal radiation and internal scatter. Some tubes also include a hood around the anode to absorb off-focal radiation and prevent it from exiting the tube.

In the United States, the law requires that the collimator and the image receptor must be linked so that the maximum settable beam size does not exceed the dimensions of the receptor. The influence of scatter can be further reduced by collimating to and using a one-dimensional fan

of radiation or a pencil beam of radiation, provided that the scatter geometrically misses the image receptor or is attenuated before it reaches the receptor. The former method is accomplished by having a small scanning detector aligned with the useful beam (e.g., in slot scanning). The latter technique involves the use of a secondary collimator between the patient and image receptor that scans synchronously with the X-ray beam. Slot scanning systems have recently become commercially available. These systems use an X-ray fan beam synchronized with a moving collimated slot of detectors to perform a planar patient scan. The effect is to provide a digital radiograph with significantly reduced scatter fractions (relative to grid systems) compared with conventional commercial cone beam systems. One drawback is a slightly longer scan with potentially increased patient motion artifacts, as well as other possible electrical or mechanical artifacts due to the use of moving parts.

The energy spectrum of the X-ray beam emerging from the target is modified (filtered) by its passage through extrinsic filtration before it enters the patient. The X-ray spectrum is further modified as it passes through the patient. These spectral modulations are due to the inherent energy dependence of the X-ray attenuation coefficients, which are intrinsic characteristics of an elemental or compound material.

An improper selection of the target material or filter, relative to the examination being performed, can result either in diminished image quality or excessive patient dose. A certain amount of filtration is beneficial to the patient. Low-energy photons that enter and are absorbed by the patient's tissues contribute to dose without contributing to the image. These photons may be removed from the X-ray beam by inserting additional filters between the tube and the patient. High-purity aluminum is the filter material that is commonly used for general radiography. Because the attenuation coefficient of aluminum decreases across the radiographic energy range, aluminum preferentially absorbs the low-energy photons emitted from the target and is therefore said to "harden" the X-ray beam.

Beam quality is usually described in terms of its half-value layer (HVL) of aluminum. This is the amount of aluminum needed to reduce the beam intensity by 50%. The U.S. Public Health Service has placed minimum HVL requirements on diagnostic X-ray beams. Too high an HVL is also undesirable as it leads to poor SNR in screen-film imaging systems. Alternatively, the molybdenum filter used for mammography functions in a different manner. As pure metals are relatively transparent to their own characteristic radiations, a molybdenum filter passes most of the K-characteristic radiation from a molybdenum target. At an appropriate tube voltage, the filter preferentially removes both low- and high-energy photons. Such a matched filter-target combination results in the transmission of a narrow energy band of X rays.

Several other methods are available for generation of narrow X-ray spectra with an inherent improved dose efficiency when the mean energy is matched to the object and imaging task. Use of heavy (i.e., 7–10 HVLs) K-edge filtration (Fig. 5) can provide a practical, low-complexity, and relatively inexpensive means of using heavy attenua-

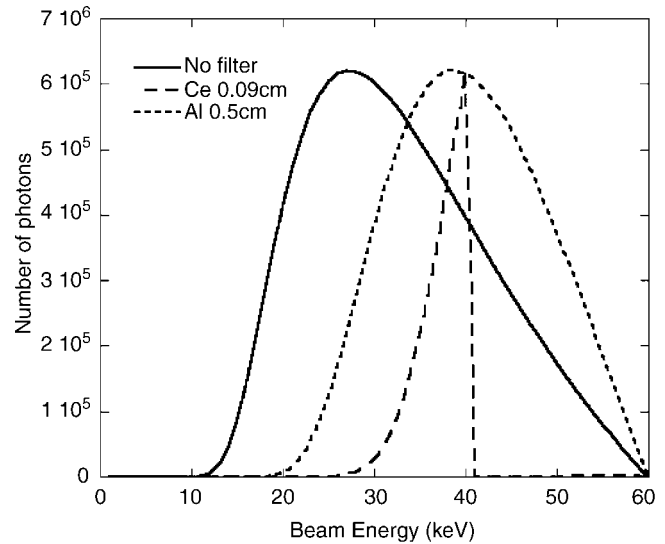


Figure 5. Sample spectra generated via simulation using 60 kVp tube potential, 1 mAs exposure with various extrinsic filters to illustrate effect on beam shape. Illustrated are spectra for no filtration, 0.5 cm thick aluminum, and 0.09 cm thick cerium extrinsic filtration. These represent unfiltered, minimally filtered (similar to standard filtration in commercial systems), and heavily K-edge filtered beams. Note that the cerium filtered beam is quasi-monochromatic (i.e., 7% full width at half the maximum height), which can, by design, provide maximal dose efficiency in a uniformly attenuating, homogeneous object.

tion for beam hardening and near-complete elimination of lower energy photons and using the K-edge to nearly completely eliminate higher energy photons, thereby producing a narrow, quasi-monochromatic beam. Higher tube exposures are necessary, accompanied by increased object heating and reduced tube life, although the actual object exposure may be similar to that with regular filtration. Bragg diffraction can also be used to produce quasi-monochromatic beams. Although Bragg diffraction can produce a very narrow beam, devices using Bragg diffraction suffer from longer scan times because only narrow fan beams can be obtained; indeed, the object needs to be moved relative to the beam because sweeping the beam would compromise beam quality. Synchrotron light sources have also been proposed for monochromatic radiation but so far have proved far too expensive and impractical for routine medical imaging applications.

Bowtie filters can be employed within the collimator attached to X-ray tubes (Fig. 6, top). These filters conform to the general contour of the patient in that they are thinnest (i.e., allow maximum flux) at the center where the patient is thickest and become thicker near the edge (i.e., allow minimum flux) where the patient contour gets thinner. This allows reduction of dose to the patient while providing a more uniform optical density (in the case of screen-film) or SNR (in the case of digital detection) across the image.

The interactions of X rays with the patient's tissues and other structures in the beam path result in the production, by the Compton process, of copious quantities of scattered radiation. A portion of this scattered radiation is directed

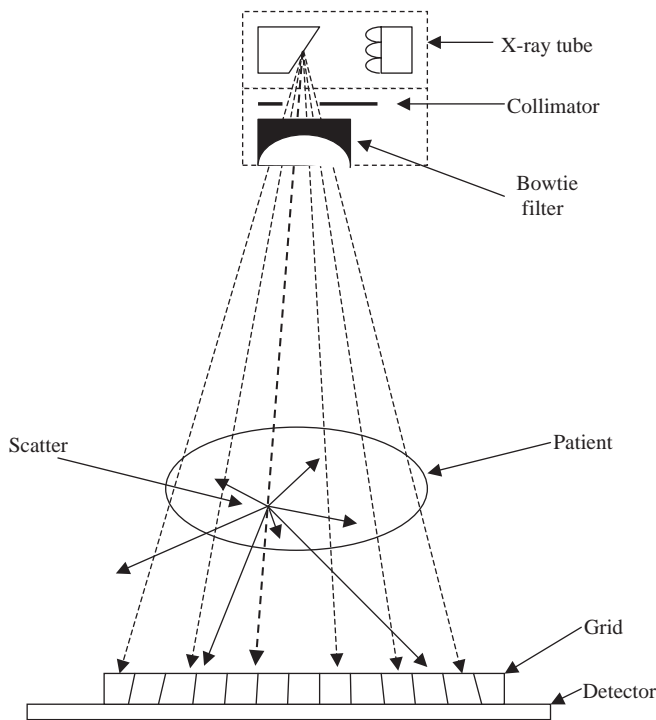


Figure 6. Schematic of the radiation trajectories and various components of interaction in the X-ray generation and detection system. Near the source is the source collimation apparatus. Distal to the source, the radiographic grid consists of thin lead strips, usually aligned with the focal length of the X-ray tube. Useful imaging photons “see” narrow grid strips and wide interspaces. Scattered photons see wider lead strips; hence, they are more likely to be absorbed by the grid.

toward the image receptor. The scattering process destroys the imaging geometry, which consists of straight lines from the focal spot to the image receptor. Scatter does not contribute to the useful imaging process, but because of its diffuse nature, scatter produces a generally uniform background intensity, reducing radiologic contrast, and decreasing the SNR of the image. In chest, for example, surprisingly most (80–90%) X rays on film are scattered radiation, although the image quality of a chest radiograph is still good and full of fine detail.

The use of a grid (Fig. 6) exploits the lack of geometrical coherence of scatter and helps to remove it. The lead strips of the grid are geometrically aligned toward the focal spot of the X-ray tube. Unscattered photons “see” only narrow grid strips and relatively wide grid interstitial spaces. Only a small fraction of these photons (the photons useful in the imaging process) are absorbed by the grid. Scattered photons produced in the patient have a wider angular distribution. On average these scattered photons see much wider grid strips and much narrower interstitial spaces. Thus, a grid selectively removes scattered radiation.

When a grid is used, however, a comparatively higher skin dose is required for the same SNR. The improvement in image quality achieved with a grid usually justifies the increased patient dose. Grids are constructed with thin lead or tungsten strips, and interstitial spaces are composed of a variety of uniformly thick materials including

aluminum, organic material (solid or foam), or air. Grids are available in many different aspect ratios (the height of the strips divided by the space between adjacent strips). A higher ratio grid will reject more scattered radiation (necessitating higher incident flux to maintain SNR) but requires more precise positioning than a lower ratio device. Grids, especially coarser ones, can also be moved during the exposure so that the grid lines are blurred, and essentially eliminated by the motion. The combination of a grid and its movement mechanism is referred to as a Potter-Bucky diaphragm. Grids are now available with varying focal lengths and material compositions and can be constructed by extrusion, layering, or chemical etching.

One functional drawback to the grid is that the solid geometry (of the lines of sight) is fixed, so the source to image (SID) cannot be changed without changing the grid. Several alternatives to the anti-scatter grid are available to the designer for dealing with scatter, some or all of which can be used in combination. These include the use of air gaps between the object and imaging device, the slot scanning technique, and algorithmic correction (including the beam stop and deconvolution methods). Employment of algorithmic correction is especially attractive in digital systems.

Image Receptors

Although digital detector usage is increasing greatly, conventional film and film-screen systems are still the most widely used of all radiologic image receptors (see SCREEN-FILM SYSTEMS). The radiographic cassette contains a relatively high density and high atomic number scintillating screen for converting X-ray energy to visible light. Visible light provides better film exposure than the direct interaction of X rays in the silver halides embedded in film. When irradiated, the screen fluoresces and exposes the film to light. Most film blackening results from this fluorescent light. A screen is held in close contact with the film in a light tight cassette so that there is an optimal transfer of the light to the film, with minimal resolution blur caused by the bloom of light generated in the scintillator. Modern film-screen systems have a detective quantum efficiency (DQE) around 50% (DQE is a measure of the ability to convert SNR presented to the film into recorded SNR and includes metrics of resolution, noise, and efficiency). This means that about half of the ultimate information in the modulated X-ray beam is detected by the system. The sensitivity (speed) of film-screen systems may be chosen by selection of screen and film types. Systems that provide proper film blackening at low dose either exhibit a great deal of quantum noise or mask the noise by blurring both the noise and the image detail. This is determined by selection of screen type. Scintillators in the screen can be amorphous (the majority of cases) or highly structured in finely grown pillars. The shape of the gray scale transfer function of the film-screen system is determined by the choice of film. Ideally, the optimum visible absorption efficiency of the film corresponds to the optimum light wavelength output of the screen, although each of these absorption and emission functions vary with wavelength. Different film characteristics are often needed for the

optimal performance of different radiographic examinations. For example, a slow film speed film with finer spatial resolution may be more useful for extremity imaging of fine bone fractures, whereas a faster speed film with coarser resolution is more useful for chest imaging where stopping or minimizing the influence of cardiac motion is desired.

There are both advantages and disadvantages to the development and use of digital X-ray detectors, but a primary advantage is that there is a decoupling of the image capture and image display process. In addition to separating image capture from image display, the digital representation of the radiographic image is key to the successful implementation of a picture archiving and communications system (PACS) system, which has already shown to have a beneficial impact on the delivery of health care. With advances in large-scale miniaturization and packaging, and production of large-area electronics structures with high yield, there are continual advances in production and manufacture of digital detectors with increasingly finer spatial resolution. Indeed, there is a projected 600% growth in demand of digital radiography and digital mammography systems by 2008, along with declining demand for screen-film and computed radiography (CR) systems sales. Digital systems are expected to surpass other film-screen based X-ray imaging systems in the near future. Nevertheless, a large number of these systems currently exist and are part of contemporary radiography.

Photostimulable phosphor systems were introduced several years ago and remain popular in radiographic practice. In a CR system, an image storage plate "traps" an image of the attenuated X-ray distribution with excited state electrons. When such a plate is scanned with a pencil-tip-sized laser beam, the electrons are de-excited and return to their ground state with the emission of visible, phosphorescent radiation. This phosphorescence can be measured, digitized, and subsequently used to produce an image (fluorescent scintillation systems and phosphorescent systems differ in the decay time τ of the scintillation event, where fluorescence occurs in $\tau \sim 10^{-9}$ s, and phosphorescence can be considerably longer, lasting $\tau \sim$ hours or days). Such CR systems are capable of reproducing >1 ; 5 line per millimeter with an acceptable DQE. The phosphors used in these systems have a linear dose-to-photoluminescence response extending over several orders of magnitude of absorbed dose. With proper reading technique, this intrinsic linearity permits a decoupling of dose from film blackening. Intentional or inadvertent exposure errors can be compensated for by the reading electronics, thus reducing retakes. Dose may be selected with regard to the image quality requirements of the individual examination.

The X-ray-sensitive image intensifier, using charge coupled devices (CCDs), is currently the universal image receptor for fluoroscopy and fluorography (see IMAGE INTENSIFIERS). This device converts the modulated X-ray beam into a minified visible light image projected onto a small, fine resolution (cooled) CCD. The conversion factor (image blackening per unit of X-ray input) of modern image intensifiers is high enough that the quantum sink in most imaging chains is the X-ray intensity detected and photoelectrons created at the tube's input screen. The DQE of

these detectors is also in the neighborhood of 50%. Most image intensifier-based image receptor systems use optical diaphragms between the output of the image intensifier tube and the input of the CCD. This diaphragm is used to restrict the light intensity reaching the CCD, thereby forcing enough X-ray intensity through the patient so as to produce a statistically meaningful image. In some systems, this diaphragm is externally adjustable, permitting the operator to set the dose to a level that is consistent with the allowable X-ray quantum noise for the examination.

Image intensifier systems commonly use video as the primary fluoroscopic display and recording medium. If the video is used only for fluoroscopic monitoring or for simple playback applications, normal broadcast video standards are employed. For systems that use the video signal for fluorographic applications, such as digital subtraction angiography (DSA) (see DIGITAL ANGIOGRAPHY), better quality nonstandard video formats may be used. A variety of photofluorographic cameras can be used for image capture and the production of hardcopy images (see CINE AND SPOT FILM CAMERAS).

Flat panel detectors are becoming increasingly popular for digital radiography (DR). Although they are more expensive than CR systems and may not be able to be retrofitted into existing gantry's, they provide much faster (i.e., instant) digital data acquisition and higher resolution (now on the order 50 μm or 20 lp/mm) than CR systems. DR detectors are generally divided into indirect and direct detector categories. With indirect detection, a scintillator (e.g., CsI, BGO, $\text{Gd}_2\text{O}_2\text{S}$, CdWO_4) is used to convert X rays into photons in the visible spectrum that are then detected by a thin-film transistor (TFT)-based photodiode array. With direct detection, X rays are detected and converted directly into electric charge by the detection layer (e.g., amorphous-Se, CdZnTe, amorphous-HgI₂) without the need for an intermediate scintillation event. There is an ongoing debate pertaining to the superiority of one approach versus the other. Suffice to say here that direct detectors can have finer spatial resolution leading to a better MTF, and until very recently, indirect detectors with higher overall stopping power had better detection efficiency and hence DQE.

With digital detectors of any kind, one must be concerned with image lag or ghosting characteristics that are specific to the type of detector chosen. Acquisition and/or correction techniques may have to be employed to minimize these residual image effects and generally do not affect image quality for planar projection images in the clinical setting. Tomographic systems that require multiple rapid image acquisitions are more likely to suffer from these degrading effects.

Control and Supervisory Logic

Radiological equipment produces X rays with tube voltages of 25 to 150 kV at power levels extending from less than 100 W to more than 100 kW. Exposure times range from a millisecond to several minutes. A wide voltage range is required to produce the required signal characteristics in the final image. Wide power and exposure time ranges are needed to deliver the appropriate dose rate and total dose

to the image receptor. In addition, digital detection systems require careful synchronization such that frames are not exposed during the frame readout period. High-speed frame-grabbing hardware is also required in specialized applications, such as tomography, where frame rates can be very high. Control, synchronization, regulation, and automation of such elements necessitate the introduction of control and feedback elements into X-ray systems (Fig. 1). In many cases, installed radiographic systems have their control primarily associated with the generator, especially in screen-film systems. For newer digital systems, including special applications such as tomography, the generator becomes a component under the logical control of the overall system host computer, which is responsible for controlling and managing all components of the system.

One example of such synchronization and control is illustrated by a timing diagram (Fig. 7). The diagram shows multiple exposure cycles as would be typical in a tomography application, using a digital detector, rotating anode X-ray tube, and movable gantry, although other implementations may exist. A signal must first be sent to the generator starter to initiate rotation of the anode. Once the anode has reached maximum speed, the detector is checked for status availability. Exposures can be initiated only if the detector is not in the process of reading out the panel. The period between readouts is known as the vertical blanking period. X rays can be exposed during this time (the length of this period depends on the frame rate of the detector). When the detector ready signal is verified, a signal is sent to the X-ray generator to expose X rays. After the X-ray pulse is complete, a signal is sent to the gantry to move the assembly into position for the next image, and simultaneously, a signal is sent to the frame grabbing circuit to acquire the image from the detector controller. Frame grabbers are specialized circuits responsible for quickly obtaining or "grabbing" the exposed frame from the digital detector.

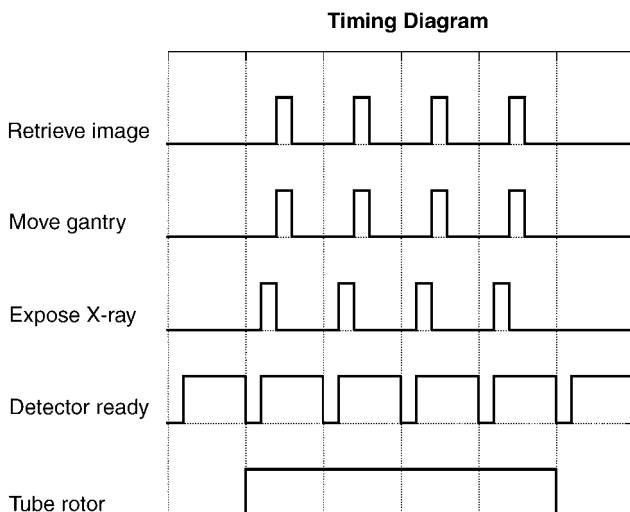


Figure 7. Timing diagram illustrating the necessity for synchronization when implementing an automated system that involves many different system components, including X-ray generator, digital detector, motion control devices, and frame grabbing devices.

The previous scenario gives maximal flexibility to the system designer. Many other options are also available. For example, the detector ready signal could be wired directly to the generator so that exposures are initiated automatically whenever the detector goes into a ready state. In addition, some tomography applications have the X-ray beam constantly on during the multiple acquisitions for tomography and, instead, use a timed, physical shutter for initiation of exposures.

The above scenario also assumes that exposure times are predetermined and will take place at the preset values when an expose signal is initiated. Another control option available on many systems is that of automatic exposure control (AEC). This involves the measurement of X-ray transmission and is used as a control element for generator switching. The commonly used feedback loop, which switches the X-ray beam at a predetermined dose, remains essentially the same as Morgan's original phototimer. Most AEC systems offer the choice of several different measuring fields, with the selection of active fields being examination and projection dependent. With proper positioning, this ensures proper exposure of critical anatomical detail. Modern systems adjust sensitivities within the AEC system to accommodate different film screen systems, compensate for reciprocity loss in radiographic film, select different measuring fields for different examinations and for other control functions. This provides a high degree of uniformity and image consistency via the automatic termination of an exposure for a predetermined desired effect. So the overall hope is that with AEC, the number of reshoots and/or call backs for additional exposures is reduced. The most common transmission measurement device used is an ionization chamber interfaced to the phototiming circuit.

The radiographic technique is the selection of appropriate X-ray exposure factors, taking into consideration patient size, shape, and physical condition, the examination and projection to be performed, and the available choice of radiological materials and supplies. This may, of course, all be done properly by a technologist using his or her own judgment with a manually controlled generator. However, this process has become less a matter of technologist preference and more a part of a departmental standard protocol, both in terms of required radiographic projections and technique factors. Many modern systems also provide anatomical programmed radiography (APR). APR typically provides the technologist with dozens of examination views for each of several different anatomical regions in a menu at the controller. By selecting a particular mode, APR automatically sets the required technique factors, including kVp, mA, time, focal spot, film speed, SID, and imaging receptor. In addition, the operator may input the actual patient thickness providing the optimal technique parameters for that examination.

Patient Support Structures

Patient safety, handling, and comfort are critical elements in the design of medical imaging equipment. The needs of the physician or technologist must also be taken into account. These are not trivial design problems. The engineer must consider, among other things, the patient's

physical condition, attached life support systems, required accessory devices (i.e., traction slings or intravenous bottles), and the patient's mental attitude. In addition, the ergonomics of the equipment should promote the expeditious performance of the examination. For example, the examination table should permit easy patient transfer with a minimum amount of staff assistance. The key component of each radiographic examination is the image receptor. To this end, the technologist must have direct and unimpeded access to the patient from a variety of directions. Nearly all of the technologist's tasks, with the exception of patient handling and positioning, can be automated. The technologist's ability to interact with a sick patient while obtaining the desired imaging results is a fine art. Equipment design and selection must be conducive to this process.

As many rooms often contain Potter-Bucky diaphragms (they are referred to as table and wall Bucky's, respectively), the room must be designed to carry such devices. A variant on the radiographic room is the use of a tilting table in place of the flat (nontilting) table. The X-ray tube is usually mounted on a ceiling suspension. A special-purpose apparatus may be used to improve the efficacy of positioning for complex examinations such as skull radiography and trauma-related procedures. C-arms may also be used and must be accommodated.

Another major requirement in room design is that of safety. Rooms must be shielded to prevent radiation from reaching unintended areas while allowing visibility between the operator, who is generally outside the permanent room or behind some radiation shielding, and the patient and scanner. Mobile leaded shields with various sized lead-doped glass are available, and some dedicated systems (e.g., in mammography) have attached operator shielding to minimize exposure from the multiple studies coordinated by the technologist. Thus, rooms must be designed to accommodate both the general requirements common to radiographic systems as well as special requirements for custom applications.

Custom Applications and Equipment Selection

The design of a radiology department in terms of the number and type of procedure rooms and the selection of equipment is an important consideration. The amount to which equipment can be differentiated depends on the size of the institution and the nature of its workload. It is a rare department that can maintain an even patient flow throughout the day. Most departments have a sharp peak early in the morning and a secondary peak in the early afternoon. Sufficient staff and equipment must be provided to handle these peaks.

Radiologic equipment is classified by its technical composition and its clinical function. The simplest apparatus is the dental, mobile, or portable generator. This consists only of the means for X-ray generation and beam control. Patient support, grids, image receptors, and so on, are not included.

In fluoroscopy, the rooms are almost universally equipped with radiographic components, a tilting table, an under-table X-ray tube, and an image intensifier. The under-table tube and image intensifier are linked together and mounted on a

spot film tower. The spot film device, located between the patient and the image intensifier, is a specialized cassette holder. Radiographic spot films, obtained during fluoroscopic examinations, are captured using the spot film device and conventional cassettes. The radiologist usually views the fluoroscopic image by means of closed-circuit television or on a computer screen. Fluorographic images are recorded using camera systems directly focused on the output of the image intensifier. The increasing use of fluoroscopic cameras makes the conventional spot film device obsolete.

Angiographic and other interventional or special-procedure laboratories are highly specialized rooms with film changers and cameras adapted to the needs of rapid imaging. Special procedures require especially easy access to the patient by the staff as well as the capability for multiple projection angles of the X-ray beam. The patient's medical condition and any attached life support devices dictate that these examinations be performed with as little patient handling as possible. In addition, the equipment should provide adequate radiation protection for the staff. The patient table and other mechanical equipment in these rooms are therefore highly specialized to meet the needs of different examinations. Under many circumstances, the apparatus permits simultaneous radiography from two directions (i.e., biplane imaging).

Tomographic imaging is another mechanically complex procedure found in radiographic rooms. In a historical tomographic apparatus, the film and X-ray tube were synchronously moved so that overlapping structures in the patient are removed. Tomographic apparatus ranged from attachments to simple Bucky tables to dedicated equipment capable of producing a wide variety of motions. Contemporary computed tomography is still mechanically complex, although manufacturers have gone to great lengths to enable easy patient and technologist access in a clutter-free system (see COMPUTED TOMOGRAPHY).

In radiotherapy, very high (MeV) tube potentials are used for radiation of tumors. These systems may include a separate but integrated keV imaging system for scouting the area for treatment planning before irradiation. MeV imaging systems are also being investigated for their ability to produce their own image without the need for a separate keV system. In the past, direct film detection was used because the incident energy and flux was high enough to sufficiently expose film, but newer flat panel digital devices are rapidly being adapted for this portal imaging, MeV application. Designers must take these into consideration.

Several other emerging areas including tomosynthesis, small animal imaging, computed mammotomography, and other dedicated or application-specific imaging technologies are developing rapidly. Thus, it behooves the designer to keep in mind potential future applications when considering equipment selection options. The primary changes have been rapidly occurring in imaging (i.e., detector) technologies, in moving from analog to digital systems, and the Internet technology associated with their image processing, sharing, and transfer. There are several initiatives to modify the X-ray sources for both improved focal spot characteristics (e.g., smaller, faster, switching capabilities) and target and filtration materials (i.e., for more optimal spectra used

in application specific imaging paradigms). To most appropriately synchronize and use these emerging, digitally based technologies, computer-based, centralized data control and acquisition technologies need to keep pace with these other emerging technologies. Only through properly using and synchronizing these various necessary X-ray system components can the lowest ionizing radiation doses be applied to obtain images that provide the most information content available to the clinician, whose proper interpretation will affect the patient under observation.

FURTHER READING

- Curry III TS, Dowdey JE, Murry Jr RC. Christensen's Introduction to the Physics of Diagnostic Radiology. 3rd ed. Philadelphia, PA: Lea & Febiger; 1984. A general overview of the physics and technology of diagnostic radiology. Many U.S. radiology residents use this book as their primary textbook and technical reference.
- Hendee WR, Chaney EL, Rossi RP. Radiologic Physics, Equipment and Quality Control. Chicago, IL: Year Book Med. Publ.; 1977. A textbook for advanced technologists.
- Johns HE, Cunningham JR. The Physics of Radiology. 4th ed. Springfield, IL: Thomas Publ.; 1983.
- Beutel J, Knudel HL, Van Metter RL. Handbook of Medical Imaging, Volume 1: Physics and Psychophysics. Bellingham, WA: SPIE Press; 2000.

See also X-RAYS, PRODUCTION OF; X-RAY THERAPY EQUIPMENT, LOW AND MEDIUM ENERGY.

X-RAY MAMMOGRAPHY. See MAMMOGRAPHY.

X-RAY QUALITY CONTROL PROGRAM

EDWARD L. NICKOLOFF
Columbia University
New York City, New York

INTRODUCTION

In order to improve or maintain an optimal achievable level of healthcare, most medical facilities usually have some type of quality assurance program. The quality assurance programs encompass a variety of functions, such as development of comprehensive safety rules, establishment of infection control guidelines, provisions for the efficient management of critically ill patients, establishment of accurate record-keeping functions, and the development of adequate educational training for the staff (1). In radiology, a portion of the overall quality assurance program is devoted to a quality control program for the X-ray equipment. The American College of Radiology (ACR) defines quality control (QC) programs as a periodic monitoring of aspects of the precision or accuracy of the X-ray equipment pertaining to the successful performance of the equipment, techniques, or tests rather than to the clinical decision making (2). In general, QC involves equipment rather than patient care.

Most X-ray QC programs have several very specific goals toward which they are directed. The list merely enumerates these items. (1) Image quality, (2) patient radiation doses, (3) safety, (4) consistency, (5) economic factors, (6) Regulatory requirements.

The type of QC testing performed is often dependent on the particular goals that are being evaluated. Usually, state or local regulatory agencies require that a QC program for X-ray equipment be conducted on a routine basis. Some regulatory agencies have very specific items that must be measured by a qualified expert or licensed medical physicist at least annually. These items may include typical patient doses (doserate) and generator calibrations. Moreover, the Federal government has published many documents in which suggested survey procedures for X-ray equipment are discussed (3–5). In 1992, the Federal government enacted the Mammography Quality Standards Act (MQSA), which requires: the accreditation of all mammography units, U.S. Food and Drug Administration (FDA) certification of mammography facilities, and at least annually physics QC testing of mammography equipment. In addition, the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO) mandates an adequate QC program that has qualified staff and maintains patient safety. Hence, QC programs for X-ray equipment are mandated by various regulatory agencies and accreditation organizations like the American College of Radiology (ACR) and JCAHO.

Some of the regulatory requirements, however, do not have a major impact upon image quality or economic factors. For example, one regulatory restriction requires that X-ray and light field be congruent with each other to within 2% of the source to image receptor distance (between the X-ray tube focal spot and the image receptor) (SID). Another regulatory requirement is directed toward checking the various interlocks in the X-ray system. Although these regulatory requirements affect radiation protection, they have little impact upon the operation of the equipment or upon image quality. Hence, some of the various QC tests mandated by regulatory agencies, and the results must comply with legal standards of performance.

Another reason for performing QC tests is to assess mechanical, electrical, and radiation safety of the X-ray equipment. These evaluations include inspections directed toward eliminating any electrical shock hazards or the failure of cables, locks, mechanical supports, or counterbalances that could result in the injury to patients or staff. The mechanical inspections require that the system components be both visually inspected and the components be manipulated to test their integrity. Radiation levels to the patients and staff must be measured to determine that there is adequate radiation safety and that radiation levels are not excessive.

Another feature of QC programs is to determine the functional performance of an X-ray system. Various controls, switches, and function buttons must be tested to make certain that they are operating properly. These tests are directed toward making certain that the equipment functions as it was designed. A common malfunction of X-ray equipment is a selector switch that indicates some change has been made when in actuality nothing happens.

The optimization of the image quality in radiology procedures is one of the most important reasons for QC programs. The QC programs establish a test procedure by which the image quality can be measured and establish a standard level of image quality one attempts to maintain (6). Thereafter, the image quality can be measured on a regular basis and corrective actions are taken as necessary. Degraded image quality can often hamper the proper clinical diagnosis and thereby have a severe negative impact on patient care. One would not like to deliver radiation doses to the patients that result in limited or no diagnostic information. Hence, a major goal of QC programs in diagnostic radiology departments would be the assessment of image quality.

Quality control programs must also attempt to maintain a consistent level of equipment performance. X-ray outputs that are not reproducible from one exposure to the next will result in over- and under-exposed images. Unusable radiographs result in unnecessary patient radiation dosages, delays in diagnosis and treatment, and economic losses in terms of technologist's and physicians time and film. Similarly, the technologists should be able to expect that adjacent rooms with similar X-ray equipment will have similar X-ray outputs and similar calibrations within reasonable tolerances attributed to differences in the models and manufacturer's designs of the equipment. Automated film processors should be adjusted to produce similar radiographs regardless of where the films are developed. Similarly, digital X-ray imaging like Computed Radiography (CR) and Digital Radiography (DR) must have consistent X-ray exposures, image processing, and display of images. Consistency of a given X-ray room and consistency in comparing the various X-ray equipment is essential to maintaining a standard level of radiographic image quality for the entire facility.

Finally, economic considerations are a major impetus for having an effective QC program. It has been estimated that each radiographic film that must be repeated costs the facility ~\$35 (1986 prices) (7). The cost of repeated films include items such as film prices, X-ray tube usage, X-ray equipment depreciation, chemistry usage, processor depreciation, technologist's and physician's time, and facilities cost (rent, heat, and electricity). A study performed by the federal government indicated that repeat rates can often be decreased by a factor of two by effective QC programs (8,9). Without QC programs, repeat rates can be expected to be 10–16%; with QC programs, repeat rates can be expected to be 5–8%. These repeat rates are due to patient motion, positioning errors, or other technologists' errors. Repeat rates can often be drastically reduced by the usage of charts in which the technical factors to be selected are listed for a variety of patient procedures and patient sizes. The development of these technique charts should be considered to be part of a good QC program. Furthermore, a good QC program often identifies and corrects equipment malfunctions before they become serious and result in excessive downtime of the equipment. Thus, a good QC program usually provides cost savings to a X-ray facility that can offset the cost of the program (10,11).

In summary, there are a number of good reasons for the establishment of a QC program for a medical X-ray facility.

First, a number of regulatory agencies mandate the need for a QC program. Second, a QC program is necessary to ensure adequate mechanical, electrical, and radiation safety. The QC program may decrease the legal problems that can be encountered from various malpractice suits. In addition, the QC program should result in a good and consistent quality of the diagnostic information obtained from the radiological procedures. Finally, an economic benefit should be realized in terms of reduced repeat rates and diminished equipment downtime.

Even though the justification for, and the goals of, the QC program can be easily identified, the description of the measurement procedures for the program is difficult for a number of reasons. First, there are often several different ways by which a given measurement can be performed with varying amounts of accuracy. The methodology chosen will often depend on the kind of test equipment available and the training of the QC personnel. A rudimentary QC program that would be restrained to have limited expenditures might utilize very basic tools that could be readily operated by inexperienced QC personnel. This type of program would utilize items, such as pocket ionization chambers, step wedges, digital kVp meters, and pennies for field alignment locators (12). A more sophisticated QC program might utilize items, such as oscilloscopes, digital exposure–exposure rate detectors with waveform outputs, high voltage bleeder units for peak kilovoltage measurement, and special test stands. The level of the QC program is usually determined by the size of the X-ray facility, the complexity of the X-ray equipment, the goals of the QC program, the funding for the QC program, the training of the QC personnel, and the caliber of the support provided by the X-ray service organizations. In the material that follows, the parameters that should be measured and acceptable levels of tolerance will be indicated. Although a method for performing the measurements may be indicated, the reader should be aware that there may be several alternative means for performing the measurements. The author has not attempted to list all QC measurement modalities (13–17).

Furthermore, there is a large diversity of X-ray equipment. Each type of X-ray installation requires certain specialized tests unique to that type of installation. In general, diagnostic radiology equipment can be categorized into the following groups: (1) simple radiographic (rad units), (2) chest radiography units, (3) mammographic units, (4) body section tomographic units, (5) dental units, (6) portable radiographic units, (7) portable C-arm fluoroscopic units, (8) ultrasound units, (9) combined radiographic and fluoroscopic (R & F) units, (10) special procedure rooms (with digital and/or cine), (11) computerized axial tomographic (CT scanners), (12) magnetic resonance imaging (MRI), (13) automated film processors, (14) CR and DR units, (15) video display terminals (monitors or VDTs), (16) picture archiving and communications systems (PACS).

Even though there is a diversity in the types of diagnostic X-ray installations there are some common components of all units; for example, all X-ray installations (regardless of their type) have an X-ray tube, collimator system, control panel, and an X-ray generator. If the QC program were described in terms of the types of X-ray units, the tests for the various components would be

repeated many times. Instead, it is more efficient to describe the QC testing in terms of the components that would be common to many different types of installations (e.g., the X-ray tube). Items that are unique to a given type of installation are listed separately.

Finally, the level of what is deemed as acceptable performance for the equipment can vary. Older equipment cannot be expected to perform as well as newer equipment. Equipment manufactured after August 1974 is required by Federal regulation to meet certain performance levels and have certain special features. This equipment is designated as compliant equipment. Automatic collimation to the cassette size (PBL, positive beam limitation) was previously required; although this requirement is no longer mandatory, many modern X-ray units still employ PBL (18). Equipment manufactured prior to this date is not as strictly regulated. Moreover, equipment that is well maintained and routinely calibrated should be expected to perform better than equipment not receiving this attention. Hence, performance standards must have some leeway for interpretation. The performance standards that will be quoted here are for newer compliant equipment that is properly maintained.

An evaluation of the performance of X-ray equipment should involve an understanding of the accuracy of the measurements, the condition and age of the X-ray equipment, and its usage. Having stated the various caveats and limitations, the material that follows provides a basic QC outline for various components in the X-ray system.

X-RAY TUBE

The X-ray tubes are the source of the X rays produced in the X-ray equipment. Figure 1 shows a sketch of the X-ray tube. Without exposure initiation, there is usually a

standby current passing through the filament. This current is sufficient to keep that filament warm, but it should be less than that necessary for thermionic emission. There are two or more different filaments. The larger filament is used when more thermionic electrons are needed for increased X-ray production. The use of larger filaments, however, results in a larger area on the anode where the electrons collide with the target; this area, as seen from the position of the image receptor, is called the effective focal spot. Large focal spot sizes can result in the degraded image quality due to geometric blur. An important part of a QC program, therefore, is the measurement of effective focal spot size.

Just prior to making an X-ray exposure, the rotor preparation switch is depressed. This switch increases the current through the filament so that sufficient electrons are boiled off in order to yield the correct amount of electron flow (tube current) between the cathode (filament) and the anode (target). Simultaneously, the anode begins to rotate, increasing to its correct speed. The rotation is needed to distribute the heat created by electron bombardment of the target. Upon depression of the exposure switch, the selected high voltage in kilovolts (kVp) is applied between the cathode and the anode. This voltage accelerates the thermionic electrons from the filament and causes them to hit the anode with high speed. The collisions result in X-ray production. The lead lining attenuates X rays in all directions except in the region of the tube port where the lead is missing. Some of the X rays do penetrate through the lead shielding and they are called leakage radiation. This radiation must be measured to determine if it is within legal limits. At the end of the exposure, the high voltage is removed. The filament current then returns to the standby level, and electromagnetic braking is provided to slow the anode to a stop (19–21).

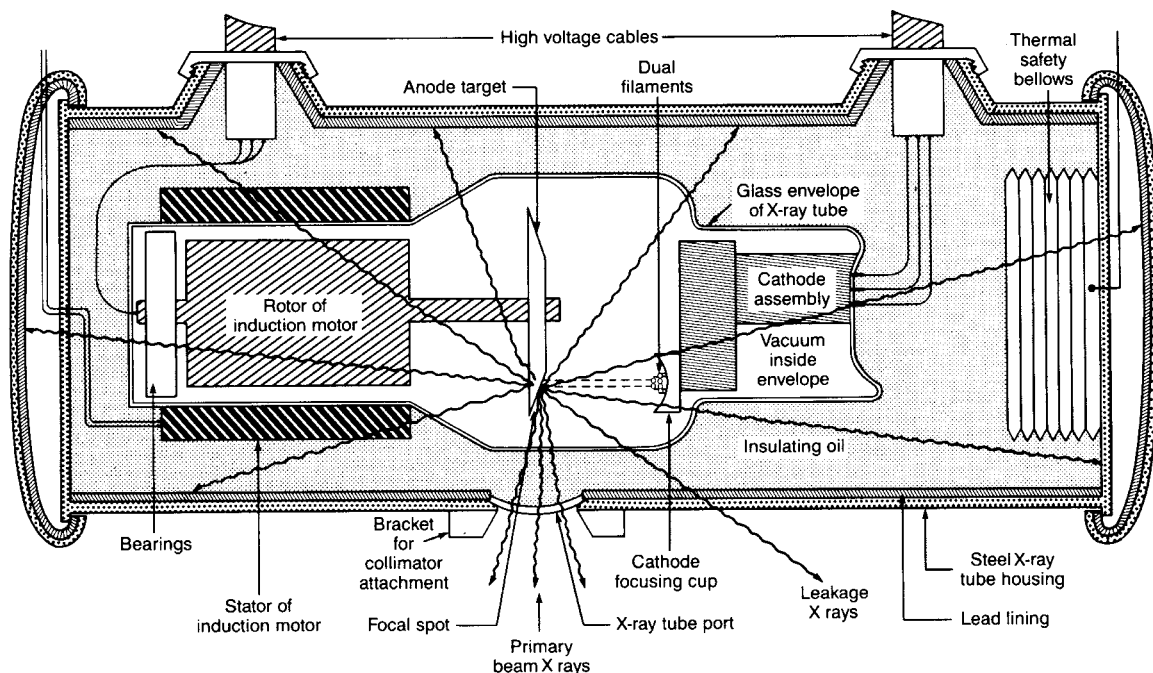


Figure 1. Diagram of an X-ray tube assembly that depicts the major components.

Several types of X-ray tube malfunctions can occur that should be assessed through QC testing. An X-ray tube that has a vacuum leak is known as a "gassy" tube. Gassy tubes have a decreased X-ray output, show spikes in the milliampere and radiation waveforms, and show decreased penetration. A tube that has been overheated repeatedly may cause metal pitting in the anode; the rough anode surface decreases X-ray output (22). The overheating of the anode can also cause metal to evaporate and desposit inside the glass envelop of the X-ray tube; this phenomenon is known as metalization. Metalization can result in electrical arcs which produce spikes in kilovolts, milliamperes, and radiation waveforms. Overheating the filament results in a thinner filament due to evaporation of the metal. This increases the resistance of the filament. Without a tube current feedback circuit, the X-ray production can be decreased. If the filament becomes too thin, the filament will break and X-ray production will cease. Another X-ray tube malfunction is destruction of the bearings upon which the anode rotates due to excessive heat and mechanical wear. As the bearings fail, the anode does not rotate at the proper speed and the heat destroys the target. Bearing problems create excessive noise during anode rotation. Improper warm-up procedures can crack the anode and cause noise during anode rotation. The QC procedures should be able to detect these malfunctions.

MECHANICAL INSPECTION

The X-ray tube housing should be inspected visually for physical damage and/or oil leakage. The alignment of the X-ray tube in the support rings (trunions) should be checked.

ELECTRICAL INSPECTION

All cables into the X-ray tube should be visually inspected for damage to the insulation. The high voltage cables should be checked to make certain that connectors are properly tightened in the X-ray tube housing.

BEARING ROTATION

There are both optical and vibration sensors that can determine the anode speeds. The typical anode speeds are 3600 rpm on low speed and 10,000 rpm on high speed. The best method for checking the bearings is to listen to the sounds made during preexposure boost and slow down. The pitch of the sound should increase for 1–2 s and stabilize as a steady, even hum. During braking, the sound should be rapid 5–10 s deceleration. A metallic clinking sound and other irregular metallic sounds are indicative of bearing failures.

CENTRAL BEAM ALIGNMENT

An X-ray tube can be misaligned in its mountings or the insert can move inside the housing. To check the central beam alignment, a Lucite block with holes drilled in it (or a

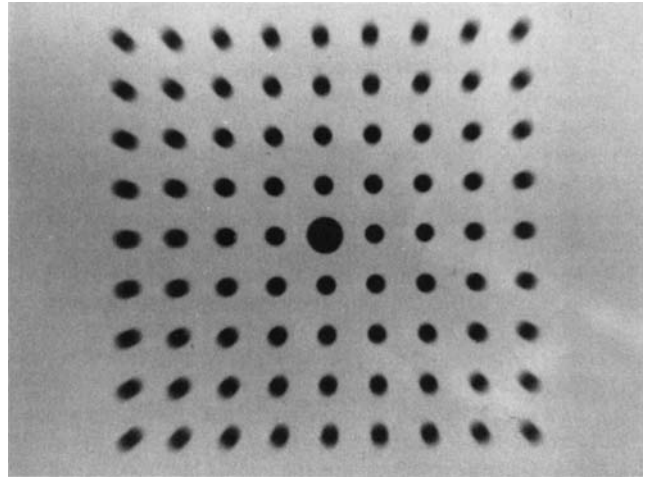


Figure 2. A radiograph of a device utilized to locate the central axis of the X-ray beam.

special hollow plastic cylinder with a central hole on both the top and bottom surface) can be used. The device is placed on a image receptor cassette. The center of the block is positioned at the center of the central beam as shown by the collimator cross-hair. The hole that appears most circular in the film image is at the central axis of the X-ray beam. Off-axis positions have oblong images on the film. This is illustrated in the X-ray image shown in Fig. 2.

X-RAY OUTPUT

The X-ray output could decrease significantly for a number of different reasons. Some sources of a decreased output include excessive X-ray beam filtration, low kilovoltage calibration, low milliampere calibration, damage to the anode from overheating, a pitted anode, or gassy X-ray tube. To test the X-ray output, a radiation detector is placed at 1 m from the X-ray tube focal spot and suspended several centimeters above the table top to decrease backscatter. The output should be measured at 100 kVp with a large-focal-spot milliampere setting and exposure times $>1/10$ s. The measured air radiation dose in milligray (mGy) is divided by the mAs used to make the exposure; the measurement can then be converted to a standardized output by expressing the value as mGy per 100 mAs. For a three-phased, 12 pulse X-ray unit, or a high frequency X-ray generator, the X-ray output should be between 7–10 mGy per 100 mAs at 100 kVp and a distance of 1 m. A single-phase, full-wave rectified unit would measure $\sim 50\%$ less. X-ray outputs that are outside these limits by more than $\pm 15\%$ should be considered to be abnormal unless the X-ray tube has significant filtration (>3 mm Al equiv or >0.1 mm Cu).

RADIATION WAVEFORM

Irregularities in the radiation waveform can be indicative of severe problems with the X-ray tube, cables, and/or X-ray generator. Some sources of irregularities in the radiation waveform include gassy X-ray tubes, pitting of

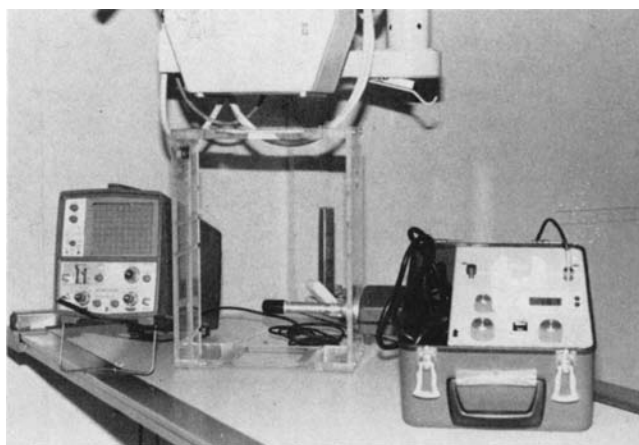


Figure 3. Equipment used to display radiation waveforms. The radiation detector has an output terminal that is connected with a coaxial cable to a storage oscilloscope in order to display the waveform.

the X-ray tube anode, arcing cable connections, malfunctioning milliamperage stabilizer, phase imbalance in the power line, rectifier failure, and other generator problems. Modern radiation detectors usually have a BNC or computer output by which the radiation waveform can be monitored. A coaxial cable can be connected to a storage oscilloscope to record the radiation waveform (see Fig. 3). A three-phase, 12-pulse X-ray tube unit usually exhibits a 5–10% ripple in the wave form; high frequency units have a variable ripple (that depends on kVp and mA used) of 3–20%; a three-phase, 6 pulse unit usually exhibits 20–25% ripple; single-phase units exhibit 100% ripple. The peak values of each of the pulses in the waveform, however, should be relatively constant; the peak values should not change by >7%. The display of spikes or dropouts in the waveform are indicative of problems.

FOCAL SPOT SIZES

Focal spot measurements are usually only performed for new X-ray tubes; focal spot measurements are not routinely performed during QC: unless some spatial resolution problems are being investigated. The size of the X-ray tube focal spots can have a significant influence upon the radiographic image quality. Objectional bur can be introduced by excessively large focal spots. One way to measure focal spots is to use a leaded star pattern that is taped to the center of the collimator face plate. For focal spots >0.6 mm, a 2° star pattern is adequate; <0.6 mm focal spot sizes, a 1° pattern should be utilized. Typically 75–80 kVp should be used along with a milliamperage value that is midrange for the focal spot size being measured. Focal spot sizes decrease as peak kilovoltage increases and milliamperage decreases. Imaging of the star pattern should use plain film in a paper cassette (without intensifying screens that is called direct film exposures). The magnification factor should be small enough that the outermost interference pattern should appear in the film image (see Fig. 4). In general, a magnification factor of ~1.5× should be utilized

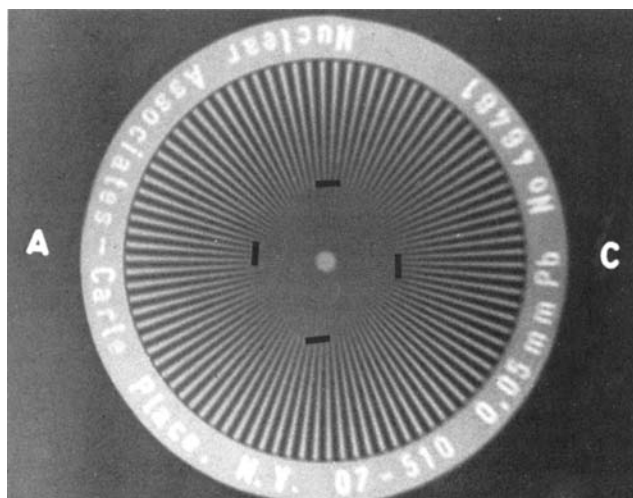


Figure 4. The radiograph of a 2° star pattern used to measure focal spot sizes. The radiograph was taken at 120 kVp and 100 mA with a magnification factor of 2.9. The manufacturer's specified focal spot size was 0.60 mm. The measured focal spot was 0.83 × 0.90 mm.

with a 1.2 mm focal spot, and a magnification factor of ~2.5× should be utilized with a 0.6 mm focal spot. The distance, d , between the outermost interference patterns is measured on the film. The effective focal spot size, α , can be calculated from the following equation (23,24)

$$\alpha = \frac{\pi\theta}{180} \frac{d}{(m-1)}$$

where θ is the number of degrees in the star pattern and m is the magnification factor. The measured focal spot sizes should be within -0 to +50% of the manufacturer's stated specifications in the small dimension (width) and within -0 to +100% in the long (length) directions. Small focal spots (<0.4 mm) often have a -0 to + 50% tolerance in both orthogonal directions.

LEAKAGE RADIATION

Since the X rays produced in the anode of the X-ray tube are nearly isotropic in direction, the lead lining within the X-ray tube housing is intended to reduce leakage radiation to acceptable levels. In order to measure the leakage radiation, the highest peak kilovoltages utilized clinically are selected. To accommodate the relatively slow response time of many exposure rate detectors, a relatively long exposure time is required (2–6 s). Federal regulations specify that the leakage radiation be measured at several locations 1 m from the focal spot and averaged; leakage techniques are specified as the highest peak kilovoltage with the highest continuous tube current. The X-ray tube can only be operated continuously at very low tube currents. These currents are typical of those used during fluoroscopy. Unfortunately, radiographic units usually cannot be operated at low currents in a continuous manner; radiographic units are designed for pulsed operations at high currents with short exposure times. To measure leakage radiation with radiographic units, a high peak

kilovoltage and low milliamperage are selected with a several-second exposure time. The adjustable collimators are closed completely and an additional lead sheet (at least 3 mm thick) is placed to block the primary X-ray beam; in this way, only leakage radiation is measured. Readings are taken with an exposure rate meter at several locations 1 m from the other focal spot and averaged. The average value is scaled down to a value corresponding to a low milliamperage continuous tube current (typically 5–10 mA). Federal regulations required that the average leakage exposure rate measured at a distance 1 m from focal spot at leakage techniques be $<100 \text{ mR}\cdot\text{h}^{-1}$ ($0.87 \text{ mGy}\cdot\text{h}^{-1}$). Higher readings are indicative of problems with the lead lining inside the X-ray tube housing. Excessive leakage radiation levels are rare occurrences. Moreover, these measurements need only be performed after X-ray tube replacement; they should not be a part of the routine QC program.

COLLIMATOR ASSEMBLY

The collimator assembly has variable lead shutters that define the size of the X-ray field. The assembly also contains a light bulb and mirror that are used to project a light field identical in size and location to the X-ray field. Collimator assemblies usually contain metallic filters (usually composed of aluminum or maybe also a thin layer of copper) that remove low energy X rays that are not very penetrating. These low energy X rays do not have sufficient energy to penetrate through the patient anatomy that is being

radiographed in sufficient quantity to influence the imager quality; these low energy X rays, however, can contribute significantly to the patient radiation dose. Therefore, the X-ray tube plus collimator assembly should contain sufficient metallic filtration to remove the very low energy X rays and harden the X-ray spectra. Figure 5 shows a typical collimator assembly. The QC procedures should evaluate the X-ray field–light field congruency and the penetration of the X-ray beam (19–21).

X-RAY FIELD–LIGHT FIELD ALIGNMENT

If the localizer light is misaligned with the X-ray field, the radiographic image can be too small, too large, or miscentered on the X-ray field. The alignment can be checked by placing a image receptor cassette at a given distance from the focal spot. The typical SID utilized in diagnostic radiology is 1 m. With the light localizer turned on and pointed at the cassette, metal markers are used to delineate the periphery of the light field; pennies are often used for the markers. A radiographic exposure of the film is made. The misalignment of the metal markers with the edge of the X-ray field is determined from the developed image. Figure 6 is an example of a misaligned X-ray–light field. Federal regulations specify that the misalignment in any one orthogonal direction should not exceed 2% of the SID. For automatic collimation (PBL), the error in any one direction must be $<3\%$ and the sum of the errors in two orthogonal directions must be $<4\%$.

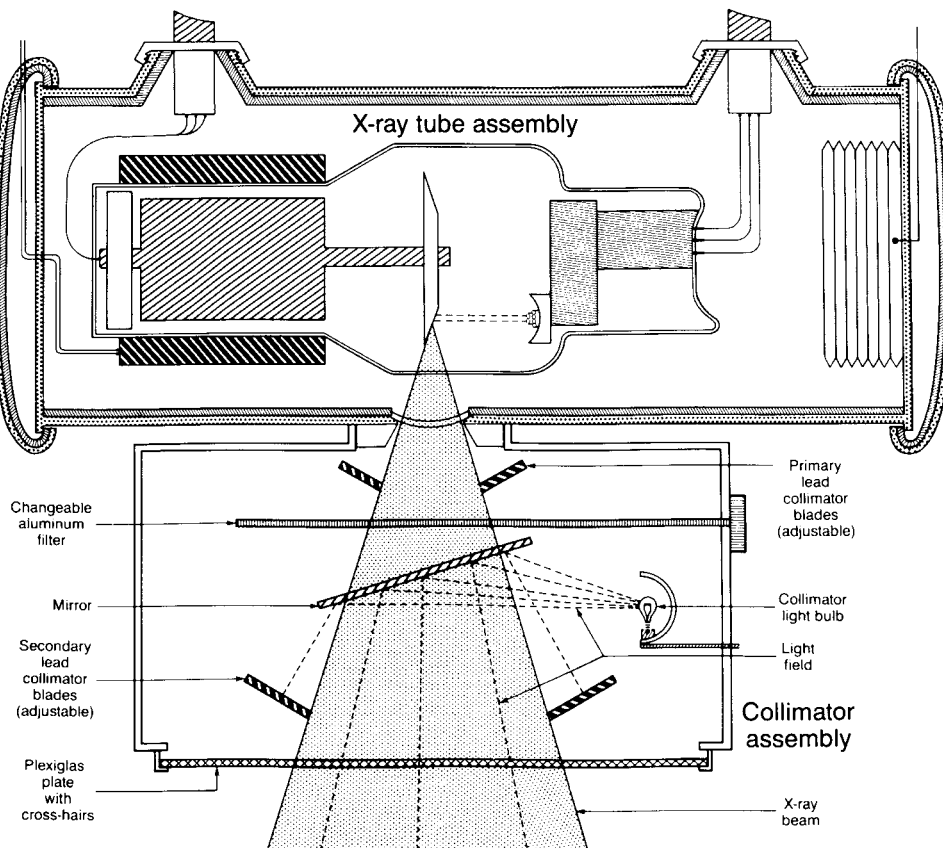


Figure 5. Diagram of an X-ray collimator assembly that depicts major components.

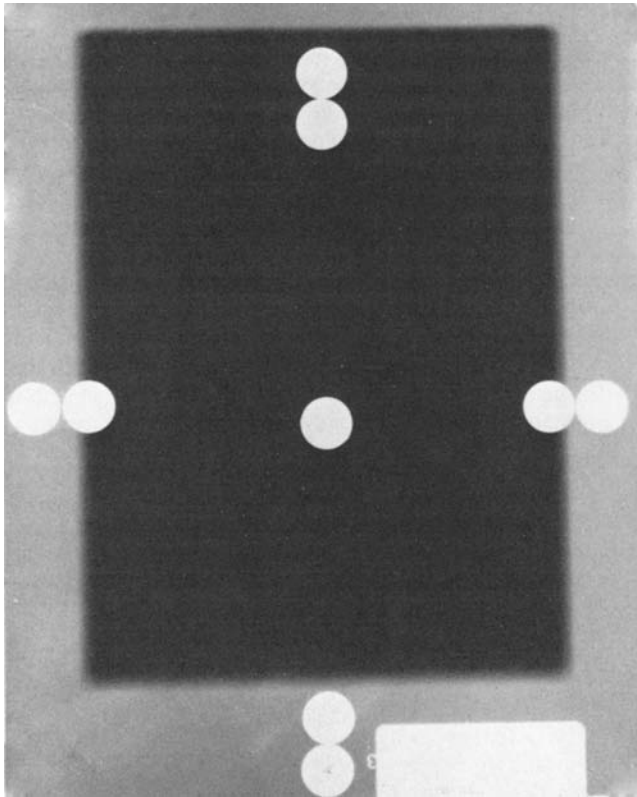


Figure 6. Radiograph of the markers used to measure X-ray–light field congruence. The fields were severely misaligned in this image.

HALF-VALUE LAYER DETERMINATION

The half-value layer (HVL) for each X-ray unit is a measurement of the penetration of the X-ray beam. The HVL is the thickness of aluminum necessary to reduce the X-ray beam intensity by 50%. The HVL is measured in order to determine whether there is sufficient filtration in the X-ray beam. Insufficient filtration will result in increased patient dose. Excessive filtration results in extra stress on the X-ray tube, a loss in radiographic contrast, and increased exposure time. Increased exposure times can result in blur due to patient motion.

The HVL is measured by placing a radiation detector in the X-ray field. Although any selection of peak kilovoltage and mAs can be chosen, usually 80 kVp at 200–400 mA and 0.10 s is selected on the X-ray control. Two exposures are taken with no added filtration in front of the detector; the average value of these two readings represents 100% transmission through the X-ray tube and collimator assembly. Sheets of 1100 aluminum (mammography uses extra high purity aluminum filters) are then placed (one at a time) between the X-ray tube and the detector. The aluminum should be placed closer to the X-ray tube, allowing at least a 30 cm air gap between the aluminum and the detector to minimize scatter effects.

After each new piece of aluminum is added, an exposure is taken and the radiation measured is recorded. When the exposure measured is <50% of the baseline value, the data are plotted on semilog paper and the HVL determined. For

80 kVp, the Federal regulations require the HVL to be >2.3 mm of aluminum equivalent (3). The following table provides the minimum acceptable HVL values for different modalities as a function of the kVp used:

Modality	Equation for Minimum HVL
Mammography	$>(\text{kVp}/100) + 0.03 \text{ mm Al}$ $<(\text{kVp}/100) + C \text{ mm Al}$ Where $C = 0.12 \text{ Mo/Mo}$, 0.19 Mo/Rh , and 0.22 Rh/Rh
Radiograph and routine fluoro	$>[2.26 (\text{kVp}/100)] + 0.48 \text{ mm Al}$
Angiography and cardiac fluoro	$>[3.5 (\text{kVp}/100)] + 0.08 \text{ mm Al}$

The maximum HVL should not be >6 mm Al equiv at 80 kVp for routine fluoroscopic, and radiographic units nor >9 mm Al equivalent at 80 kVp for angiography and cardiac fluoroscopy. For CT scanners at 120 kVp, typical values of the HVL are between 6 and 10 mm Al equiv.

ILLUMINATION OF THE LIGHT LOCALIZER

It is important that the collimator light be bright enough to visualize the borders of the light field during normal light conditions in an X-ray room. Otherwise, the X-ray field would not be properly determined and collimated. A visual inspection of the collimator localizing light can be conducted under normal room lighting. If a light meter is available, the intensity of the collimator light should be measured with the room lighting turned off. Federal regulations require that the light intensity of the collimator localizer >160 lux (15 ft·C) at a distance of 100 cm.

POSITIVE BEAM LIMITATION

Some X-ray equipment has an automatic collimation system for the X-ray beam. This automatic collimation system is designated as PBL. The PBL restricts the X-ray field size to the size of the cassette placed into the bucky tray. If this system malfunctions, the X-ray field size could be larger than the cassette and result in unnecessary X-ray dose to the patient. A smaller field size due to malfunctions could result in a loss of diagnostic information toward the periphery of the film. If the PBL system is present, some state regulatory agencies have performance criteria for the collimation accuracy.

To test the PBL system, two identical cassettes can be used. The collimator field size is opened to its maximum extent, and one cassette is placed into the bucky tray. The localizer light is turned on in the collimator, and the tray is inserted into position. One should visually observe the light field decrease to the approximate size of the cassette. The second cassette is placed upon the table and centered in the light field. The light field should be slightly inside all borders of the cassette on the tabletop. In fact, for a 1 m SID and an 8 cm distance between the cassette in the bucky tray, and the tabletop, the light field should be 7.5% smaller

than the cassette in each of the two orthogonal directions. A typical criteria for the PBL collimation is an accuracy of $\pm 3\%$ of the SID error in either orthogonal direction and $\pm 4\%$ of the SID sum error without regard to sign for both directions.

MINIMUM SOURCE-TO-SKIN DISTANCE

In order to prevent placement of the X-ray tube too close to the patient, the federal regulations specify the minimum distance that the bottom of the collimator must be from the focal spot. This regulation usually does not apply to overhead X-ray tubes where the patient would never be placed next to the collimator under routine clinical conditions. This distance can be determined from geometric magnification effects. If an object of known size (OS) is placed at the bottom of the collimator and radiographed, the image size (IS) will be magnified. The distance from the object to the film should be measured (OFD). The source-to-skin distance (SSD) can be determined from the measurements.

$$SSD = [OFD(OS/IS)]/[1 - (OS/IS)]$$

For radiographic units, the federal regulations specify that the SSD be >30 cm (12 in.) for mobile systems. For fluoroscopic units, the Federal regulations specify that the minimum SSD should be >38 cm (15 in.) for stationary certified fluoroscopic units, 30 cm (12 in.) for mobile fluoroscopic units, 19 cm (7.5 in.) for mobile fluoroscopy c-arm with an SID = 45 cm that are used for extremity imaging and 20 cm (8 in.) for image intensifier fluoroscopic units used for specific surgical applications. In general practice, the minimum SSD for stationary certified fluoroscopic units is usually 45 cm (17.7 in.).

X-RAY GENERATOR

The X-ray generator controls X-ray production. The timer circuit determines the duration of the X-ray exposure. The milliamperage selector determines the number of electrons per second that bombard the anode. Hence, the product of milliamperage and time determines the number of X rays produced; this product is known as mAs. kVp selector provides the high voltage between the cathode and anode that accelerates the electrons in an X-ray tube. The peak kilovoltage determines both the energy spectrum and the number of X-ray produced. As peak kilovoltage increases, the penetration of the X-ray beam increases (19–21).

In order to be able to consistently take acceptable radiographs, it is important the X-ray generators be properly calibrated. Moreover, the exposure should be reproducible and the various X-ray equipment should be calibrated to the same standard. In order to meet these criteria, the QC procedures should carefully evaluate the generator performance. Most extremity imaging and some other projections are performed using manual X-ray generator settings.

X-RAY REPRODUCIBILITY

If an X-ray generator is operating properly, the amount of radiation produced for a given selection of technique

factors should be constant. If there is a problem with milliamperage stabilization, peak kilovoltage waveform, arcing in the X-ray tube, and/or timer viability, the measured X-ray output at any fixed location will vary. A radiation detector should be placed in the central ray ~ 30 – 70 cm from the bottom of the collimator. Five-to-ten separate X-ray exposures are recorded for some fixed selection of technique factors (e.g., 80 kVp, 300 mA, and 0.10 s). The average value and the standard deviation for these exposures is calculated. The ratio of one standard deviation divided by the average exposure is also computed. The ratio is called the coefficient of variation (CV). Federal regulations require that the CV be <0.050 for compliant X-ray units and it is desirable that this value really be <0.02 for properly functioning X-ray generators.

TIME ACCURACY

Since the exposure duration timer circuitry partly determines the mAs used, inaccuracy in the timer could result in variations in the radiographic density for film images and mottle for digital images. The duration of X-ray exposures can be easily measured by either special electronic X-ray timers or some digital exposure detectors. One merely places these devices in the X-ray beam and reads the measured exposure time. Other methods utilize the display of a radiation waveform on an oscilloscope and determining the exposure time from the width of the displayed waveform full width at half-maximum (fwhm). It is recommended that the timers be accurate to within $\pm 5\%$ or ± 1 ms, whichever is larger at all timer settings on the control panel; Single-phase equipment should only be accurate to 1 pulse or ± 8.3 ms. Instead of testing all the timer settings, however, 5–10 timer settings from the shortest to the longest should be checked. The X-ray tube chart should be consulted to avoid exposures that would overheat the tube.

TIMER REPRODUCIBILITY

One reason for the lack of exposure reproducibility is an inconsistently functioning X-ray timer circuit. The time for five identical exposures should be measured. Federal regulations require that average exposure time (T) for any fixed selection of technique factors be greater than five times the worse excursion in the exposure duration

$$[5(T_{\max} - T_{\min})]$$

MILLIAMPERAGE ACCURACY

Another factor that influences the selected mAs is the milliamperage calibration. Again, inaccuracies among the milliamperage stations can result in improper radiographic densities on the film. Moreover, it is desirable that various rooms in a large X-ray department have a similar calibration. Therefore, technique charts for procedures throughout the department can be standardized.

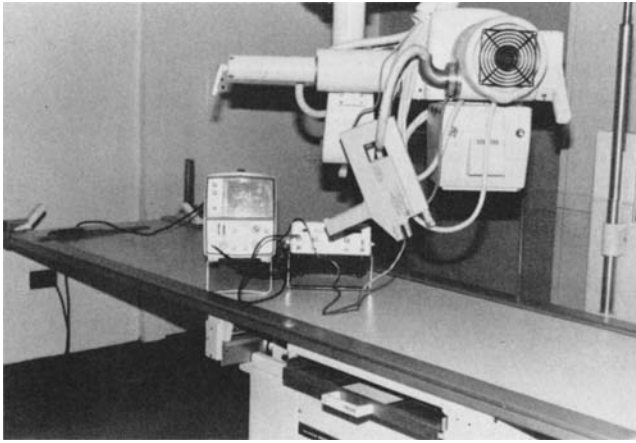


Figure 7. Equipment used to measure and display milliamperage of the generator. The noninvasive probe shown on the X-ray cable measures the milliamperage indirectly by sensing the induced magnetic field around the cable. The storage oscilloscope is used to display the milliamperage waveform.

Several different ways to measure the milliamperage accuracy are available. A simple method is the use of noninvasive current probes (25). These probes clamp around the high voltage cables to the X-ray tube and measure milliamperes indirectly by sensing the induced magnetic field around the cable. Figure 7 shows this equipment. This method is usually better for measurement of the larger milliamperage settings. It is recommended that all milliamperage stations be calibrated to within $\pm 5\%$ or ± 1 mA (whichever is larger) of the indicated value on the control panel. Furthermore, the shape of the milliamperage waveform should be observed on an oscilloscope to detect problems, such as arcing, loading, lack of stabilization, rectification problems, and excessive ripple. The waveform should be regular and the peaks in the milliamperage waveform should not vary by $>7\%$.

MILLIAMPERE LINEARITY

The assessment of the relationship between the various milliamperage settings and their corresponding measured X-ray output is called milliamperage linearity. If the milliamperage is increased by a factor of 2 with all other settings unchanged, the X-ray output should also increase by a factor of 2. If the milliamperage settings on the control are miscalibrated, the actual milliamperage will not increase linearly with the selected values.

In order to make milliamperage linearity measurements, one merely places a radiation detector in the X-ray field along the central ray. The detector is kept at a fixed location. One peak kilovoltage and time setting is selected (e.g., 80 kVp and 0.1 s). A series of fixed exposures are measured for each of two (or more) different milliamperage settings (e.g., 200 and 400 mA). The average exposure value for each of the 2 mA stations is calculated (E_1 and E_2). The average exposures are divided by the mAs values used to produce the exposures mGy per 100 mAs (or mR mAs⁻¹). The term ($X_1 = E_1 \cdot \text{mAs}^{-1}$) should be the same for

all selections of milliamperage employed. A linearity factor (LF) is then calculated as follows:

$$LF = |X_1 - X_2| / |X_1 + X_2|$$

Federal regulations require that the LF to be <0.10 . Actually, most X-ray generators are capable of LF values <0.05 for the 2 mA settings with the largest discrepancies (if properly calibrated and functioning correctly).

PEAK KILOVOLTAGE CALIBRATION

Improper peak kilovoltage calibration of the X-ray generator has major influences upon image quality and patient dose. High peak kilovoltages reduce patient contrast, increase scatter, and increase subject latitude; they also result in reduction of patient dose. Low peak kilovoltages have the opposite effect. Moreover, miscalibration results in non-uniform X-ray quality throughout a large X-ray department. Thus, a standard for the determination of peak kilovoltage accuracy is important.

There are a number of methods by which peak kilovoltage calibration can be measured. These methods include high voltage bleeder resistor networks and noninvasive kVp meters (26–29). A simple method often utilized is noninvasive kVp meters, such as one shown in Fig. 8. These meters are placed in the X-ray beam and display a digital number for the measured peak kilovoltage. The meters contain several solid-state radiation detectors, each with different amounts of copper filtration. By measuring the amount of penetration of X-ray beam through these filters the peak kilovoltage can be computed. These noninvasive peak kilovoltage meters are usually accurate to within ± 2 kVp. Even though more accurate peak kilovoltage instrumentation exists, the relative ease with which is noninvasive peak kilovoltage measurements can be made is their principal advantage. It is recommended that peak kilovoltage calibration be accurate to within $\pm 5\%$ or ± 2 kVp, whichever is larger at all milliamperage

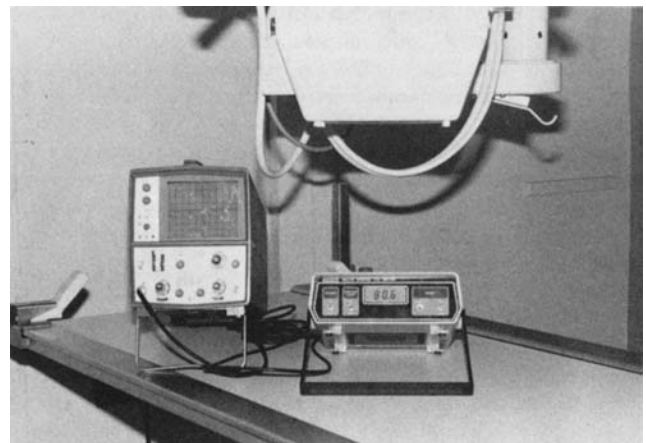


Figure 8. Equipment used to measure and display to peak kilovoltage of the generator. The device shown below the X-ray tube is a noninvasive peak kilovoltage meter. The storage oscilloscope is used to display the peak kilovoltage waveform.

stations for a range of peak kilovoltage settings. For radiographic units, the checks are usually made at 60, 90, and 120 kVp, using a variety of milliamperage settings. The shape of the kVp waveform can also be observed by a BNC output connection from the noninvasive kVp meter to an oscilloscope. The peak kilovoltage waveform is monitored to detect problems with arcing, excessive or irregular ripple, stabilization problems, rectification failures, phase imbalances, and noise. The peaks in the kVp waveform should not vary by $>7\%$.

OPERATOR'S CONTROL PANEL

The control panel contains the selection switches that determine the technique factors for the X-ray exposure. The QC program should determine that selection switches function properly and that regulatory requirements are met.

FUNCTION CHECKS

The various selection switches should be tried; it needs to be determined that the switches actually function. Furthermore, all lights, meters, and digital displays should be checked to make certain they are operational.

EXPOSURE CONTROL

Recommendations suggest that the exposure control should be secured at least 40 in. (1 m) from the edge of the secondary radiation protection barrier. It should not be possible to carry the exposure switch into the room and/or make an exposure while inside the X-ray room.

REGULATORY CHECKS

The operator's control panel is required to have certain features. The QC procedures should check for these regulatory items. There must be a warning label with a caution about the hazards of X-ray irradiation. The technique factors must be indicated on the control panel prior to an exposure. During an exposure, there should be a visible beam-on indicator and an audible indicator of exposure termination. Following the exposure, there should be an automatic resetting of the exposure termination timer. For fluoroscopic units, there should be a 5 min timer; after 5 min of fluoroscopy, there must be an audible indicator at the end of each 5 min of fluoroscopy. The 5 min timer alarm must continue to sound until the timer is reset (18). However, the reset of the alarm must not affect the cumulative timer that accumulates the total fluoroscopy time during each study.

TABLE AND TUBE STAND

Foremost, the mechanical functions of the table and tube stand motion should be checked. All typical positions utilized clinically should be tried. The accuracy of distance

indicator for the SID should be determined; a tape measure can be used to check the SID indicator. The mechanical safety of the table and tube stand should be visually inspected. Anticollison devices should be checked.

AUTOMATIC EXPOSURE TERMINATION

Automatic Exposure Control (AEC) or phototiming utilizes a radiation detector to measure the amount of X-ray radiation incident upon the film-screen or CR cassette. When the detector measures sufficient radiation to produce adequate density on the image or a level of mottle on digital image receptors that is deemed appropriate, the detector sends an electronic signal to the X-ray generator, which terminates the exposure. In this manner, with properly functioning phototimers, radiographs should have satisfactory density (film) or minimally acceptable noise (CR/DR) to prevent degradation of image quality. In order to check the AEC system, the phototimer tracking should be measured.

PHOTOTIMER TRACKING

An accurate AEC circuit can result in a decrease in the repeat rate and a consistent image quality for a variety of patient sizes. However, phototimer circuits can be susceptible to variations with changes in the peak kilovoltage, non-linear responses of film-screen systems, a limited range of patient sizes, or a lack of reproducibility. The best method to check the phototimer is to use acrylic plastic sheets to stimulate the patient and to take radiographs of the plastic. The acrylic sheets are placed in front of the phototimer to simulate various patient sizes. For various peak kilovoltage settings or changes in the plastic thickness, the density of the exposed and developed film (as determined with a densitometer) or Exposure Index Number for CR or DR cassettes should remain constant. If the phototimer circuitry is performing properly, it is recommended that the measured film density be within ± 0.15 du at one standard deviation (or $\pm 10\%$ for digital system index number) of the average value for all variations in peak kilovoltage or simulated patient thickness. Furthermore, for conditions of fixed peak kilovoltage and thickness, reproducibility of the measured film density should be within ± 0.05 du (one standard deviation) or $\pm 5\%$ for the digital system index number.

BUCKY TRAY

The tray into which the film-screen or CR cassette is inserted is called the bucky tray. Above the tray, there is usually a moving grid that is used to remove some of the scattered X rays. It is important that the grid moves properly during the exposure and that it does not contain any defects. A common problem is that the wrong grid is inserted into the table. The detente mechanism that positions the X-ray tube over the cassette may also not be properly centered. The QC tests should attempt to investigate for possible problems.

VISUAL INSPECTION

If possible, the grid should be removed from the bucky assembly on an annual basis. It should be visually inspected for damage and for spills of contrast material. The focal distance and grid ratio marked on the grid should be checked to determine if they are appropriate for the intended radiographic procedures. It should also be determined that the correct side of the grid faces the X-ray tube; inverted grids result in cutoff of image density toward the edges.

GRID UNIFORMITY

In order to ascertain that there are no defects in the grid and that there are no irregular cutoffs, a cassette should be placed in the bucky tray and a rapid exposure made to result in a film density of ~ 1.20 du (or appropriate digital image index number). Exposure times should be < 20 ms in order to stop grid motion. The developed film should have uniform density throughout and the grid lines should be visible. Irregularities in the grid lines or nonuniform density may be indicative of a grid problem.

BUCKY MOTION

Sometimes, the device that causes the grid to oscillate during the exposure malfunctions or is not properly synchronized with the X-ray exposure. To test for bucky motion, a cassette is placed in the bucky tray and an exposure is taken with a longer exposure time selected. Exposure times should be > 100 ms. The developed image should have a uniform density and grid lines should not be visible. The bucky motion should blur all the grid lines on the image. If the grid lines still appear, there is a problem with the bucky motion.

CENTERS ALIGNMENT

The miscentering of the X-ray field and the center of the image receptor cassette could result in the loss of diagnostic information. For this test, the detente and collimator light systems are used to position the X-ray tube assembly over the film–screen cassette that has been placed in the bucky tray. The X-ray field size must be smaller than the cassette. An X-ray exposure of the cassette is made and the centers of both the X-ray field and the image are determined. The misalignment of the center of the cassette with the center of the X-ray field must be $< 2\%$ of the distance between the X-ray tube focal spot and the cassette (SID).

FLUOROSCOPIC X-RAY EQUIPMENT

Some of the components in fluoroscopic X-ray equipment are similar to those in radiographic equipment. The same X-ray generators are often used to operate several X-ray tubes in a single radiology room; often a fluoroscopic and radiographic X-ray tube are connected to the same X-ray generator. However, in addition to the standard radio-

graphic components, fluoroscopic X-ray equipment usually includes additional components, such as image intensifier, cassette spot film devices, television systems, 100 mm (or digital) fluoroscopic spot film cameras, and a tilting table. The fluoroscopic system is designed to operate as a real-time dynamic imager. The X-ray tube operates with low milliamperage values for continuous fluoroscopy so that X-ray production can continue for minutes without overheating the X-ray tube. Pulsed fluoroscopy utilizes large milliamperage values (> 100 mA), but the duty cycle (actual X-ray production) is $\sim 5\text{--}30\%$. The image intensifier stops the X rays transmitted through the patient and converts them to a light image. The light from the image intensifier can then be recorded by the television system, cassette spot film, or the spot film camera. The QC tests should also include those components that are specific to fluoroscopic systems. The QC procedures for the X-ray generator and tube have been described previously (30).

TRACKING TEST

The image-intensifier assembly (IIA) on fluoroscopic X-ray systems can be positioned at various heights above the table. It is important that the collimator on the X-ray tube adjusts to track the image receptor properly, otherwise, the X-ray field size could possibly exceed the image receptor size. This function is observed by placing an object in the fluoro field and moving the IIA up and down. The collimator should adjust automatically. Federal regulations require compliant fluoro systems to have tracking with SID adjustment.

INTERLOCK TEST

When the IIA is placed in the park position, fluoroscopic X-ray production should be inhibited. If fluoroscopic X-ray production were not inhibited, the patient could receive a radiation dose without any image being recorded (i.e., unnecessary dose). The test is performed by placing the IIA in the park position and then placing a radiation detector on the table surface. The fluoroscopic “on” switch is depressed and the detector indicates whether X rays were produced. IIA interlocks are required for compliant units.

PRIMARY BARRIER TRANSMISSION

The IIA acts as a primary radiation barrier. Any X ray impinging upon it should be drastically attenuated. Transmission of the X rays should be minimal. The test is performed by placing a 1.5 in. (3.8 cm) aluminum penetrometer on the table surface in the primary X-ray beam and depressing the fluoro switch. The entrance radiation exposure rate into the penetrometer is measured as well as the exposure rate of the radiation transmitted through the IIA. Federal regulations require that the exposure rate transmitted through the IIA should not exceed $2 \text{ mR}\cdot\text{h}^{-1}$ at 10 cm above the IIA surface for each roentgen per minute of entrance exposure.

MAXIMUM ENTRANCE EXPOSURE RATE

In order to limit the patient dose, the federal government has established regulations to limit the maximum fluoroscopic entrance exposure rate. The measurement can be performed with no backscatter material according to regulatory specifications; however, more realistic conditions can be simulated by placing a 1.5 in. (3.8 cm) aluminum penetrometer on the table with a radiation detector next to the penetrometer on the side toward the X-ray tube. Once the detector is centered in the X-ray beam, a lead sheet (3 mm thick) is placed on the penetrometer side toward the IIA. Both the Automatic Brightness Control (ABC) and the manual adjustments are used serially to obtain the highest X-ray radiation levels. The radiation detector is used to measure the exposure rate. Federal regulations require that the entrance rate at the point where the center of the useful beam enters the patient shall not exceed $87 \text{ mGy}\cdot\text{min}^{-1}$ in air ($10 \text{ R}\cdot\text{min}^{-1}$) in the ABC mode, except during the recording of the fluoroscopic image (cassette or spot exposures) or when provided with an optional high level control. When provided with the optional high level control, the equipment shall not be operable at any combination of tube potential and current that will result in any exposure rate in excess of $87 \text{ mGy}\cdot\text{min}^{-1}$ in air ($10 \text{ R}\cdot\text{min}^{-1}$) at the point where the center of the useful beam enters the patient unless the high level control is activated. In the higher level fluoroscopy mode, a special buzzer or chime must sound and the maximum patient entrance exposure rate must be $<174 \text{ mGy}\cdot\text{min}^{-1}$ in air ($20 \text{ R}\cdot\text{min}^{-1}$). Furthermore, equipment that does not employ ABC shall not be operable at any combination of peak kilovoltage or milliamperage that will result in an exposure rate in excess of $43 \text{ mGy}\cdot\text{min}^{-1}$ in air ($5 \text{ R}\cdot\text{min}^{-1}$) at the point where the center of the useful beam enters the patient, except during recording of the fluoroscopic images or when provided with an optional high level control.

TYPICAL PATIENT ENTRANCE EXPOSURE RATES

In order to assess the normal operation of fluoroscopic or cine equipment, typical patient entrance exposure rates should be measured. These rates can be influenced by the improper functioning of the equipment and by adjustments made by the service personnel. A 3.8 cm (1.5 in.) aluminum penetrometer to which 0.5 mm copper has been added can be utilized to simulate an average male patient of $\sim 75 \text{ kg}$ weight; the measurement procedure is the same as that for maximum dose rate, except that the lead is not placed in the X-ray beam. It is recommended that the simulated patient entrance exposure at the point the X-ray beam enters the phantom should not exceed $8.7\text{--}26.1 \text{ mGy}\cdot\text{min}^{-1}$ in air ($1\text{--}3 \text{ R}\cdot\text{min}^{-1}$) in the 23 cm (9 in.) Field-of-View (FoV) with ABC. During these measurements, the indicated fluoroscopic peak kilovoltage should $>55 \text{ kVp}$ and $<90 \text{ kVp}$. During digital cine recording in the 23 cm (9 in.) mode, the measured entrance exposure should be less than $200 \mu\text{Gy}$ per frame (23 mR per frame) with the same phantom of 3.8 cm (1.5 in.) of aluminum plus 0.5 mm of copper. Some local regulatory agencies require that

typical patient radiation levels be measured annually for all X-ray equipment.

LIMITATION OF THE X-RAY FIELD TO THE IMAGE RECEPTOR SURFACE

The intent of this test is to limit the X-ray field size to the IIA surface such that the patient is only irradiated over a surface for which a TV image can be seen. If the X-ray field size were larger, the patient would be unnecessarily receiving a radiation dose over an area that would not be imaged. In order to perform the measurement of X-ray field size, a leaded ruler is placed in two orthogonal directions upon a film in a cardboard cassette. The film with the ruler can be placed in the X-ray beam anywhere between the tabletop and the IIA surface. The film and ruler are then fluoroscoped. The TV image is viewed to determine the amount of the scale that is visible on the TV monitor. The exposed film is then developed and the size of the fluoro X-ray field is indicated by the darkened portion on the film. Federal regulations specify that the misalignment between the X-ray field size and the image receptor in the fluoroscopic mode must be $<3\%$ of the SID in either orthogonal direction and less than a sum of 4% of the SID for the two orthogonal directions combined.

FLUOROSCOPIC X-RAY OUTPUT

The fluoroscopic X-ray output is valuable information utilized in several ways. These data can then be used to yield patient dose estimates. Furthermore, miscalibrations in peak kilovoltage and/or milliamperage and filtration problems would become apparent in the fluoroscopic X-ray output measurements. Additionally, malfunctions in the X-ray tube, such as anode deterioration, should cause drastic changes in the fluoroscopic output. The measurement is performed by placing an X-ray detector on the table surface (without backscatter) in the X-ray beam to measure the air dose rate in $\text{mGy}\cdot(\text{mA}\cdot\text{min})^{-1}$ [or exposure rate in $\text{R}\cdot(\text{mA}\cdot\text{min})^{-1}$] during fluoroscopy for various peak kilovoltage settings. It is recommended that the fluoroscopic X-ray output measured at the table surface ($\sim 46 \text{ cm}$ or 18 in.) from the focal spot should be between 13 and $22 \text{ mGy}(\text{mA}\cdot\text{min})^{-1}$ [or 1.5 and $2.5 \text{ R}\cdot(\text{mA}\cdot\text{min})^{-1}$] at 100 kVp , except for heavily filtered X-ray beams.

RADIATION INTO INPUT RECEPTOR(S)

The operation of fluoroscopic X-ray equipment is closely aligned to levels determined by feedback control loops that measure parameters associated with the input radiation exposure levels. Thus, proper equipment performance necessitates that the levels be adjusted appropriately. Within the guidelines listed below, the levels should be high enough to limit fluoroscopy image noise to tolerable levels, but low enough to limit patient dose and contrast losses associated with the ABC driving to high peak kilovoltage settings. These tests are performed with a 3.8 cm (1.5 in.) aluminum plus 0.5 mm copper in the

X-ray beam and the brightness being controlled by the ABC mode. The IIA should be placed high above the aluminum and copper attenuator in order to limit scattered X-rays from being measured. The radiation detector should be placed at the IIA surface with the grid removed (on some units the grid cannot be easily removed). It is recommended that the following radiation levels for input exposure rates the image receptor be established:

1. Fluoroscopy for 23 cm (9 in.) mode (no grid) $\leq 6 \text{ mR}\cdot\text{min}^{-1}$ ($100 \text{ R}\cdot\text{s}^{-1}$ or $0.9 \text{ Gy}\cdot\text{s}^{-1}$)
2. Fluoroscopy for 15 cm (6 in.) mode (no grid) $\leq 12 \text{ mR}\cdot\text{min}^{-1}$ ($200 \text{ R}\cdot\text{s}^{-1}$ or $1.8 \text{ Gy}\cdot\text{s}^{-1}$)
3. Digital cine recording for 23 cm (9 in.) mode (no grid) $\leq 20 \text{ R/frame}$ or $0.17 \text{ Gy}\cdot\text{f}^{-1}$
4. Digital subtraction angiography radiographic exposures (no grid) $\leq 1.0 \text{ mR}$ per image or 8.7 Gy per image for 23 cm (9 in.) mode

Excessive input radiation levels are indicative of problems, such as the selection of too small an aperture (a diaphragm placed behind the objective lens of the IIA) or deterioration of the image intensifier.

SPATIAL RESOLUTION MEASUREMENTS

The image quality analysis of fluoroscopic X-ray systems requires the measurement of the spatial resolution capabilities of the various imaging modalities. Unfortunately, the TV chain usually degrades the spatial resolution available from the output phosphor of the image intensifier (II). Hence, measurements of II spatial resolution require the removal of the TV camera; direct observation of the II output phosphor with a special telescope is necessary. This measurement is rarely done. Furthermore, the spatial resolution is dependent upon: the magnification mode for the II [6 or 9 in. (15 or 23 cm) field size], the use of a grid, amount of scatter, the type of test pattern, the location of the test pattern on the II surface, the peak kilovoltage employed, and the focal spot size. Focal spot influences are minimized by the placement of the bar test pattern as close as possible to the II input surface; however, clinical usage has the object of interest displaced away from the II input surface. Therefore, a second test should be performed with the test pattern located at a position corresponding to the patient's location. In this case, the combined effect of focal spot and II distortion is included in the spatial resolution measurement.

A variety of measurement procedures for assessing the spatial resolution of the II are utilized; the following methodology, however, is suggested. A 3.8 cm (1.5 in.) aluminum penetrometer should be placed on the tabletop at the center of the X-ray beam. A 0.10 mm lead bar pattern (0.5–5.0 line pairs per millimeter) should be taped at the II surface with the grid removed. The equipment should be operated in the ABC mode with 60–80 kVp. The output phosphor of the II should be observed through a special telescope. An alternative procedure would be to record the image of the bar pattern with the fluoroscopic spot film camera. The mea-

sured spatial resolution for the II should be > 3.8 – 4.0 line pairs per millimeter in the 15 cm (6 in.) mode and 2.5–2.7 line pairs per millimeter in the 23 cm (9 in.) mode. The TV camera should then be reattached and the spatial resolution on the TV monitor should be examined. The lead bars should be oriented at 45 degree to the raster lines on the TV system. For a 525 line TV system, the spatial resolution should be at least 1.8–2.0 line pairs per millimeter in the 15 cm (6 in.) mode and at least 1.2 line pairs per millimeter in the 23 cm (9 in.) mode. For 1023 line TV units, the measured spatial resolution is typically 50–80% better. Flat panel image receptors typically have resolutions of 2.2–2.8 line pair millimeters in all FoV.

CONTRAST RATIO

Due to light scattering from adjacent areas (veiling glare) in the fluoroscopy image, contrast can be degraded. It is therefore important to measure the maximum contrast achievable. To perform this measurement, a lead disk [at least 0.32 cm (1/8 in.) thick] is placed at the center of the input surface. The diameter of the disk should be 10% of the II input area selected. It is suggested that a thin penetrometer be placed in the X-ray field. For modern digital systems (without film recording) a light meter can measure the intensity levels behind the lead disk and the surrounding background on the display monitor using standardized contrast and brightness control settings. The contrast ratio is a ratio of the light intensity in the area surrounding the disk to the area directly under the disk. For satisfactory fluoroscopy image receptors, the contrast ratio measured with a light meter should be $>40:1$ – $60:1$. For film type measurements, the ratio should be $>20:1$.

CONVERSION GAIN

The conversion gain of an II defines the efficiency by which the device can convert incident X-ray radiation at the II entrance into light output at the output phosphor of the II. Conversion gain (G_x) is measured in units of nit ($\text{cd}\cdot\text{m}^{-2}$) of light output per $\text{mR}\cdot\text{s}^{-1}$ of X-ray radiation input. The conversion gain degrades with age and usage; typical amounts for the decrease in conversion gain are 5–10% year^{-1} .

As the II conversion gain decreases the patient radiation dose will increase and the X-ray tube heat loading will increase. Although II conversion gain measurements can be performed by medical physicists, it is usually better to request the manufacturer's service personnel to do the measurement (31,32). Conversion gain should be measured immediately after installation when the II is new. By the time that the conversion gain drops to 50–60% of its initial value, replacement of the II would be recommended.

SCATTERED RADIATION LEVELS

Information about the scatter radiation levels is relevant to the assessment of radiation protection criteria for radiology

personnel. In order to limit the radiation to the radiologists to 870 Gy (100 mR) per week, the measured levels should be $<170\text{--}220\text{ Gy}\cdot\text{h}^{-1}$ ($20\text{--}25\text{ mR}\cdot\text{h}^{-1}$) for an estimated maximum of 4–5 h of actual fluoro time per week.

The scatter radiation measurements are performed by placing 25 cm of acrylic in the fluoro beam with the collimator wide open. The ABC is used to control the radiation levels. A portable survey meter is used to measure the scatter radiation levels. It is assumed that protective curtains and flaps will be utilized where they are available. Radiation levels are measured for eye, chest, and gonadal levels at standard positions of staff in the fluoroscopic rooms.

MAMMOGRAPHIC X-RAY EQUIPMENT

Mammographic equipment is a special category of X-ray units. These units can either be used with special film-screen cassettes or as digital units. Mammographic units intended for both types of units typically have a molybdenum X-ray tube target with a beryllium window and either a molybdenum or rhodium X-ray beam filter. (Some mammography units are available with either rhodium or tungsten X-ray tube targets.) Digital units typically contain an image receptor with a scintillator phosphor coupled to photodiode array or a direct conversion amorphous selenium radiation detector array. Both types of mammography units operate at low peak kilovoltages usually with X-ray tube potentials of 25–35 kVp. The measured HVLs will depend on the type of unit being tested. Federal regulations specify that the HVL at 30 kVp must be at least 0.33 mm of aluminum equivalent using ultra pure aluminum filters (not 1100 Al) (see table in Half Value Layer Determination).

At these low peak kilovoltages, small irregularities in the compression plate, grid, or cassette can produce artifacts on the radiographs. Even a small piece of tape (or dirt) inadvertently placed on the compression plate can create objectional artifacts. Another consideration is that the female breast is radiation sensitive. Therefore, the radiation dosage during mammography should be kept as low as reasonably possible (ALARA). Finally, mammography strives to achieve better spatial resolution than other types of film-screen radiographs. To achieve these goals, an X-ray tube with a small focal spot size is used to minimize geometric blur. The film-screen cassette is also designed to provide excellent film-screen contact. Beyond the aforementioned factors, the mammographic equipment has many features in common with the standard radiographic units. The QC procedures for mammographic equipment must be expanded to encompass their special features. Specialized tests required for mammographic units are listed below (33–35)

TYPICAL PATIENT DOSE

There are a number of methods to specify the patient dose. One can either measure a skin entrance dose, a mid-breast dose, or an average glandular dose; the average glandular dose is being recommended due to its biological significance

(34). Because of the ease of measurement, however, the skin entrance exposure value is most often determined. The measurement is dependent on breast size, breast composition, peak kilovoltage used, anode and filter combination used, and the type of image being recorded (digital or film-screen). To simplify the measurement, 4–5 cm of BR-12 plastic or acrylic plastic can be used to represent a typical breast. The exposure should be taken to obtain normal densities or digital exposure settings for the image. The skin entrance exposure is then measured with a radiation detector that has a good low energy response. The measured radiation with grid should be 5.2–13 mGy (600–1500 mR) dependent on: the type of mammography unit, film-screen or digital selection, image processing conditions, film density, anode type, filtration type, and kVp used. The average glandular radiation dose for a 4.2 cm compressed breast must be $<3.0\text{ mGy}$ (300 mrad). Typical values for both film-screen and digital units with a 4.2 cm ACR phantom are $\sim 1.5\text{--}2.0\text{ mGy}$ (150–200 mrad). Although the use of grids increases the measured exposure values by 2.0–2.5 times, grids are essential to obtain good image contrast.

SPATIAL RESOLUTION

To assess the spatial resolution of the imaging system, a bar pattern (same one described previously in radiography and fluoroscopy section) should be placed on top of a 4–5 cm plastic block (BR-12 or acrylic) that simulates a typical breast. An image should be recorded and the spatial resolution seen on the image should be determined. This measurement combines the effects of the focal spot and image system blur. In general, the spatial resolution of most mammography systems is excellent. It should be anticipated that the measured spatial resolution should be >13 line pairs per millimeter with bars along the anode-cathode direction and 11 line pairs per millimeter with bars in the perpendicular direction for film-screen systems. The spatial resolution for digital systems is limited by the detector pixel size; and it is usually in the 5–10 line pairs per millimeter range for current technology.

ARTIFACTS

To identify artifacts, an image of the 3–5 cm acrylic sheets should be imaged. The density should be appropriate and uniform across the surface. Any changes in density are indicative of potential artifacts.

IMAGE QUALITY

Good mammographic image quality involves an assessment of numerous factors. The QC procedures should require that a radiographic image of a phantom that includes several different types of test objects be taken. Figure 9 is an image of the ACR phantom that contains fibrils, calcium specks, simulated masses, and a plastic disk for contrast assessment. For film-screen units, at least 4 fibrils, 3 speck groups, and 3 masses must be clearly

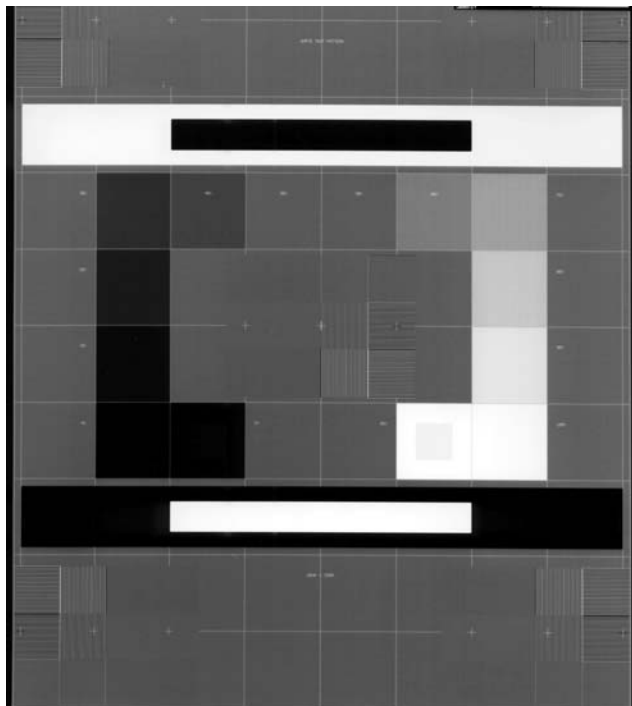


Figure 9. CT image of ACR CT phantom showing plastic rods for CT number linearity and wire for slice thickness evaluation.

visible for a total score of 10 objects (according to ACR and MQSA standards). Digital mammography systems should be able to visualize even more objects in the phantom. The ACR Mammography Accreditation Manual should be consulted for details about many of the mammography equipment tests.

TRAY TRANSMISSION

Since the patient's lap is often situated beneath the cassette tray, it is important that the back surface have sufficient lead lining to significantly attenuate or block the X-ray beam. A radiation detector should be used to measure the amount of radiation transmitted during typical mammographic exposures. It is recommended that the transmission be <0.01% of the primary beam.

BODY-SECTION TOMOGRAPHIC EQUIPMENT

The purpose of the equipment is to image thin planar sections through the patient's anatomy using film-screen or CR cassettes. The X-ray tube and cassette move in opposite directions about a pivot (or focal) plane to blur objects outside this focal plane. The motion is done such that the anatomy at a preselected depth in the patient will remain in focus on the image receptor cassette. All other anatomy outside the focal plane will be blurred by the motion.

The QC procedures of these tomographic units should include additional tests directed toward analyzing the tube motion and the planar radiographs (36).

PINHOLE IMAGES

The angle through which the X-ray exposure is made influences the thickness of the imaged tomography planar sections. Moreover, it is important that the motion be uniform and symmetric about a central point. All these factors can be assessed through a pinhole image.

To do this test, a lead sheet with a small hole [$\sim\frac{1}{8}$ in. (0.32 cm)] drilled in it should be suspended several inches above the tabletop at the height of the focal plane. A film-screen cassette should be placed on the table top beneath the hole in the lead sheet. With the X-ray tube at zero degree tomo angle, the pinhole should be placed along the central ray and a short exposure taken with no tomo motion to mark this position. Then, a routine tomographic exposure should be made and the cassette film developed. The track on the film should be of uniform density and symmetrical about the zero tomo angle point. The tomographic angle can be computed from the height of the pinhole above the cassette (h) and the width of the track on the film (W). For linear tomography, the tomographic angle is given by the following equation:

$$\theta = 2 \tan^{-1} \frac{W}{2h}$$

The measured exposure angle should correspond to the selected angle within $\pm 1^\circ$.

SLICE THICKNESS

A standard tomographic phantom has been designed to measure the slice thickness and image quality for the body section units (37). The phantom is merely placed on the table and a preselected depth is chosen. A routine tomographic scan is taken and the image developed. Figure 10 shows a tomographic image through this phantom. Examination of the portion of the image that is not blurred yields the depth accuracy, slice thickness, and image quality. It is recommended that both the location of focal plane depth and the slice thickness be accurate to within ± 1.0 mm.

COMPUTED TOMOGRAPHY

Computed tomography scanners are composed of many major subsystems. The hardware contains a variety of complex equipment, such as the X-ray source, the rotating mechanical assemblies in the gantry, radiation detectors, signal-processing electronics, computer systems, display systems, and input-output devices (38). The X-ray source consists of a standard X-ray tube and generator. This portion of the equipment can be tested in the routine manner to check the peak kilovoltage, milliamperage, and time calibration. Because it is difficult to evaluate each of the other CT subsystems individually, the QC tests assess the overall image quality through the use of phantoms. Other tests are directed toward measurement of the patient dosimetry, slice geometry, and table motion. There are several different types of CT phantoms that can be used for QC testing (39-42). The procedures that will be described here are based upon use of the ACR CT Accreditation

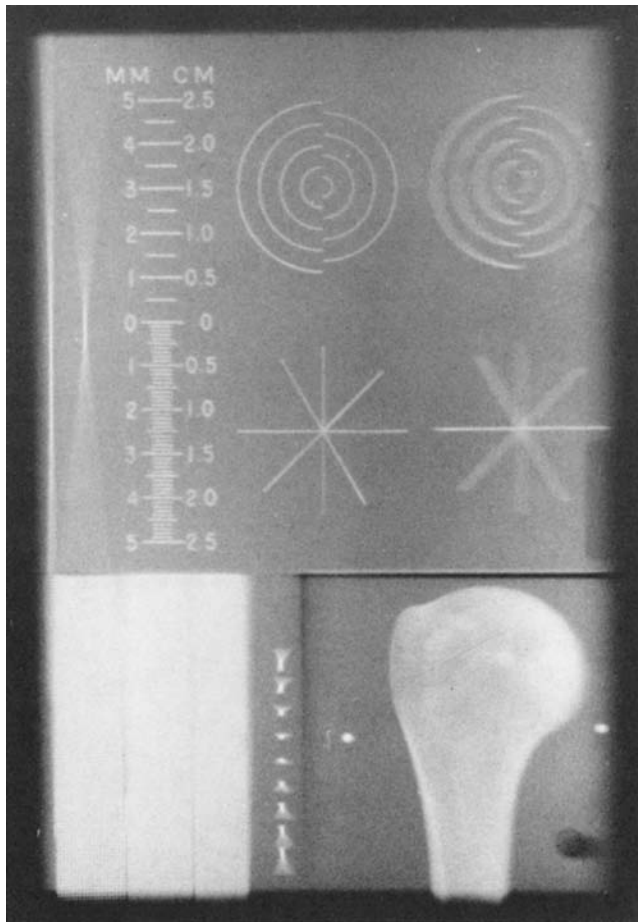


Figure 10. Radiograph of a tomographic phantom used to assess performance of linear tomographic units.

phantom. Similarly, there are many methods by which the CT dosimetry can be performed; however, most often a specialized CT ionization chamber with 100 mm length is used. The chamber is placed in holes drilled in cylindrical acrylic phantoms with a 16 cm diameter cylindrical head and 32 cm diameter body phantom are used for this purpose.

CT NUMBER LINEARITY

The CT numbers of a computerized axial tomographic unit represent the effective linear attenuation coefficients of the various materials being analyzed. In order to be able to differentiate various kinds of body tissues, the CT scanner must be able to measure the effective linear attenuation coefficients correctly for a wide range of materials from air to compact bone, that is, the CT scanner must be able to detect small differences in the way different substances stop X-rays. Usually, the CT numbers (in Hounsfield units) represent the difference in the linear attenuation coefficients between air and water as 1000 units. The CT number linearity is determined by plotting the CT numbers for known materials (usually plastics) versus their effective linear attenuation coefficients, as published in the literature. Linear attenuation coefficients are a function of the photon energy, as well as the X-ray beam filtration. A

typical average energy for a well filtered 120 kVp X-ray beam is ~ 70 keV; this effective energy is often used for this assessment.

One section of the CT phantom contains various rods of plastic embedded in a water equivalent plastic cylinder that has a diameter of 20 cm. The rods consist of air, polyethylene, acrylic, and simulated bone. The anticipated CT numbers for these materials at 120 kVp are approximately -1000 , -95 , 120 , and 955 Hounsfield Units (HU), respectively. Water (or water equivalent plastic) has a reference value of zero. The CT numbers for the various substances are plotted against their attenuation coefficients. The linear regression fit to the data should have a correlation coefficient of >0.99 . In order to compare different types of tissue by their CT numbers, it is important to eliminate both CT number offsets and/or slope differences in the linearity scale.

CONTRAST SCALE

The contrast scale for any given CT scanner can be arbitrary. If calibrated properly, however, the CT number of air should be -1000 and the CT number of water should be zero. Compact bone should have a large positive value near $+1000$ or more. In this manner, an acceptable gray scale will be established for the various anatomical substances shown in scans of patients.

The contrast scale is the change in the effective attenuation coefficient per CT number unit. Usually, measurements are performed with reference to plexiglass and water. The contrast scale (CS) is represented by the following equation:

$$CS = \frac{\mu_{\text{plex}} - \mu_{\text{water}}}{(\text{CT No.})_{\text{plex}} - (\text{CT No.})_{\text{water}}}$$

For most CT energies, the numerator of the equation is equal to $\sim 0.024 \text{ cm}^{-1}$. The CS value should be ~ 0.0002 .

In order to determine whether the contrast scale is suitable, one must compute the anticipated CT numbers for air and compact bone based upon the contrast scale measurements.

$$(\text{CT No.})_x = \frac{\mu_x - \mu_{\text{water}}}{CS}$$

A number of references list the linear attenuation coefficients for air and compact bone. For 120 kVp, $\mu_{\text{air}} = 0.000208 \text{ cm}^{-1}$ and $\mu_{\text{bone}} = 0.414 \text{ cm}^{-1}$. For water, μ_{water} is usually taken to be 0.190 cm^{-1} . Hence, based upon the measurement of CS, the CT numbers of air, water, and bone can be estimated. Usually the CT units set air at -1000 and compact bone at $+1000$ HU. (Bone actually has a wide range of CT number values due to the various densities it may have in the human body.) A comparison of the measured values with the ideal values indicates whether the CS is satisfactory.

COMPUTER TOMOGRAPHY NOISE LEVELS

When one images a uniform material in a CT scanner, the CT numbers for a localized region should all be the same. In

practice, the CT numbers vary around some average value. The reason for these variations are changes in the X-ray output, small differences and drifts in the detectors, extraneous electronic noise, and small errors in the computer processing algorithm. All of these parameters introduce variations in the CT numbers. The variations are lumped collectively into a term called the noise of the CT system.

The noise is determined by scanning a uniform water-bath section of the CT phantom and evaluating all the different CT numbers in a small region. The mean value (X) and the standard deviation (σ) of the CT numbers are computed. The CT noise is then expressed as a percentage variation of the linear attenuation coefficient of water.

$$\% \sigma_{\text{water}} = \frac{(100 \sigma \text{CS})}{\mu_{\text{water}}}$$

The noise value should be measured at various locations in the phantom; both the average noise value for the various locations and the worst case value should be quoted (see Fig. 11).

The noise values for CT scanners are very important performance parameters. The difference in the attenuation coefficients between normal and pathological tissue is small. High CT noise values tend to obscure small tissue differences. For example, differences in the CT values between white and gray matter in the brain are ~ 0.5 – 0.6% relative to water (1% on most CT scanners are ~ 10 CT numbers). In fact, most body tissues have CT numbers of nearly the same value, $\pm 8\%$ (except for bone). Furthermore, the use of lower peak kilovoltages also tends to differentiate tissue better. CT noise levels for a 10 mm slice thickness with 0.50 mm pixel sizes should be below 0.5% of the linear attenuation value of water.



Figure 11. Society of Motion Picture and Television Engineers (SMPTE) test pattern for the evaluation of VDTs used with digital imaging systems.

The measured CT noise levels, however, are a function of many parameters. The noise level decreases as the radiation dose is increased. As the slice thickness is increased, the noise level decreases. The noise levels are also higher for large patients. The relationship to the various factors to the noise level is usually given as the following equation:

$$\% \sigma_{\text{water}} \propto [B/(HW^3D)]^{0.5}$$

where B is percentage attenuation, H is slice thickness, W is pixel width size, and D is patient dose per scan. In many modern CT units, the noise changes little with pixel size. CT noise is also closely related to the type of reconstruction kernel chosen. The CT noise is greater for edge enhancement kernels, and CT noise is less for smoothing kernels. To maintain quality control of the CT scanner, noise levels should be monitored on a regular basis. An increase in the noise level is indicative of a developing problem in the CT scanner.

SPATIAL UNIFORMITY

It is extremely important that the response of a CT scanner be uniform across the field being scanned. A CT scan of a uniform material should only exhibit random fluctuations due to system noise. No areas of nonuniform response nor artifacts should be appearing in the scan. This feature is tested by scanning uniform water equivalent plastic section in the CT phantom and examining the CT numbers in different portions of the image. The mean CT numbers should be the same in all portions of the scan. The difference between the average CT number in the various portions of the image should not differ by > 2 – $3 \times$ values the standard deviation of the CT numbers (or ~ 10 CT numbers).

HIGH CONTRAST SPATIAL RESOLUTION (HOLE PATTERN)

One of the performance criteria of CT scanners is their ability to image small objects. The ability of the scanner to image small objects depends on the subject contrast. The CT scanners have the best spatial resolution with high subject contrast. The ability of the CT scanner to image small, high contrast objects can be measured by a section of the ACR CT phantom; this section has many plastic bars and spaces of different sizes. These bars are of vastly different CT number from the plastic in the spaces to create high contrast, spatial resolution test objects. For each set of bars, a spatial frequency in line pair per cm can be assigned. The frequency ranges from 4 to 12 line pairs per cm. The CT display window and level settings are adjusted to best display bars and spaces. The just barely discernable bar pattern is identified and recorded. It has been suggested that a scanner be capable of resolution approximately twice its smallest pixel size (for pixel-limited resolution). With a 512×512 matrix and a 25 cm FoV, the spatial resolution would be limited by pixel size to ~ 10 line pairs per cm. For small FoV, the spatial resolution is limited by factors, such as the focal spot size, detector size, and reconstruction algorithms. The modern CT units are capable of spatial resolution of ~ 8 – 10 line

pairs per cm for high contrast objects using image zoom display features.

HIGH CONTRAST SPATIAL RESOLUTION (IMPULSE RESPONSE)

Another way of determining scanner resolution is the scanning of a metal pin in a water bath. This scan will produce a point spread response function (PSF). One usually plots the numbers and draws a smooth curve. The point spread for a CT unit can usually be fit with a Gaussian function given below:

$$F(x) = A \exp[-\alpha^2(x - \epsilon)^2]$$

This equation can be rearranged in order to determine the value of α and ϵ . The modulation transfer function (MTF) can be computed directly using the value α (43).

LOW CONTRAST RESOLUTION

The CT scanner must not only be able to detect small objects with high contrast, but it must also be able to image small objects with low subject contrast. The difference between normal tissue and pathology is usually small differences in subject contrast (1–2% differences). A good CT scanner should be able to detect small low contrast lesions, that is, low contrast is more important than high contrast resolution.

In order to test low contrast resolution, CT scans are taken through a section in the phantom composed of a plastic with CT numbers only slightly different from water. This section contains various size holes of a slightly different plastic than the surrounding material. The contrast between the plastic and the water is usually 0.4–0.6%. The detectability is dependent upon the radiation dose used, but it would be desirable to have low contrast discrimination of 4–6 mm holes at 0.5–0.6% contrast.

BEAM HARDENING

X-ray beams are polychromatic. Therefore, filtration changes the X-ray spectrum and its effective energy. When CT scanner X-ray beams pass through a lot of bone or long paths through tissue, the X-ray spectrum is altered. The effective linear attenuation coefficients for these hardened X-ray beams change because attenuation is a function of X-ray energy. Therefore, the CT numbers of tissue behind extensive sections of bone may decrease appreciably due to beam hardening. These effects are highly undesirable as some tissues appear abnormal in CT number. This beam hardening artifacts in CT numbers are undesirable.

In order to evaluate beam-hardening effects, 1 in. (2.5 cm) diameter plastic rods with high attenuation coefficients are taped on the opposite sides of the surface of the water section of a CT phantom. This section is scanned, and the CT numbers along the line between the rods are measured. CT scanners should be designed in a manner that the water values are not depressed more than three times the noise standard deviation.

IMAGED CT SLICE THICKNESS

The CT image slice thickness (i.e., sensitivity profile) is measured by scanning a section of the phantom that contains small wires that are displaced along the longitudinal axis of the phantom by a 0.5 mm distance and angled. By counting the number of wires visible in the image, the image slice thickness can be assessed. The measured sensitivity slice thickness should be within ± 0.50 mm of the selected slice values.

PATIENT TABLE TOP INDEXING

It is extremely important that the patient table top moves by the selected distance, otherwise either clinical information or the patient radiation exposure is compromised. This motion is tested by taping a radiation therapy verification film to the outside surface of a cylindrical phantom. Afterward, a series of scans are taken. Each slice thickness setting should be tried; the table motion should be equal to the slice thickness. Three or more consecutive scans should be taken for each collimator setting; the table should index at least 20 mm between different slice thickness settings. Following these scans, the film is developed and examined. There should neither be gaps nor overlaps for consecutive scans. Moreover, the table motion should be accurate to within 1.0 mm for 100 mm of motion in 10 or more steps.

ARTIFACTS

Artifacts can be misinterpreted as anatomical abnormalities in patient scans. Various software and scan techniques should be used to scan the water section of the CT phantom. Since water equivalent plastic should be homogeneous, any artifacts should be clearly visible in the CT image. The various window level and width settings should be utilized to view the image.

PATIENT DOSIMETRY

A special CT ionization chamber (100 mm long) is inserted into cylindrical acrylic phantoms designed to simulate a typical head and body of a patient. The head CT phantom has a 16 cm diameter and the body phantom has a 32 cm diameter. Each phantom has holes drilled at the center and at other locations with different distances from the surface of the phantom. At least one hole is 1 cm inside the outer surface. The ionization chamber is placed in both the center hole and later at the periphery hole. Empty holes are filled with acrylic rods. A single, axial CT slice is taken in the middle of the cylinder, and the exposure measured by the ionization chamber (E_m) is recorded. The radiation dose is then computed by the following equation:

$$CTDI = (C \times f \times E_m \times 100 \text{ mm}/XW)$$

where C = calibration factor for detector, f = exposure to absorbed dose conversion factor, E_m = measure exposure, XW = X-ray beam width specified.

The f -factor for acrylic is $0.78 \text{ cGy}\cdot\text{R}^{-1}$, and the f -factor for tissue is about $0.92 \text{ cGy}\cdot\text{R}^{-1}$. The f -factor for air is $\sim 0.87 \text{ cGy}\cdot\text{R}^{-1}$. A weighted CTDI (CTDI_W) is calculated using (1/3) of the center value and (2/3) of the peripheral value.

For these data, the actual radiation dose values can be scaled using clinical setting of kVp, milliampere-second, and pitch. Pitch is the table movement between rotation of the X-ray tube divided by the width of the X-ray beam. The $\text{CTDI}_{\text{volume}}$ is equal to the CTDI_W divided by the pitch. It is recommended that the $\text{CTDI}_{\text{volume}}$ for clinical head CT scans of patients be $<6.0 \text{ cGy (rad)}$ and that the body value is $<3.5 \text{ cGy (rad)}$.

FILM PROCESSORS

Because the film processing affects the end product of film-screen radiographic images, this is one of the most important aspects of the QC program for diagnostic radiology. Film processing conditions can affect film contrast, the speed of the film-screen system, and base plus fog levels. Indirectly, inadequate processing usually causes the technologist to alter kilovoltage or mAs that could result in an increased dose to the patient.

In general, a 1°F (0.6°C) temperature change in the developer chemistry produces a $0.03\text{--}0.07$ density change in the developed radiograph. The film density decreases for a decrease in the developer temperature. For an $8\text{--}18^\circ\text{F}$ ($4.4\text{--}10^\circ\text{C}$) decrease in the developer temperature, a 50% increase in the patient exposure is necessary. Similarly, increasing the developer temperature by 5°F (2.8°C) can cause a latitude film to appear to have increased contrast. Contamination of the processing chemistry also causes changes in the film density and contrast. The magnitude of the effects are dependent on the type of film, the kind of chemistry, and the processor design.

In order to maintain consistency in the film development, a routine maintenance program for the processor should be established. The replenishment rates should be adjusted properly for the workload. The rollers should be cleaned and inspected on a regular basis and the chemistry should be changed as required. Moreover, steps should be taken to minimize inconsistencies in the chemistry. Within reason, all processors should operate from the same batch of fixer and developer; this can be accomplished by mixing large batches for delivery or using electronic mixers with modular chemistry canisters.

The QC program for processors should utilize daily sensitometry strips that are run through each processor (44). In conjunction with the sensitometry, the processor replenishment rates and temperatures should be checked. Light sensitometers are recommended for making the sensitometry strips (45,46) (see Fig. 12). The film used for sensitometry should be the fastest high contrast film routinely used in the facility. Since variations can be found in different film emissions batches of the same kind of film, strips from both old and new emulsion batches should be run simultaneously during transition periods. The processed sensitometry strips should have the densities measured with a densitometer. A sensitometer step

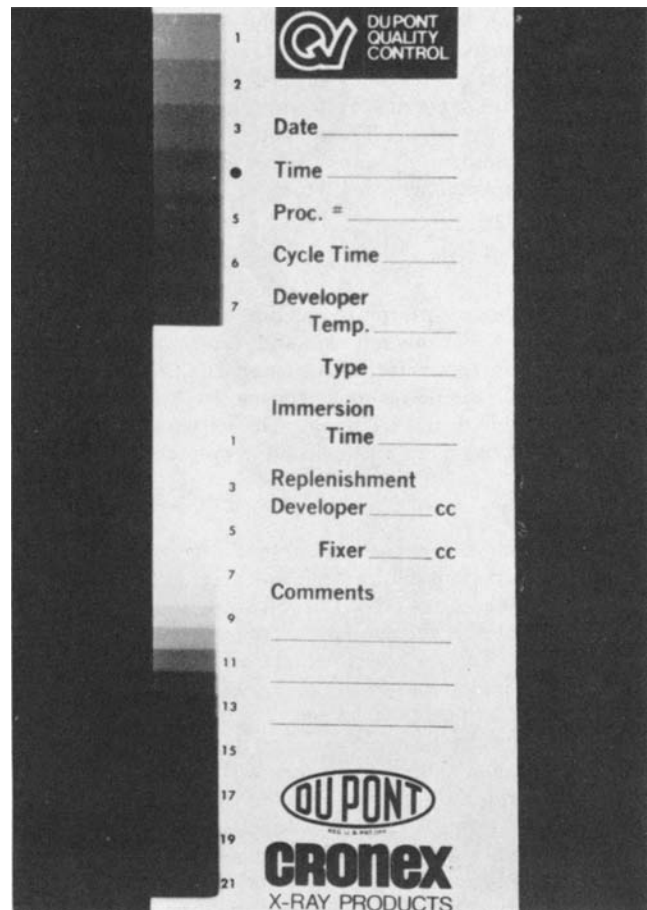


Figure 12. A reproduction of a developed sensitometry strip that is used to evaluate processor performance. The steps used for speed point and contrast index are shown on the strip.

with a density $\sim 1.0 \text{ du}$ above the base plus fog level should be selected as the speed point. The difference in densities of two points in the linear portion of the characteristic curve should be used to determine a contrast index. The base plus fog level on the sensitometry strips should also be measured. These three parameters should be compared to optimal values supplied by the film manufacturer and average values should be established for the facility. The speed point should not be allowed to change by more than $\pm 0.15 \text{ du}$ about the average value. The contrast index should not change more than ± 0.15 units. Since the base plus fog is not very sensitive to variations, a small change is indicative of significant abnormalities in the processing. With good maintenance and QC procedures, the film processors should be very stable and consistent.

DIGITAL IMAGE RECEPTORS

Digital image receptors could be either CR Cassettes with their electronic readers or DR image receptors (47). With both systems, a QC program should be routinely implemented to ensure consistency, identify artifacts and evaluate physicians display monitor (48). Most CR and DR

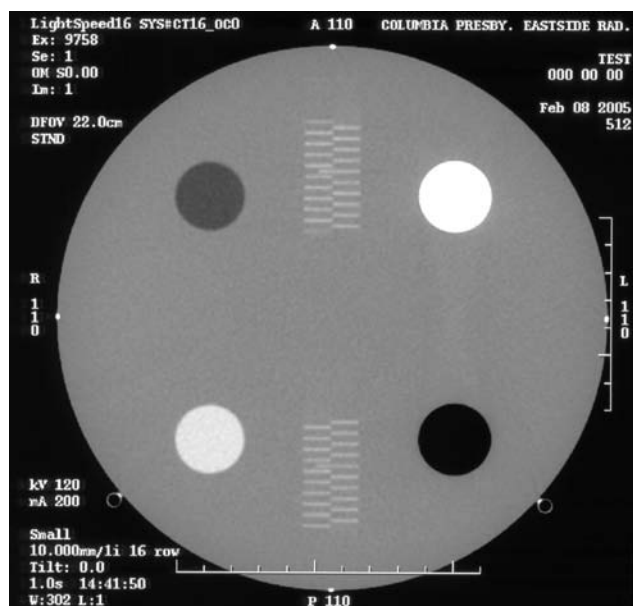


Figure 13. Radiographic image of an ACR Mammography phantom showing fibrils, calcium, speck group, and simulated masses.

systems have an index number to ensure sufficient radiation is utilized to avoid excessive quantum mottle. The QC program should assess that the AEC is properly adjusted to maintain proper radiation exposure. For CR systems, cassette should be exposed to measure spatial linearity, spatial resolution, uniformity of density and proper size and collimation. Both CR and DR units should be evaluated for artifacts and dirt. The physicians' review video monitor should be checked with a video pattern such as the SMPTE (Society of Motion Picture & Television Engineers) for low contrast visibility, spatial linearity, spatial resolution and a reasonable contrast scale. A light meter can then be employed to ensure that maximum luminance and contrast scale are adjusted similarly on all video monitors (Fig. 13).

SUMMARY

There are numerous reasons for providing a QC program in diagnostic radiology. Quality control is mandated by several regulatory agencies; hence a radiology facility must demonstrate that some type of QC is being performed. Another reason for QC programs is to ensure electrical, mechanical and radiation safety for patients and staff. Moreover, the QC program should provide better image quality and a consistency in the radiographs. Finally, the QC program should result in financial savings and improved equipment performance.

The effectiveness of the QC program will depend on the extent of testing performed. A cursory QC program is usually conducted to just meet minimum regulatory requirements. A more extensive program will attempt to establish and maintain a high standard for equipment performance.

Because of the diversity in the diagnostic equipment, the QC program should be directed toward evaluating the special features of each type of equipment. Hence, there are

many specialized aspects to QC testing. Although other testing could be used to supplement and strengthen the QC procedures described, the information given in this article provides a basis for establishing an effective QC program for diagnostic radiology equipment.

BIBLIOGRAPHY

1. Joint Commission on Accreditation of Hospitals. Accreditation Manual for Hospitals. Chicago, IL. Available at <http://www.JCAHO.org>. 2005.
2. American College of Radiology (ACR). Standards. Reston (VA) ACR, 2001–2002 or <http://www.acr.org>.
3. U.S. Department of Health, FDA, Center for Devices & Radiological Health (CDRH). Routine Compliance Testing Procedures for Diagnostic X-ray Systems. Available at <http://www.fda.gov/cdrh/radhlth/xraytestproc.html> 2003.
4. U.S. Department of Health, FDA, Center for (CDRN) Devices and Radiological Health (CDRN). Mammograph facility survey, equipment evaluation and medical physical qualification requirements under MQSA. Available at <http://www.fda.gov/cdrh/dmqrp/6409.pdf> 2000.
5. U.S. Department of Health, FDA, Center for Devices and Radiological Health (CDRH) Resource manual for compliance test parameter of diagnostic X-ray systems. Available at http://www.fda.gov/cdrh/comp/rad_medical.html 2005.
6. Ostinelli A, et al. Quality assurance in diagnostic radiology: Practical outcomes. *Rad. Medica* Oct 2003;106(4):413–419.
7. Noyes RS. The economics of quality assurance. *Radiol/Nucl Med Mag* Dec 1980.
8. U.S. Department of Health, Education and Welfare. Quality assurance programs for diagnostic radiology facilities (FDA) 80-8110, Washington (DC): U.S. Govt. Printing Office; 1980.
9. Peer S, et al. Comparative reject analysis in conventional film-screen and digital storage phosphor radiography. *Rad Protection Dosim* 2001;94(1-2):68–71.
10. Nelson RE, Barnes GT, Wittten DM. Economic analysis of a comprehensive quality assurance program. *Radiol Technol* 1997;49:129–134.
11. Fields T, Griffith CR, Hubbard LB. What price quality? A quality assurance program for diagnostic radiology. *Appl Radiol* Jan–Feb 1980;57–64
12. U.S. National Council on Radiation Protection and Measurements (NCRP). NCRP Report No. 99 quality assurance for diagnostic radiology, Bethesda (MD); 1980.
13. Waggner RG, Wilson CR, editors. Quality Assurance in Diagnostic Radiology. New York: American Institute of Physics; 1980.
14. Gray JE, Winkler NT, Stears J, Frank ED. Quality Control in Diagnostic Imaging. Baltimore: University Park Press; 1983.
15. American College of Radiology (ACR). Accreditation Programs (Mammography, MRI, CT, Fluoroscopy and Ultrasound). Available at <http://www.acr.org>. 2005.
16. American Association of Physicists in Medicine (AAPM). Quality Control in Diagnostic Radiology. Madison (WI): Medical Physics Publishing; 2002.
17. Hospital Physicist's Association. Quality Assurance Measurements in Diagnostic Radiology. Rep. No. 29, London: HPA; 1979.
18. U.S. Department of Health, Education and Welfare. Regulation for the administration and enforcement of radiation control for health and safety act of 1968. (FDA) 75–8003, Washington (DC): U.S. Govt Printing Office; 1976.
19. Seibert JA, Barnes GT, Gould RG. Specifications, Acceptance Testing and Quality Control of Diagnostic X-ray Equipment. New York: Springer-Verlag; 1997.

20. Hendee WR, Ritenour ER. *Medical Imaging Physics*. 4th ed New York: Wiley; 2002.
21. Bushberg JT, Siebert JA, Leidholdt EM. *Essential Physics of Medical Imaging*. 2nd ed. Baltimore: Williams & Wilkins; 2001.
22. Stears JG, Felmlee JP, Gray JE. Half-value layer increase owing to tungsten build-up in the x-ray tube: Fact or fiction. *Radiology* 1986;160:837–838.
23. Spiegler P, Breckinridge WC. Imaging of focal spots by means of the star test pattern. *Radiology* 1972;102(3):679–684.
24. Hendee WR, Chaney EL. X-ray focal spots: Practical consideration. *Appl Radiol* 1974;3(3):25–29.
25. Atherton JV, Nickoloff EL. Non-invasive X-ray tube current measurement. *Med Phys* 1987;14(2):258–261.
26. Chaney EL, Hendee WR. An Instrument with digital readout for indirect determination of kVp. *Med Phys* 1978;5(2):141–145.
27. Ramirez-Jamenez FJ, et al. Consideration of the measurement of practical peak voltages in diagnostic radiology. *BJR* Sept., 2004;77(921):745–750.
28. Healey T, Dickson DG, Greenwood MWB. A calibration system for x-ray generators and tube factor. *Br J Radiol* 1979;52:44–50.
29. Giarratano JC, Waggner RG, Hevezi JM, Shalek RJ. Comparison of voltage-divide, modified Ardran-Crooks cassette, and Ge (Li) spectrometer methods to determine the peak kilovoltage (kVp) of diagnostics X-ray units. *Med Phys* 1976;3:142–147.
30. U.S. Department of Health, FDA. *Quality Assurance for Fluoroscopy X-Ray Units and Associated Equipment*, FDA 80-8095, Washington (DC): U.S. Government Printing Office; 1979.
31. Holm T, Moseley RD. The conversion factor for image intensifier. *Radiology* 1964;82:898–904.
32. Hay GA, Clarke OF, Coleman NJ, Cowen AR. A set of X-ray test objects for quality control in television fluoroscopy. 1985;April *BJR* 58(688):335–344.
33. U.S. National Council on Radiation Protection and Measurements (NCRP). *NCRP Report No. 85. A user's guide to mammography*. Washington (DC); 1986.
34. Hendrick RE, Botsco M, Plott CM. Quality control in mammography. *Radiol Clin N Am* Nov 1998;33(6):1041–1057.
35. Haus AG. Screen-film mammography update: X-ray units, breast compression, grids, screen-film characteristics and radiation dose. *Proc SPIE* 1984;486.
36. U.S. Department of Health Education and Welfare. *Quality assurance for conventional tomographic X-ray units*. (FDA) 80-8096. Washington (DC): U.S. Govt. Printing Office; 1979.
37. Littleton JT. A phantom method to evaluate the clinical effectiveness of a tomographic device. *Am J Roentgenol Rad Ther* 1970;58:139–145.
38. Nickoloff EL. What to look for when buying CT equipment? *Appl Radiol* 1982;11(3):69–74.
39. American Association of Physicists in Medicine (AAPM), *AAPM Report No. 39 Specification and acceptance testing of computed tomography scanner*. College Park (MD): American Institute of Physics (AIP); 1993.
40. McCollough CH, et al. The phantom portion of the American College of Radiology (ACR) CT accreditation program: Practical tips, artifacts, examples and pitfalls to avoid. *Med Phys* Sept 2004;31(9):2423–2442.
41. Morin RL, Gerber TC, McCollough CH. Physics and dosimetry in computed tomography. *Cardiol Clinics* Nov 2003;21(4): 515–520.
42. Johns HE, et al. Physics of CT scanners: Principles and problems. *Int J Rad Oncol Biol Phys* 1977;3:45–51.
43. Nickoloff EL, Riley R. A simplified approach for modulation function determinations in computed tomography. *Med Phys* 1985;12:437–442.
44. Gray JE. *Photographic quality assurance in diagnostic radiology, nuclear medicine and radiation therapy*. Vols. 1 and 2 (FDA) 76-8043 and (FDA) 77-8018, Washington (DC): U.S. Govt Printing Office; 1976.
45. Blendl C, Buhr E. Comparison of light and x-ray sensitive response of double emulsion films for different processing conditions. *Med Phys* Dec 2001;28(12):2420–2426.
46. Nickoloff EL, Leo F, Reese M. A comparison of five methods of monitoring the precision of automated X-ray film processors. *Radiology* 1978;129:509–514.
47. Rowlands JA. The physics of computed radiography. *Phys Med Biol* Dec 2002;47(23):R123–R166.
48. Samei E, et al. Performance evaluation of computed radiography systems. *Med Phys* March 2001;28(3):361–371.

See also CODES AND REGULATIONS: RADIATION; PHANTOM MATERIALS IN RADIOLOGY; SAFETY PROGRAM, HOSPITAL.

X-RAY SCREEN-FILM SYSTEMS. See SCREEN-FILM SYSTEMS.

X-RAY SIMULATOR. See RADIATION THERAPY SIMULATOR.

X-RAY THERAPY EQUIPMENT, LOW AND MEDIUM ENERGY

C-M CHARLIE MA
Fox Chase Cancer Center
Philadelphia, Pennsylvania

INTRODUCTION

X-rays were used to treat cancer patients within a year of their discovery by Wilhelm Roentgen in November 1895 (1). However, it was not until the invention of the hot cathode tube by Coolidge in 1913 that X-ray beams could be delivered in a stable manner and, hence, the output controlled and measured with any precision. Further improvements in the accuracy of X-ray beam dosimetry occurred with the publication of the first central-axis depth dose tables in 1922 (2) and with the definition of the Roentgen and its use as a standard unit for X-ray dosage in 1928 as recommended by the Second International Congress of Radiology in Stockholm, Sweden.

In those early years, the energy of X-ray beams was limited to 140 kV, and therefore only relatively superficial lesions could be successfully treated. The situation was improved in the early 1920s as higher energy (up to 400 kV) X-ray units became available to permit the treatment of deep-seated lesions (3). Since cobalt-60 treatment units (average γ -ray energy = 1.25 MeV) came into use in the late 1950s and electron linear accelerators capable of producing high energy (≥ 4 MeV), X rays in the early 1960s, the use of kilovoltage X-rays in radiation therapy has drastically declined for deep-seated tumors.

The use of low energy X rays for the treatment of superficial lesions is still popular, even though the use of electrons has replaced them for treating such lesions in

many cancer clinics. Electron beams have finite ranges that are ideal for treating shallow tumors to spare distal normal tissues. However, clinical accelerators that produce electron beams for radiation therapy are more expensive than X-ray units. There is an increased use of low energy X-rays for intraoperative radiation therapy, such as for stereotactic brain irradiation or treatment of the dura, for endocavitary irradiation of rectal cancers, and treatment of other malignant skin lesions. The depth-dose properties of X-rays are considered by some oncologists to be more favorable than electron beams for adequate surface coverage and for treating macroscopic diseases at depths. Medium energy X-rays still play a role in many developing countries, where the technology is easier to support and maintain than electron linear accelerators and safer than cobalt-60 teletherapy units. Medium energy X-rays are also widely used in radiation biology research for *in vitro* cell irradiation or *in vivo* animal experiments.

X-RAY DEVICES FOR RADIATION THERAPY

Medium energy X-rays (also called *orthovoltage* X-rays) are generated by X-ray devices at an accelerating potential between 150 and 500 kV. Low energy X-rays (also called *superficial* X-rays) are generated by X-ray devices at an accelerating potential between 50 and 150 kV. Very low energy X-rays that are generated at an accelerating potential <50 kV have mainly been used in *contact* therapy.

The manner in which X-rays are generated is similar between therapeutic and diagnostic X-rays. Electrons emitted from a heated filament are accelerated in an evacuated tube onto a tungsten anode (also called target), producing bremsstrahlung photons with peak energy up to the accelerating voltage. Because of the intended applications, however, therapeutic and diagnostic X-ray tubes have some very different design features as discussed below.

One difference is the target design for heat dissipation. Diagnostic tubes are generally run for shorter times (from

a fraction of a second to several seconds) at high current (up to 500 mA) leading to high instantaneous rates of heat production. Tungsten is used as the target material since it has a high melting point (3370°C). In addition, rotating targets are used in diagnostic tubes to spread the heat deposition over a large area, thereby increasing the maximum permissible loading of the tube. Therapeutic tubes are often run for longer times (several minutes being common) at relatively low currents (5–30 mA). The instantaneous rate of heat production is low for therapeutic tubes, but the total amount of heat generated during a therapy session can be quite high. Therefore, therapeutic tubes use stationary targets that can be cooled by oil or water. The target is often made of tungsten and is mounted on a massive copper stem that serves as a good heat conductor.

The second difference is the target angle, which effectively defines the X-ray source size. Small target angles in the range of $6\text{--}20^{\circ}$ are used in diagnostic tubes in order to produce small focal spots that will result in a small geometric penumbra of the image. In therapeutic tubes, however, geometric penumbra is not of primary concern. Therefore, therapeutic tubes can use larger target angles in order to produce radiation beams with large field sizes (up to 20×20 cm) at target-to-surface distances (SSD) that are usually much shorter than the SSDs employed in diagnostic tubes. The resultant X-ray source size of a therapeutic tube may be as large as 1.0 cm in diameter.

Another difference is the target shielding for secondary electrons. For therapeutic tubes running at an accelerating potential >200 kV, damages to the tube may result if secondary electrons from the target are allowed to reach the glass envelope. The target in a therapeutic tube is often surrounded by a double-layered shield as shown in Fig. 1 (4). The inner layer is made of copper that stops most electrons while its low atomic number ($Z = 29$) minimizes bremsstrahlung production. The outer layer is made of tungsten ($Z = 74$), which absorbs most stray bremsstrahlung photons produced in the copper layer. A thin beryllium ($Z = 4$) window in the shield below the target absorbs most secondary electrons, but allows the useful X-ray beam

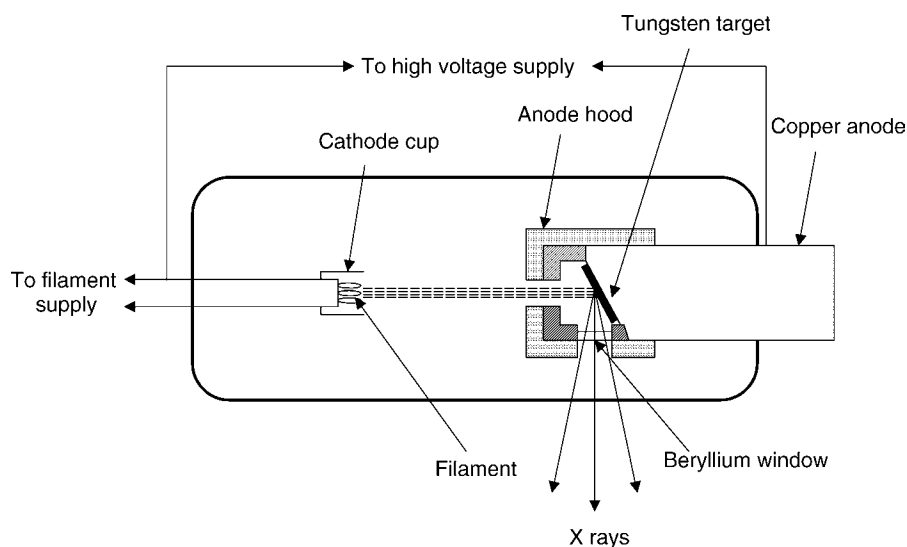


Figure 1. An X-ray tube for radiation therapy constructed with a hooded anode design.

to pass through with little attenuation. This type of shielding design for therapeutic tubes is called a hooded anode. The X-ray tube is enclosed in an oil shield that serves two purposes: (a) to cool the tube and (b) to attenuate stray X rays.

Many major manufacturers have ceased manufacturing kilovoltage X-ray therapy units (5). The three currently available orthovoltage treatment units are Therapax DXT300 manufactured by Pantak, Inc. (East Haven, CN) and Gulmay D3300 and Gulmay D3225 manufactured by Gulmay Medical Ltd. (Mississauga, ON). The same two vendors also produce two superficial therapy units, Therapax HF 150 and Gulmay D2000. Some kilovoltage units are no longer available, but are still in wide use, such as the Philips RT-250 and RT-305 units and the Siemens Stablipan. The most popular contact unit still in clinical use is the Philips RT-50, formerly manufactured by Philips Medical Systems (Shelton, CN). A new, and novel, contact therapy device available is the Photon Radiosurgery System, Model PRS400, manufactured by Photoelectron Corp. (Lexington, MA).

Orthovoltage Units

Most orthovoltage units operate at 200–300 kV with a tube current of 10–20 mA and have various filters to provide beams with half value layers (HVL) between 1 and 4 mm Cu. The X-ray generators used in these units may be single phase, that is, self-rectified or half-wave rectified, or may employ a Villard-type voltage-doubling circuit or a constant potential circuit (6). The constant potential circuit maintains the accelerating potential at a nearly constant value so that X rays can be produced continuously rather than in pulses. This type of circuit allows for treatments at a shorter treatment time with a higher average X-ray energy.

The energy absorbed by the anode of an X-ray tube is proportional to the product of tube current, accelerating potential, and operating time. For orthovoltage units, which may be operated continuously at high tube voltages, the anode must be cooled efficiently. The copper anodes of orthovoltage units are usually massive to serve as heat conductors. Oil circulated cooling is used instead of water circulated cooling, as shown in Fig. 2 (7), where a hollowed-out anode

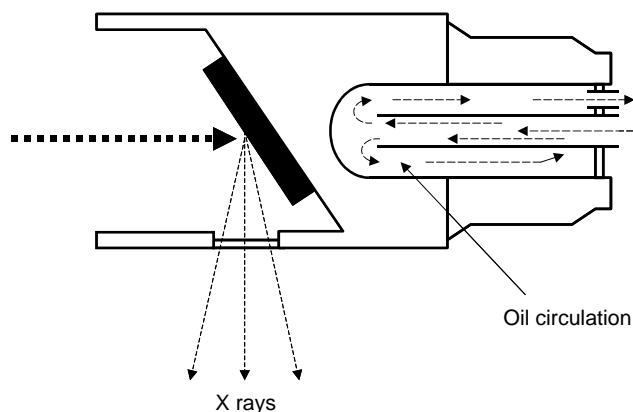


Figure 2. The anode of an orthovoltage X-ray tube. The back of the anode is hollowed out to permit oil circulation.



Figure 3. The Therapax DXT300 orthovoltage unit manufactured by Pantak.

provides efficient oil circulation for heat dissipation. The oil then flows through a reservoir where heat is removed by either circulating water or forced-air cooling. Orthovoltage units usually require a warm-up time of several minutes, which allows the tube potential to increase gradually in order to minimize the strain of high voltage on the tube.

Figure 3 shows a currently available microprocessor-controlled orthovoltage treatment unit, Therapax DXT300 (5). This unit features a dual channel dosimetry system in which the primary means of termination of treatment dose can be either a timer or the integrated signal from a transmission ionization chamber. It features a metal ceramic tube, a control panel that incorporates microprocessor technology, a Diamentor dose monitor, and a Cockcroft–Walton multiplier generator that gives a very stable output. The unit can be ceiling or stand mounted. The Therapax DXT300 unit offers energies and filtrations from 300 kV (3 mm Cu) down to 30 kV (0.1 mm Al). The treatment field shape and size can be defined by a variable collimator (rectangular or square fields up to 20 cm × 20 cm defined at 50 cm SSD) or by fixed square or circular cone-type applicators: open-ended cones (30 cm SSD) or close-ended cones (50 cm SSD).

Superficial Units

Superficial units operate at an accelerating potential 50–150 kV with a tube current 5–15 mA. The X-ray generators in these units are generally half-wave rectified. The X-ray tube has a hooded anode and is enclosed in an oil shield that is constructed with expansion bellows to allow oil to expand as temperature increases. Superficial units may also operate continuously for long treatment times, and therefore require additional cooling in parallel to heat conduction by the copper anode stem and

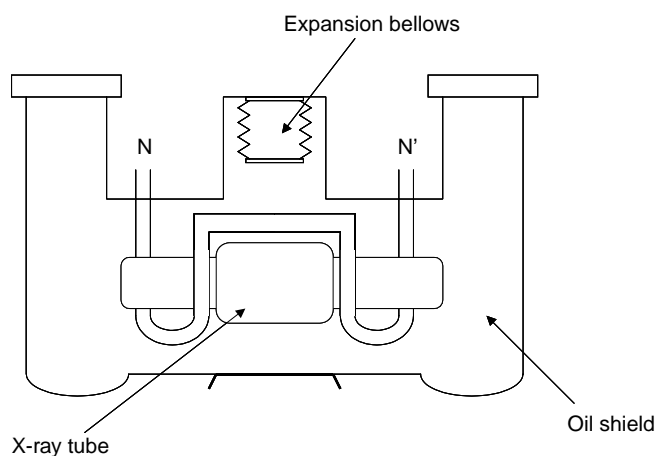


Figure 4. The cooling system for superficial X-ray tubes. Water circulates through the pipe N–N' to remove the heat from the oil.

convection and absorption by the oil shield. Figure 4 shows a cooling system that employs circulating water to remove heat from the oil continuously (7).

Orthovoltage tubes can also produce X-ray beams in the superficial energy range with matching HVLs. The principal difference between a superficial unit and an orthovoltage unit is that a superficial unit typically has shorter SSD cone applications and, consequently, higher dose rates (5). For example, the Therapax HF150 (Fig. 5) is the companion device to Therapax DXT300 and produces X rays in the range 20–150 kV with HVLs from 0.1 to 10.78 mm Al. The DXT300 can also produce X rays down to 30 kV with an HVL of 0.1 mm Al. The HF150 has 15 cm SSD cone applicators while the DXT300 has 30 and 50 cm SSD cone applicators.

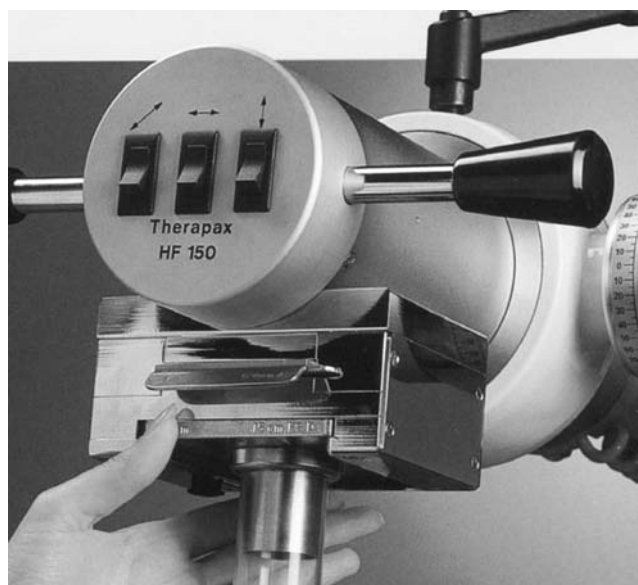


Figure 5. The Therapax HF150 superficial unit manufactured by Pantak.

Contact Units

Both orthovoltage and superficial units can produce X-ray beams in the contact X ray range with complete overlap of available energies and HVLs. The Philips RT-50 unit is strictly a contact therapy unit, operating at energies between 10 and 50 kV. This unit has a slim, long body measuring 43 cm in length and 6 cm in diameter, tapering down to 3 cm at the tip. The SSD is only 4 cm and its dose rate in air is up to $20 \text{ Gy} \cdot \text{min}^{-1}$. These features make it especially suitable for endocavitary irradiation (e.g., for treating rectal cancers).

The Photon Radiosurgery System (PRS) is a special treatment unit (Fig. 6) consisting of an X-ray source, a control console, and associated calibration devices (8). The X-ray source is small, measuring $17.5 \times 11 \times 7 \text{ cm}$ and weighing 1.6 kg, with X rays being produced at the tip of a 10 cm long, 3.2 mm diameter probe. A plastic, biocompatible sheath covers the probe tip to avoid direct tissue contact when used interstitially. The X-ray energy can be 30, 40, or 50 kV with a tube current of 5, 10, 20, or 40 μA . The whole unit is small enough to be hand-held, though it is designed to work in conjunction with a conventional stereotactic head frame. The system has a built-in internal radiation monitor to measure the output of the beam and to determine the length of the treatment. Additional monitors include a timer and an external radiation monitor that has to be calibrated at the start of each treatment since its output is geometry dependent. Other dosimetry devices are provided to measure the isotropy of the beam, the straightness of the probe, and the relative in-air output of the beam during routine calibration and in the operating room immediately prior to the procedure.

X-RAY BEAM QUALITY

X-ray beams with different energy spectra will differ in their ability to penetrate in a medium such as tissue. The quality of an X-ray beam refers to its penetrating power, with high beam quality indicating deeper penetration. The exact knowledge of beam quality plays an important role in clinical radiotherapy to ensure proper treatment target coverage and adequate skin sparing. It is desirable to use more than one beam quality parameter to specify an X-ray beam (9). The usual quantities used are the accelerating potential and the HVL. However, most dosimetric data for kilovoltage X rays have been given using one beam quality parameter, usually HVL.

The peak voltage across the X-ray tube determines the maximum photon energy in an X-ray spectrum. The accelerating potential is expressed in kilovolts for a kilovoltage X-ray beam to indicate that the X-ray beam actually has an energy spectrum with its maximum energy up to its accelerating potential. Higher accelerating potentials produce more penetrating X-rays and, therefore have a direct effect on the beam quality.

The HVL is defined as the thickness of an absorber that reduces the air kerma rate of a narrow unidirectional X-ray beam at a point distant from the absorber to 50% of that of the nonattenuated beam (9). For orthovoltage X rays, both aluminum (Al) and copper (Cu) can be used as the

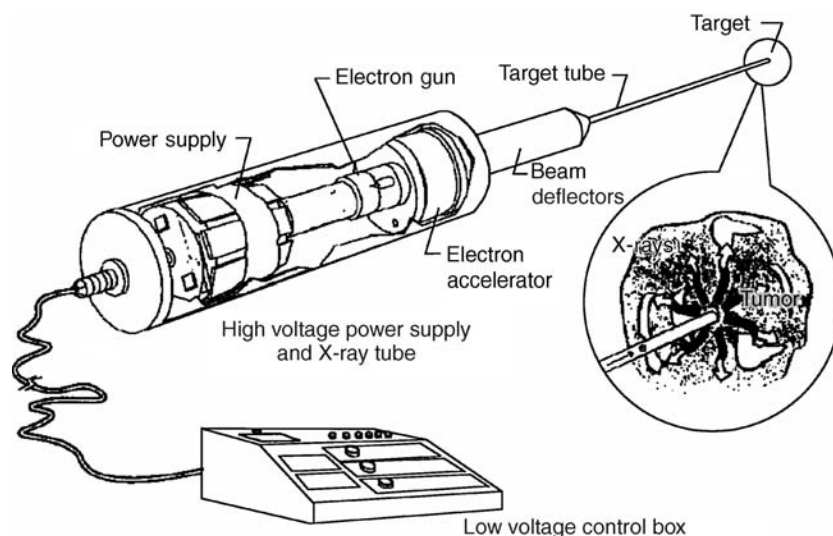


Figure 6. A schematic diagram showing the Photon Radiosurgery System for interstitial treatment.

absorbing material for the determination of HVL although Cu is more frequently used. The beam quality is expressed in mm Al or mm Cu. For superficial and contact X-ray beams, only mm Al is used to express HVL.

The HVL is measured using an ionization chamber with a narrow beam setup, in which a collimating aperture is placed half-way between the X-ray target and the measurement chamber. The field size is reduced just enough to encompass the chamber. The absorbing materials are placed at least 50 cm from the measurement chamber and close to the collimating aperture. The absorber should be made of high purity material and the thickness of the absorber should be measured accurately. The thickness that reduces the air kerma value to one-half is obtained by interpolation. A monitor detector is used to correct fluctuations of air kerma rate. The measurement chamber must have limited beam quality dependence (within 5% between 40–300 kV) for accurate HVL measurements. Thin-walled chambers are used for lightly filtered beams to measure the low energy X-ray components accurately (HVL errors of up to 10% were observed for a lightly filtered 100 kV beam using a Farmer-type cylindrical chamber).

DOSIMETRY CALIBRATION

Methodology

For kilovoltage X-ray beams, the absorbed dose to water is usually determined with an ionization chamber calibrated in air in terms of either exposure or air kerma. The commonly used ionization chambers are generally considered to be “photon detectors” as the well-known Bragg-Gray cavity theory no longer applies to this energy range (10). To determine the dose to water, the measured air kerma is converted to water kerma using the ratio of mass energy absorption coefficients for water to air, evaluated using the energy spectrum at the position of interest. The dose to water is considered to be the same as water kerma for this energy range due to the negligible difference between kerma and collision kerma (electrons have very small ranges) assuming that quasicharged particle equi-

librium exists. This is generally true for measurement points at depths equal to or greater than the depth of maximum dose for kilovoltage X rays.

Two different methods have been recommended for kilovoltage X-ray dosimetry calibration by national and international dosimetry protocols (11–14). The in-air method measures the air kerma free in air and then converts it to dose to water with the ratio of mass energy absorption coefficients for water to air and a backscatter factor. The in-phantom method measures the air kerma at the reference depth in water under reference conditions and then converts it to the absorbed dose at the depth of the center of the chamber in undisturbed water using the ratio of mass energy absorption coefficients for water to air and other beam quality- and chamber-dependent correction factors.

The ICRU Report 23 was the first to recommend the in-phantom method for orthovoltage X-ray beams since it was difficult to make accurate measurements in regions at or close to the surface of a phantom, and the dose distribution there, unlike at greater depths, was considerably affected by the details of the beam collimation system (11). For this reason, the *British Journal of Radiology Supplement 11* gave two distinct sets of depth-dose tables for close-ended applicators and open diaphragms, respectively (15). It was also clear that by normalizing the depth-dose curves at a depth rather than at the surface the differences in the recommended depth-dose curves would be virtually eliminated. The dose values at greater depths were more clinically relevant since orthovoltage X-ray beams were primarily used for treating deep-seated tumors in the 1970s.

Several more recent dosimetry protocols still recommended the in-phantom method for reference dosimetry for orthovoltage X-ray beams, but used a reference depth of 2 rather than 5 cm (12–14). This was aimed at reducing the measurement uncertainty at lower X-ray energies. However, the in-air method has been the commonly used method for the whole kilovoltage energy range in clinical radiotherapy, especially in North America (16). This may be explained by the fact that nowadays orthovoltage beams

are mainly used for treating tumors close to the skin. The primary point of interest is the dose near the surface rather than at greater depths. Another reason is that it is more convenient to perform routine beam calibration free in air than in a water phantom.

The American Association of Physicists in Medicine (AAPM) recommended that both methods be used for absorbed dose determination for orthovoltage X-ray beams depending on the measurement point of interest (9). To improve measurement accuracy, the in-air method should be used if the primary point of interest is at the phantom surface. On the other hand, if one is more interested in the dose at large depths than at the surface, the in-phantom method should be used. Better agreement in measured PDD curves at depth = 1 cm and greater can be achieved when they are normalized to the values at the 2 cm reference depth than normalized to the surface values. The AAPM recommended that only the in-air method be used for superficial X-ray beam dosimetry (9).

Measurement Equipment

Phantoms. Water is the recommended phantom material when the in-phantom method is used in orthovoltage X-ray beam dosimetry (9). Conversion factors are used to convert from the air kerma in the sensitive chamber cavity to the dose to water. Plastic phantoms are not recommended for in-phantom reference dosimetry for orthovoltage X rays as the chamber correction factors and conversion factors to derive dose at a depth in water for these phantoms are poorly known. Some water equivalent plastics are commercially available for orthovoltage X-ray measurements. However, their properties must be investigated before they can be used as water substitutes.

Dosimeters. Air-filled ionization chambers are recommended for kilovoltage X-ray beam reference dosimetry (9). Measurements for orthovoltage X rays are performed either free in air (in cases where the surface dose is the primary concern) or at a 2 cm depth in water (dose at greater depths is the primary concern). Cylindrical chambers that have a calibration factor varying with the beam quality by <3% in this energy range are recommended for the beam output measurement. For superficial X rays, only the in-air method is used. Cylindrical chambers that have a calibration factor varying with the beam quality by <5% between 50 and 150 kV are recommended. Extensive studies have been carried out on the correction factors for the commonly used Farmer-type chambers for the in-phantom measurement (16–21). Other cylindrical chambers may also be used. However, correction factors must then be determined by comparing these chambers with a chamber with known correction factors. If measurements are performed in water, a thin waterproofing sleeve should be used and appropriate correction factors should be applied depending on the material and thickness used. If measurements are made in air, the thimbles of cylindrical chambers are thick enough, so no build-up cap is required. For X-ray energies <70 kV, calibrated parallel-plate chambers with a thin entrance window are recommended. Thin plastic build-up foils should be added to the entrance window, if

Table 1. Thickness of Build-Up Material for Thin Window Parallel-Plate Chambers Used for In-Air Calibrations with X-Ray Energy <100 kV

Generating Potential, kV	Thickness of Foil, mg cm ⁻²
50	1.5
60	3.0
70	4.7
80	6.6
90	8.7
100	10.9

necessary, to provide full electron buildup and to eliminate electron contamination (see Table 1). All measurements should be corrected for temperature, pressure, ion recombination, and polarity effect.

Electrometers. Electrometers for kilovoltage X-ray therapy measurements should be capable of reading currents on the order of 0.1 nA, with an accumulated charge of 50–100 nC. They are calibrated by a standards laboratory with proper correction factors applied in the measurement. These correction factors are generally close to 1.000; but can occasionally be 5% different from 1.000. Electrometers and ionization chambers can be calibrated either together or sometimes separately and if so their calibration factors must be combined.

Detector Calibration

Modern dosimetry protocols for kilovoltage X rays are based on the air kerma, K_{air} , which is defined as the kinetic energy, ΔE , transferred from X rays to charged particles per unit air mass, Δm (i.e., $K_{\text{air}} = \Delta E/\Delta m$). Assuming that K_{air} is the air kerma at the reference point in air for a given beam quality and M the reading (corrected for temperature, pressure, recombination, and polarity effect) of an ionization chamber to be calibrated with its center at the same point, the air kerma calibration factor, N_K , for this chamber at the specified beam quality is defined as

$$N_K = \frac{K_{\text{air}}}{M} \quad (1)$$

Previous dosimetry protocols for kilovoltage X rays were based on the exposure, X , which is defined as the total electric charge, ΔQ , of all ions of one sign (e.g., electrons), produced in air by X rays in a unit mass of air, Δm (i.e., $X = \Delta Q/\Delta m$). A similar equation can be used to derive the exposure calibration factor, N_X :

$$N_X = \frac{X}{M} \quad (2)$$

The relation between the air kerma calibration factor and the previous exposure calibration factor is given by

$$N_K = N_X \left(\frac{W}{e} \right)_{\text{air}} (1 - g)^{-1} \quad (3)$$

where $(W/e)_{\text{air}}$ has the value 33.97 J/C (or 0.876 cGy·R⁻¹) for dry air, $(1 - g)$ corrects for the effect of radiative losses by charged particles to bremsstrahlung photons (g is the

fractional energy lost to bremsstrahlung photons in air, which is practically zero for X-ray beams below 300 kV).

The N_K or N_X factors can be obtained from a standards laboratory for a number of available X-ray beam qualities to match the beam energies in a clinic. In the calculation of dose to a patient for a kilovoltage X-ray beam, the accuracy of the chamber calibration determines the final accuracy. Each chamber involved in reference dosimetry of the clinical kilovoltage setup should be calibrated, since there can be up to 8% differences between the calibration factor, N_K , for different chambers of the same type. In addition, each chamber should be calibrated at a number of beam qualities to determine the energy dependence of its response. One generally expects this energy dependence to be at most 6% between 40 and 300 kV for well-designed chambers.

Determination of Absorbed Dose

Formalism for the In-Air Method. The in-air method can be used for both superficial and orthovoltage X-ray dose determination. If the point of interest is at the phantom surface the in-air method is more accurate than the in-phantom method. In this method, the measurement is performed with the chamber free in air and the dose to water at the phantom surface (theoretically, this only applies for depths beyond the range of the contaminant electrons and where quasicharged particle equilibrium has been established) can be calculated by

$$D_w = MN_K \left(\frac{\bar{\mu}_{en}}{\rho} \right)_{air}^w B_w P_{stem,air} \tag{4}$$

where M is the corrected ionization chamber reading, N_K the chamber air kerma calibration factor at the user’s beam quality, $(\bar{\mu}_{en}/\rho)_{air}^w$ the ratio of mass energy absorption coefficients for water to air, evaluated over the X-ray energy spectrum free-in-air, in the absence of a phantom, B_w the backscatter factor that accounts for the effect of the phantom (water) scatter, and $P_{stem,air}$ the chamber stem correction factor that accounts for the change in photon scatter from the chamber stem between the calibration and measurement (mainly due to the change in field size). Here, $(\bar{\mu}_{en}/\rho)_{air}^w$ is independent of the field size used since it is evaluated over the primary beam only (22). The AAPM recommended values of $(\bar{\mu}_{en}/\rho)_{air}^w$ and B_w for some typical beam qualities are given in Tables 2 and 3 (taken from Ref. (9)). The parameter $P_{stem,air}$ is taken as unity for a calibrated chamber if, for a given beam quality, the same field size is used in the calibration and the measurement. Otherwise, the $P_{stem,air}$ factor can be measured by inter-comparing the chamber with unknown $P_{stem,air}$ with a reference chamber for which $P_{stem,air}$ is known (9). A Farmer-type cylindrical chamber with flat response can be used as a reference chamber for the measurement of $P_{stem,air}$ of another chamber since its stem effect varies little with field size (<1%).

Formalism for the In-Phantom Method. For orthovoltage X rays, the in-phantom method is recommended (9) if the point of interest is at a depth in the patient. For the 2 cm reference depth in water the dose to water for orthovoltage

Table 2. Ratios of Average Mass Energy Absorption Coefficients for Water to Air to Convert Air Kerma to Water Kerma, Free-in-air for Both Superficial and Orthovoltage X Rays

HVL		$(\bar{\mu}_{en}/\rho)_{air}^w$ (free-in-air)
mm Al	mm Cu	
0.03		1.047
0.05		1.046
0.08		1.044
0.10		1.044
0.3		1.035
0.5		1.028
0.8		1.022
1.0		1.020
3.0		1.021
5.0		1.029
8.0		1.045
	0.1	1.020
	0.3	1.035
	0.5	1.050
	0.8	1.068
	1.0	1.076
	3.0	1.100
	5.0	1.109

X rays can be calculated using

$$D_w = MN_K P_{Q, cham} P_{sheath} \left(\frac{\bar{\mu}_{en}}{\rho} \right)_{air}^w \tag{5}$$

where M is the corrected ionization chamber reading, N_K the chamber air kerma calibration factor at the user’s beam quality, and $(\bar{\mu}_{en}/\rho)_{air}^w$ the ratio of mass energy absorption coefficients for water to air, evaluated over the X-ray energy fluence at 2 cm depth in water, in the absence of the chamber. The overall chamber correction factor $P_{Q, cham}$ accounts for the effect of the change in chamber response due to photon energy and angular variation between chamber calibration in air and measurement in water, the effect of chamber stem between calibration in air and measurement in water and the effect of displacement of water by the chamber in the measurement in water. The sheath correction factor P_{sheath} is needed when a waterproofing sheath is used (23), which is not directly related to the individual chamber type. The values of $P_{Q, cham}$, P_{sheath} and $(\bar{\mu}_{en}/\rho)_{air}^w$ for some typical beam qualities are given in Tables 4–6. The values were taken from Ref. 9.

CLINICAL DOSIMETRY

Percentage Depth Dose

Percentage depth dose (PDD) and lateral dose profiles are difficult to measure for kilovoltage X rays because of the significant variation of X-ray energy (especially for lightly filtered beams) and angular distribution with depth and field size. Detectors for the PDD and dose profile measurement should have high spatial resolution and constant energy and angular response. Solid detectors, attractive because of the small size of their sensitive volume (diode detector, TLD, film), usually show significant beam quality

Table 3a. Water Kerma-Based Backscatter Factors, B_w , for Different Field Diameters (d) and SSD for Superficial X Rays

SSD (cm)	d (cm)	HVL, mm Al										
		0.03	0.05	0.08	0.1	0.3	0.5	0.8	1.0	3.0	5.0	8.0
20	1	1.005	1.007	1.011	1.014	1.028	1.036	1.043	1.046	1.061	1.058	1.053
	3	1.005	1.008	1.014	1.019	1.049	1.069	1.092	1.105	1.158	1.165	1.158
	5	1.005	1.008	1.014	1.019	1.054	1.080	1.112	1.131	1.215	1.234	1.236
	10	1.005	1.008	1.014	1.019	1.057	1.088	1.129	1.155	1.291	1.334	1.354
	15	1.006	1.008	1.014	1.019	1.058	1.090	1.133	1.162	1.321	1.380	1.414
	20	1.006	1.008	1.014	1.019	1.058	1.091	1.136	1.165	1.334	1.402	1.444
30	1	1.005	1.007	1.011	1.015	1.027	1.035	1.043	1.047	1.063	1.059	1.053
	3	1.005	1.008	1.014	1.019	1.048	1.069	1.093	1.107	1.164	1.168	1.158
	5	1.005	1.008	1.014	1.019	1.053	1.079	1.111	1.130	1.221	1.242	1.237
	10	1.006	1.008	1.014	1.019	1.057	1.088	1.130	1.157	1.298	1.350	1.367
	15	1.006	1.008	1.014	1.019	1.058	1.091	1.136	1.165	1.332	1.403	1.434
	20	1.006	1.008	1.014	1.019	1.058	1.091	1.138	1.169	1.350	1.428	1.472
50	1	1.005	1.007	1.011	1.014	1.027	1.035	1.042	1.045	1.065	1.059	1.052
	3	1.005	1.007	1.013	1.018	1.049	1.070	1.093	1.106	1.163	1.169	1.160
	5	1.005	1.007	1.013	1.018	1.054	1.081	1.113	1.132	1.226	1.241	1.242
	10	1.006	1.007	1.013	1.018	1.057	1.091	1.134	1.159	1.309	1.352	1.375
	15	1.006	1.007	1.013	1.018	1.058	1.093	1.140	1.169	1.346	1.411	1.448
	20	1.006	1.007	1.013	1.018	1.058	1.094	1.142	1.173	1.363	1.443	1.493
100	1	1.005	1.007	1.011	1.014	1.028	1.036	1.042	1.044	1.062	1.059	1.053
	3	1.005	1.008	1.014	1.019	1.050	1.070	1.092	1.104	1.163	1.169	1.162
	5	1.006	1.008	1.014	1.019	1.055	1.082	1.113	1.131	1.225	1.240	1.243
	10	1.006	1.008	1.014	1.019	1.058	1.091	1.134	1.158	1.311	1.351	1.381
	15	1.006	1.008	1.014	1.019	1.059	1.094	1.140	1.169	1.354	1.417	1.460
	20	1.006	1.008	1.014	1.019	1.059	1.095	1.143	1.172	1.375	1.451	1.508

dependence or large experimental uncertainties. Well-designed cylindrical chambers have nearly constant energy response in this beam quality range and are suitable for in-phantom measurements. However, the mea-

surement depth is limited to no less than the outer radius of the chamber. Parallel-plate chambers have been used for measurements at smaller depths. Those chambers designed for electron beams usually have a calibration

Table 3b. Water Kerma-Based Backscatter Factors, B_w , for Different Field Diameters (d) and SSD for Orthovoltage X rays

SSD (cm)	d (cm)	HVL, mmCu							
		0.1	0.3	0.5	0.8	1.0	3.0	5.0	
20	1	1.061	1.055	1.053	1.048	1.045	1.024	1.018	
	3	1.158	1.168	1.155	1.147	1.140	1.082	1.057	
	5	1.214	1.242	1.233	1.219	1.209	1.127	1.088	
	10	1.290	1.352	1.353	1.339	1.326	1.204	1.141	
	15	1.320	1.407	1.415	1.403	1.389	1.251	1.174	
	20	1.333	1.434	1.447	1.436	1.421	1.278	1.194	
30	1	1.063	1.056	1.052	1.047	1.044	1.024	1.018	
	3	1.164	1.169	1.155	1.146	1.139	1.084	1.055	
	5	1.220	1.242	1.235	1.221	1.211	1.130	1.087	
	10	1.297	1.363	1.367	1.347	1.332	1.214	1.147	
	15	1.330	1.417	1.438	1.422	1.405	1.270	1.189	
	20	1.348	1.446	1.478	1.464	1.446	1.302	1.213	
50	1	1.065	1.054	1.052	1.047	1.045	1.025	1.018	
	3	1.163	1.170	1.157	1.148	1.140	1.084	1.057	
	5	1.225	1.247	1.240	1.226	1.214	1.131	1.089	
	10	1.308	1.367	1.376	1.360	1.344	1.222	1.152	
	15	1.345	1.433	1.452	1.446	1.428	1.285	1.195	
	20	1.361	1.471	1.499	1.495	1.478	1.325	1.226	
100	1	1.062	1.055	1.052	1.047	1.045	1.025	1.018	
	3	1.163	1.170	1.160	1.150	1.142	1.085	1.057	
	5	1.224	1.245	1.241	1.227	1.217	1.132	1.090	
	10	1.310	1.370	1.383	1.369	1.353	1.226	1.155	
	15	1.353	1.447	1.463	1.458	1.441	1.291	1.204	
	20	1.373	1.490	1.513	1.516	1.499	1.334	1.237	

Table 4. Ratio of Average Mass Energy Absorption Coefficients for Water to Air at 2 cm Depth in Water for a 100 cm² Field Defined at 50 cm SSD for Orthovoltage X rays

HVL		$(\bar{\mu}_{en}/\rho)_{air}^w$
mm Cu	mm Al	
0.1	2.9	1.026
0.3	6.3	1.037
0.5	8.5	1.046
0.8	10.8	1.055
1.0	12.0	1.060
2.0	15.8	1.081
3.0	17.9	1.094
4.0	19.3	1.101
5.0	20.3	1.105

factor varying with beam quality by 20–40% for kilovoltage X rays. Significant corrections with depth may be required for the PDD measurement with these chambers. Specifically designed parallel-plate chambers for low energy X rays usually have a flat energy response in air but not at a depth in a phantom. For example, >10% variations in chamber response have been observed for the Capintec PS-033 chambers. Thus, a depth-related correction factor may be required for these chambers to be used in accurate PDD measurements.

The suitability of X-ray detectors for relative dosimetry measurement has been evaluated extensively (16,24,25). As a general requirement for the evaluation of a specific detector, the relative in-air chamber response and the relative in-phantom response should be compared with a well-behaved cylindrical chamber at depths, where reasonable measurements with the cylindrical chamber can be performed. Diamond detectors and the NACP parallel plate chamber (type) have been found to require relatively small depth-dependent correction factors for orthovoltage X-ray beams (16).

Care must be exercised in the measurement of PDD and dose profiles for kilovoltage X-ray beams because of their significant depth and field size dependence. A water tank with a small-volume scanning ionization chamber is ideal for the PDD and profile measurement. A monitor chamber is often needed to eliminate the effect of erratic fluctuations in dose rate. If a thin window parallel plate chamber is used, the chamber must have sufficient buildup material placed over its entrance window (see Table 1). Because of the finite size of the ionization chamber in the beam direction, it is necessary to extrapolate the dose to the surface. Since the dose distribution may be nonlinear near the phantom surface, caution should be exercised in extrapolating over the last few millimeters (24,25).

If desired, the depth-dose distribution can be measured by placing thin sheets of water-equivalent material over

Table 5. Overall Chamber Correction Factors $P_{Q, cham}$ for Commonly used Cylindrical Chambers in Orthovoltage X-Ray Beams^a

Chamber Type HVL, mm Cu	NE2571	Capintec PR06C	PTW N300 01	NE2611 or NE2561
0.10	1.008	0.992	1.004	0.995
0.30	1.023	1.008	1.021	1.017
0.50	1.025	1.010	1.023	1.019
0.80	1.024	1.010	1.022	1.018
1.0	1.023	1.010	1.021	1.017
2.0	1.016	1.007	1.015	1.011
3.0	1.009	1.005	1.010	1.006
4.0	1.004	1.003	1.006	1.003
5.0	1.002	1.001	1.002	1.001

^aThe data applies to the in-phantom method for a chamber at 2 cm depth in water and a 100 cm² field.

Table 6. Correction Factors for PMMA Sheaths When Using Cylindrical Chambers for In-Water Measurements in Orthovoltage X-Ray Beams^a

HVL		PMMA (Lucite)			
(mm Cu)	(mm Al)	$t = 0.5$ mm	$t = 1$ mm	$t = 2$ mm	$t = 3$ mm
0.1	3.0	0.998	0.995	0.991	0.986
0.3	6.1	0.998	0.997	0.994	0.991
0.5	8.5	0.999	0.998	0.996	0.994
0.8	11.0	0.999	0.998	0.997	0.996
1.0	12.1	1.000	0.999	0.998	0.997
2.0	15.2	1.000	1.000	1.000	0.999
3.0	17.6	1.000	1.000	1.000	1.000
4.0	19.4	1.000	1.000	1.000	1.000
5.0	20.9	1.000	1.000	1.000	1.000

^aThe data applies to 2 cm depth in water and a 100 cm² field.

the chamber in a phantom and moving the chamber back by the same amount to maintain a constant SSD. The water-equivalence of the material in the energy range of interest must be verified. The Poly (methyl methacrylate) (PMMA) is not suitable for this purpose. Strictly speaking, an ion chamber measures the depth-ionization distribution rather than the depth-dose distribution. However, the difference between them is small (16). If a suitable detector for relative dosimetry cannot be identified in the clinic it is recommended to use the data from the *British Journal of Radiology Supplement 25* (26) or published data that match the kV and HVL values of the user's beams (27–31).

Output Factors

Output factors are required for all combinations of SSD and field size used for kilovoltage X-ray treatment. The output factor is defined as the surface dose value for a given SSD and field size relative to that under the reference conditions. Since the scatter contribution from the inside of a cone applicator may be significant, it is not sufficiently accurate to estimate output factors for different applicators using the ratio of the backscatter factors corresponding to the respective field sizes. The output factor for each individual applicator must be measured at each beam quality. If the in-phantom calibration method has been used for orthovoltage X rays it is necessary to obtain the PDD in order to determine the dose at the surface. Note that this may result in large uncertainties in the output factors since the uncertainties are generally high in the PDD values near the surface.

Isodose Curves

Isodose curves are obtained by joining together the points, which have the same dose values. This can be done using the depth dose and lateral profile data if they already exist. A water tank with a small volume ionization chamber is recommended for the dose distribution measurement. A monitor chamber is needed to eliminate the effect of dose rate fluctuations. Dose values at various depths across the beam are measured and the points of equal dose are connected to give the isodose curves. Published dose distributions may be used as reference if a suitable detector for relative dosimetry cannot be identified in the clinic. However, the published data must have the same kV and HVL values as those of the clinical beams.

Quality Assurance

A well-established quality assurance (QA) program will ensure the safe and accurate delivery of radiation therapy treatments using kilovoltage X rays. The safe delivery of radiation therapy is ensured by machine interlocks and strategically placed emergency-off buttons. Accurate delivery of radiation treatment is ensured by maintaining the mechanical accuracy within the specification of treatment unit, maintaining the beam quality in its original condition, and maintaining the accuracy of dosimetry calculation and measurements for treatment planning and plan verification. All dosimetry data and mechanical

limits are established during the initial machine commissioning period before the system is first used for treating a patient and thereafter annually, or after any change which may significantly alter the dosimetry data and the mechanical limits. Once the baseline is established, maintaining the baseline data becomes the mission of the dosimetry QA program. Documentation for each full yearly calibration is maintained for 5 years after the completion of calibration. Documentation of weekly and monthly spot check measurements is maintained for 2 years.

The test frequency of each mechanical component and dosimetry data for a treatment machine is determined by the significance of the tests that indicate beam dosimetry changes and mechanical tolerance changes within the limited amount of time. Unlike clinical linear accelerators, the number of combinations of energy and filter for a kilovoltage X-ray machine are high, but only a few combinations are used on a routine basis. The number of patients that are treated using kilovoltage X-ray beams are much less than for megavoltage radiotherapy. Therefore, the check of the beam dosimetry should depend on the frequency of the beam usage. The dosimetry QA items and their frequencies are summarized below:

Daily Checks.

- Beam output constancy for energy and filter combinations in use.
- Functionality of the audiovisual monitor.
- Door and energy interlock circuits and emergency stops.

Monthly Checks.

- Items included in the daily checks.
- Beam flatness and symmetry.
- Timer operation.
- Light-radiation congruence.

Annual Checks.

- Items included in the monthly checks.
- Dose rate for all energy and filter combinations.
- Output factors for each of the applicator (cone).
- Timer accuracy (verification of the timer error).
- Accuracy of the light localizer system.
- Accuracy of distance measuring devices.
- Beam quality.
- Accuracy of depth dose data and isodose charts.
- Accuracy of field size dependence data.
- Agreement of dose rate with distance from target.
- Attenuation in lead for patient block thickness.

BIBLIOGRAPHY

1. Glasser O, editor. *The Science of Radiology*. London: Bailliere; 1933.

2. Voltz F. Dosage Tables for Deep-Therapy. London: Heine-
mann; 1922.
3. Paterson R. The Treatment of Malignant Disease by Radium
and X-Rays. Baltimore: Williams and Wilkins; 1949.
4. Philips Laboratories, X ray Research. Holland: PL; 1937-
1950.
5. Van Dyk J, editor. The Modern Technology of Radiation
Therapy. Madison, WI: Medical Physics Publishing; 1999.
6. Jaundrell-Thompson F, Ashworth WJ. X ray Physics and
Equipment. Philadelphia: Davis; 1965.
7. Moran EF. Roentgen rays: Generators. In: Glasser O, editor.
Medical Physics. Vol. 1, Chicago, IL: Yearbook Publishers;
1944.
8. Dinsmore M, et al. A new miniature X-ray source for inter-
stitial radiosurgery: Device description. *Med Phys* 1996;23:45-
52.
9. Ma C-M, et al. AAPM protocol for 40-300kV x-ray beam
dosimetry for radiotherapy and radiobiology. *Med Phys*
2001;28:868-893.
10. Ma C-M, Nahum AE. Bragg-Gray theory and ion chamber
dosimetry in photon beams. *Phys Med Biol* 1991;36:413-428.
11. ICRU (International Commission on Radiation Units and
Measurements). Radiation Dosimetry: Measurement of
Absorbed Dose in a Phantom Irradiated by a Single Beam of X- or Gamma Rays. ICRU Report 23. Washington (D.C.):
ICRU; 1973.
12. IAEA (International Atomic Energy Agency). Absorbed Dose
Determination in Photon and Electron Beams; An Interna-
tional Code of Practice, Vol. 277 of Technical Report Series,
Vienna, Austria: IAEA, 1987.
13. Klevenhagen SC, et al. The IPEMB code of practice for the
determination of absorbed dose for x-rays below 300 kV
generating potential (0.035 mm Al-4mm Cu HVL; 10-300
kV generating potential). *Phys Med Biol* 1996;41:2605-2625.
14. NCS (Netherlands Commission on Radiation Dosimetry).
Dosimetry of Low and Medium Energy X-rays, a Code of
Practice for Use in Radiotherapy and Radiobiology. NCS
Report 10. Delft, The Netherlands: NCS; 1997.
15. Central axis depth dose data for use in radiotherapy. *Br J
Radiol (Suppl)* 1972; 11.
16. Ma C-M, Seuntjens JP, editors. Kilovoltage X ray Beam
Dosimetry for Radiotherapy. Madison (WI): MPP; 1999.
17. Seuntjens JP, Thierens H, Schneider U. Correction factors for
a cylindrical chamber used in medium energy x-ray beams.
Phys Med Biol 1993;38:805-832.
18. Ma C-M, Nahum AE. Monte Carlo calculated stem effect
corrections for NE2561 and NE2571 chambers in medium-
energy x-ray beams. *Phys Med Biol* 1995;40:63-72.
19. Ma C-M, Nahum AE. Calculations of ion chamber displace-
ment effect corrections for medium-energy x-ray dosimetry.
Phys Med Biol 1995;40:45-62.
20. Seuntjens JP, Verhaegen F. Dependence of overall correction
factor of a cylindrical ionization chamber on field size and depth
in medium energy X ray beams. *Med Phys* 1996;23:1789-1796.
21. Seuntjens J, et al. Determination of absorbed dose to water
with ionisation chambers calibrated in free air for medium
energy X rays. *Phys Med Biol* 1988;33:1171-1185.
22. Nahum AE, Knight RT. Consistent formalism for kV x-ray
dosimetry. Proceedings of the IAEA International Symposi-
um on Measurement Assurance in Dosimetry. Vienna,
Austria: IAEA; 1994. p 451-459.
23. Ma C-M, Seuntjens JP. Correction factors for water-proofing
sleeves in kilovoltage x-ray beams. *Med Phys* 1997;24:1507-
1513.
24. Li XA, Ma C-M, Salhani D. Measurement of percentage depth
dose and lateral beam profile for kilovoltage x-ray therapy
beams. *Phys Med Biol* 1997;42:2561-68.
25. Ma C-M, Li XA, Seuntjens J. Consistency study on kilo-
voltage x-ray beam dosimetry for radiotherapy. *Med Phys*
1998;25: 2376-2384.
26. Central axis depth dose data for use in radiotherapy. *Br J
Radiol (Suppl)* 1996; 25.
27. Podgorsak EB, Gosselin M, Evans MDC. Superficial and
orthovoltage x-ray beam dosimetry. *Med Phys*
1998;25:1206-1211.
28. Gerig L, Soubra M, Salhani D. Beam Characteristics of the
Therapax DXT-300 orthovoltage therapy unit. *Phys Med Biol*
1994;39:1377-1392.
29. Butson MJ, Mathur J, Metcalfe P. Dose characteristics of a
new 300 kVp orthovoltage machine. *Aust Phys Eng Sci Med*
1995;18:133-138.
30. Aukett RJ, Thomas DW, Seaby AW, Gittins JT. Performance
characteristics of the Pantak DXT-300 kilovoltage X ray
treatment machine. *Br J Radiol* 1996;69:726-734.
31. Kurup RG, Glasgow GP. Dosimetry of a kilovoltage radio-
therapy x-ray machine. *Med Dosimetry* 1993;18:179-86.

See also COBALT 60 UNITS FOR RADIOTHERAPY; RADIATION DOSIMETRY FOR ONCOLOGY; RADIOTHERAPY, INTRAOPERATIVE.

X-RAYS: INTERACTION WITH MATTER

ANIL SETHI
Department of Radiation
Oncology Loyola University
Medical Center
Maywood, Illinois

INTRODUCTION

X-ray interaction with matter forms the backbone of medical physics. The basic interaction and its applications can be found in many subfields of medical physics. Some of the more common examples are applications in diagnosis and treatment of cancer, shielding design of accelerators, radiation detectors, biological response to radiation, and radiation protection. A thorough understanding of the interaction process is therefore mandatory to appreciate many exciting phenomena encountered in medical physics. Although the title of this article refers to X-ray interaction with a medium, the physics of the process applies equally well to gamma (γ) rays. Furthermore, in this article, all electromagnetic radiation is referred to as photons regardless of whether their origin is atomic (X rays) or nuclear (γ rays). How a given photon will interact with matter depends on the energy of the photon and not on its birth-place.

When a beam of X rays passes through the medium, it undergoes attenuation or loss of intensity. In other words, some photons are removed from the beam. This attenuation may be due to either absorption or scattering of photons by the medium. In absorption, the energy of the photon is completely transferred to the atoms, whereas in scattering, the X-ray beam undergoes a change in direction that may be accompanied by a change in its energy. In both absorption and scattering, the net result is the transfer of energy to the medium. The amount of energy deposited per unit mass of the medium is called dose (measured in the SI

unit, gray, $1 \text{ Gy} = 100 \text{ cGy}$), which is an important quantity in radiotherapy applications. How much energy will be transferred to the medium depends on the medium composition and the energy of X rays.

X-ray energy transfer to the medium is a two-step process. First, incident photons interact with the medium and release electrons. Next, in traveling through the medium, these electrons lose their energy via excitation or ionization of atoms. Therefore, photons are labeled as “indirectly ionizing” radiation, in contrast to charged particles, such as, electrons that are “directly ionizing.” It is the latter that act as a vehicle for photon energy transfer to the medium. If the electrons have sufficient energy, they may also eject secondary electrons from atoms that form their own tracks known as δ rays.

MODES OF PHOTON INTERACTION

A photon can interact with the medium in any one of several competing but independent processes. The main modes of interaction are as follows:

1. Coherent (or Rayleigh) scattering
2. Photoelectric effect
3. Compton scattering
4. Pair production or triplet production
5. Photodisintegration

Each of the above processes is characterized by the photon interacting with a different subatomic particle in the medium. For example, the photon interaction may be with the entire atom (as in photoelectric effect, coherent scattering), atomic electrons (Compton scattering), atomic nucleus (photodisintegration), electric field of electrons (triplet production), or nuclear field (pair production). As a result, the photon may undergo elastic (coherent) scattering, inelastic (incoherent) scattering, or complete absorption. For a beam of photons emanating from a linear accelerator and having a mixture of energies (polyenergetic beam), all of these interactions may be taking place simultaneously. Which interaction process will dominate and govern the fate of most photons in the beam depends on the energy of photons and the type (atomic number, Z) of attenuating material.

The interaction processes listed above are in an increasing order of importance as the photon energy increases. For example, coherent scattering is the most important interaction at very low photon energies, whereas photodisintegration occurs only at very high photon energies. In the energy domain most common in medical physics (few Kiloelectronvolts to several Megaelectronvolts), the interactions of greatest interest are photoelectric effect, Compton scattering, and pair production. It is these interactions that will be the primary focus in this article.

SOME DEFINITIONS OF INTEREST

Before describing each interaction in further detail, it is useful to define terms that are commonly used in describing a photon beam.

Photon Fluence (Φ)

The photon fluence of an X-ray beam is the number of photons crossing a unit area. If we find that N photons pass through an area A perpendicular to the beam, then the photon fluence through the medium is

$$\Phi = \frac{N}{A}$$

Fluence Rate (Flux Density, ϕ)

The rate at which photons pass through an area is called fluence rate. If N photons pass through an area A in time t , then the fluence rate is

$$\phi = \frac{N}{At}$$

Energy Fluence (Ψ)

If the energy of each photon is $h\nu$ (where ν is the frequency and h is Planck's constant, $h = 4.135 \times 10^{-31} \text{ MeV/s}$), then the energy fluence is

$$\Psi = \frac{EN}{A} = \frac{dE}{dA} = h\nu \frac{dN}{dA}$$

Energy Fluence Rate (ψ)

Also called the intensity of the beam, and it is defined as ψ (or I)

$$I = \frac{EN}{At} = \frac{d\Psi}{dt}$$

Photon Attenuation

Linear Attenuation Coefficient (μ). Suppose that a monoenergetic beam consisting of N_0 photons is incident on an attenuator of thickness x (Fig. 1). The number of photons N that will pass through the attenuator and get registered by the detector may be written as

$$N = N_0 e^{-\mu x} \quad (1)$$

where μ is known as the *linear attenuation coefficient* of the attenuator and represents the fraction of incident photons that interact in a unit thickness of the medium. Therefore, μ represents the probability that a photon will

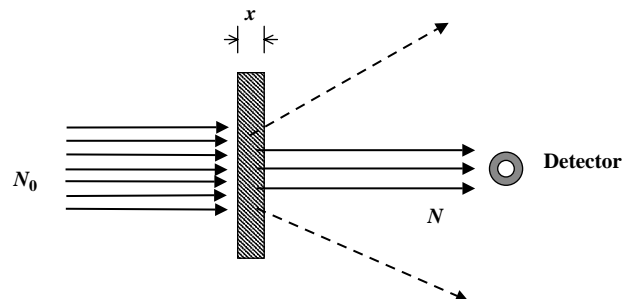


Figure 1. Experimental setup to determine photon beam attenuation in an absorbing medium.

interact in the absorber medium. The attenuation coefficient is a function of both the absorber material and the incident photon energy. As the incident beam is assumed to be monoenergetic, μ is a constant. Because μx must be a dimensionless number, when x is expressed in centimeters, the linear attenuation coefficient μ has units of 1/cm.

Equation 1 may also be written in terms of the beam intensity:

$$I = I_0 e^{-\mu x} \tag{2}$$

where I_0 is the intensity of incident photons and I is the transmitted intensity. The plot of intensity (I) versus thickness of absorber (x) is an exponential curve on a linear graph or a straight line on a semilogarithmic graph (Fig. 2). The plot shows that the percentage of photons removed from the photon beam is the same with each unit increase in absorber thickness.

Half-Value Layer. The thickness of the absorber that will attenuate a photon beam to 50% intensity is known as the half-value layer (HVL). It can be shown from equation 2 that

$$\text{HVL} = \frac{\ln 2}{\mu} = \frac{0.693}{\mu} \tag{3}$$

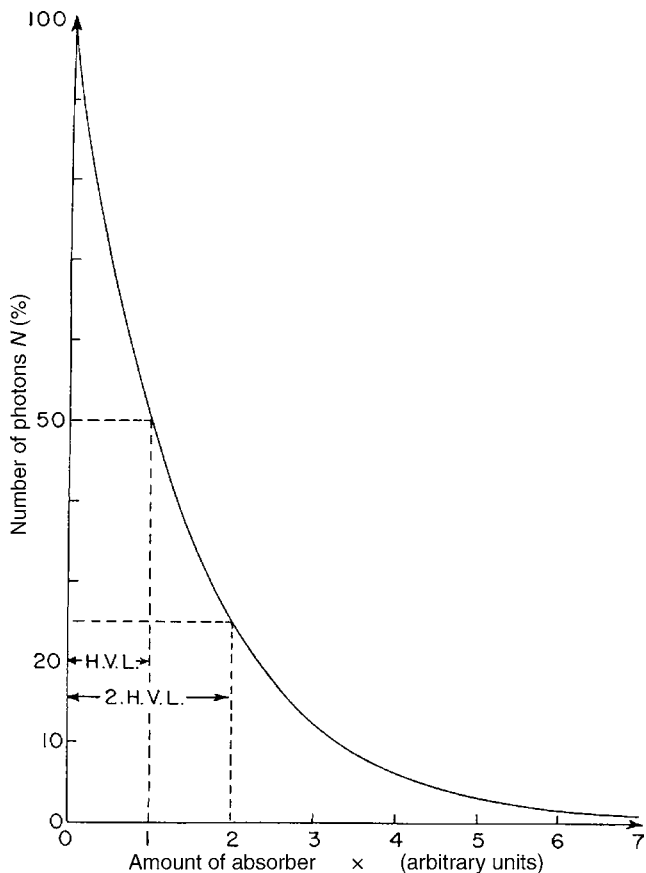


Figure 2. A graph of number (%) of photons transmitted through an absorber versus absorber thickness. Each HVL of absorber reduces the incident number of photons by half.

HVL is a measure of the penetrability of the beam: a high-energy X-ray beam (with low μ) will have a large HVL.

Tenth Value Layer. The tenth value layer (the thickness of the absorber required to reduce the transmitted intensity to one tenth of the original intensity) or TVL can be written as

$$\text{TVL} = \frac{\ln 10}{\mu} = \frac{2.302}{\mu} = 3.323(\text{HVL}) \tag{4}$$

The beam generated from an X-ray tube or a linear accelerator is not mono-energetic, but it is composed of a spectrum of energies. The average energy of such a beam of photons is approximately one third of the maximum photon energy. When this poly-energetic beam passes through the absorber, it is the low-energy X rays that are absorbed or filtered out first. After the low-energy photons have been removed, the filtered beam becomes more energetic or “harder.” As the filter thickness increases, the average energy of the beam increases. Therefore, for an X-ray beam, the plot of beam intensity *versus* absorber thickness is not quite a straight line (Fig. 3). Instead, we see that the first HVL in the absorber is smaller than the second HVL; the second HVL is smaller than the third HVL, and so on.

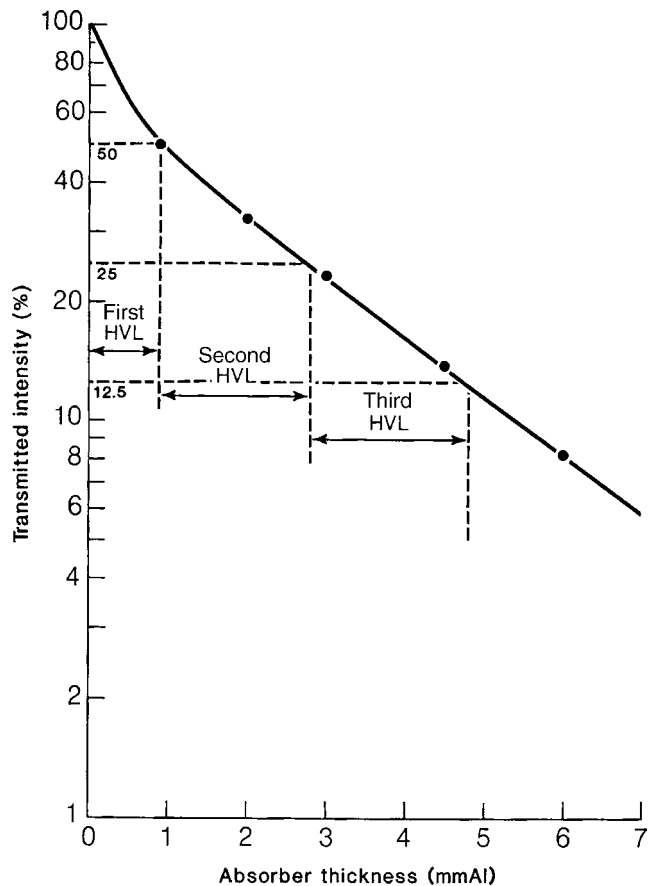


Figure 3. Transmitted intensity for a polyenergetic X-ray beam incident on an absorber. As the beam gets “harder” in passing through the absorber, the corresponding HVL gets larger.

Mean Free Path. As stated above, μ is the probability that a photon will undergo an interaction in a unit thickness of the absorber; therefore, $1/\mu$ can be thought of as the average distance a photon will travel between two successive interactions. Hence, the mean free path (MFP) is

$$\text{MFP} = \frac{1}{\mu} \quad (5)$$

It can be observed from equation 1 that approximately 37% of photons are left in the photon beam after traveling one MFP in the absorber.

Mass Attenuation Coefficient (μ/ρ). The linear attenuation coefficient (μ) varies with the density of the medium. As an example, water, water vapor, and ice have the same atomic composition, but their linear attenuation coefficients are all different due to the difference in physical density. However, if we divide the linear attenuation coefficient μ by the density ρ of the medium, we get a more fundamental quantity, known as the mass attenuation coefficient (measured in squared centimeter/gram):

$$\mu_m = \mu/\rho \quad (6)$$

The absorber thickness, in this case, is ρx and has the units of gram/squared centimeter (i.e., mass of the absorber per unit area). Here the density dependence of the attenuation coefficient has been removed and μ_m is a function of the material composition only. In the aforementioned example, water, water vapor, and ice all have the same mass attenuation coefficient μ_m .

The attenuation coefficient of a mixture of elements (such as water) can be calculated from the attenuation coefficients of individual elements:

$$\left(\frac{\mu}{\rho}\right) = \sum_i w_i \left(\frac{\mu}{\rho}\right)_i \quad (7)$$

where w_i is the fractional weight of the i th element in the mixture.

As photon beam attenuation depends on the number of electrons and atoms in the path of the beam, the corresponding attenuation coefficients can be defined as follows.

Electronic Attenuation Coefficient. When the absorber thickness is expressed as electrons/squared centimeter, the corresponding attenuation coefficient is the electronic attenuation coefficient. If N_e is the number of electrons/gram, then

$$\mu_e = \frac{\mu}{\rho} \left(\frac{1}{N_e}\right) = \frac{\mu}{\rho} \left(\frac{A}{N_A Z}\right) \quad (8)$$

where A is the atomic mass number, Z is the atomic number, and N_A is the Avogadro's number = 6.02×10^{23} atoms per atomic weight in grams. The electronic attenuation coefficient, μ_e (expressed in squared centimeter/electron) represents the probability that an incident photon will have an interaction with an electron in the absorber. As $Z/A \cong 0.5$ for all materials (except hydrogen), the number of electrons/gram is a constant for all materials: $N_e = (N_A Z/A) = 3 \times 10^{23}$ (Table 1). As we will see

Table 1. Number of Electrons Per Gram

Material	Density (g/cm ³)	Atomic Number (Z)	Number of Electrons/Gram
Hydrogen	0.0000899	1	6.00×10^{23}
Carbon	2.25	6	3.01×10^{23}
Oxygen	0.001429	8	3.01×10^{23}
Aluminum	2.7	13	2.90×10^{23}
Copper	8.9	29	2.75×10^{23}
Lead	11.3	82	2.38×10^{23}
<i>Effective Z</i>			
Fat	0.916	5.92	3.48×10^{23}
Muscle	1	7.42	3.36×10^{23}
Water	1	7.42	3.34×10^{23}
Air	0.001293	7.64	3.01×10^{23}
Bone	1.85	13.8	3.00×10^{23}

From Johns and Cunningham, *The Physics of Radiology*.

later, the near constancy of N_e across the periodic table has important implications for the Compton effect.

Atomic Attenuation Coefficient. Similarly, the atomic attenuation coefficient may be defined as the probability that the photon will have an interaction with an atom in the absorber. The coefficient is related to linear attenuation coefficient via

$$\mu_a = \frac{\mu}{\rho} \left(\frac{Z}{N_e}\right) = \frac{\mu}{\rho} \left(\frac{A}{N_A}\right) \quad (9)$$

The atomic attenuation coefficient μ_a is expressed in squared centimeter/atom, when the absorber thickness is in the units of atoms/squared centimeter.

We now look at each interaction in more detail.

COHERENT OR RAYLEIGH SCATTERING

Coherent scattering refers to the elastic scattering of a photon beam with atomic electrons, in which no energy transfer to medium takes place. The incident photon interacts with bound electrons as a whole (hence called coherent scattering). The electrons are temporarily set in motion (or oscillation) by the photon's electromagnetic field, and they return back to their original state by emitting a photon with the same energy as the incident photon (Fig. 4). The incident photon energy is too small to break free any electron from its shell. The scattered photon is emitted at a small angle relative to the incident photon. When a single electron from an atom participates in the process, the reaction is called Thomson scattering.

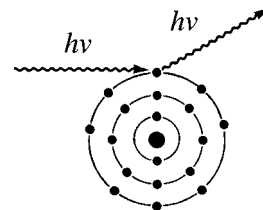


Figure 4. Illustration of coherent scattering: A photon with energy $h\nu$ scatters off an atom without transferring any energy. The scattered photon has the same energy as the incident energy.

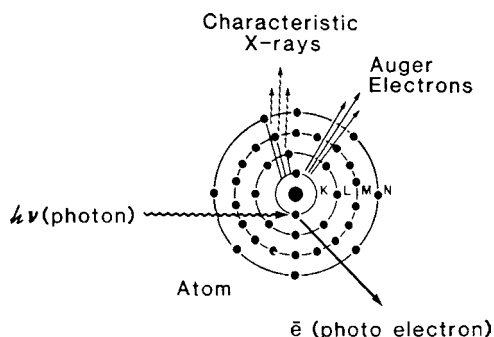


Figure 5. Photoelectric effect: A photon incident on an atom is absorbed with the emission of photoelectron. In the process, characteristic X rays and Auger electrons may also be emitted.

The probability of coherent scattering (σ_{coh}) is highest when photon energy is very small and the medium atomic number Z is very large. Likewise, the probability of coherent scattering is lowest when Z of the medium is small and the photon energy is large. Consequently, coherent scattering is not very likely to occur in tissue and therefore is not of much interest in medical physics.

PHOTOELECTRIC EFFECT

In the photoelectric effect, the incident photon interacts with an atom and ejects one of the bound electrons (from K, L, M, or N shells) (Fig. 5). The photon disappears by transferring all of its energy to the atom. Some of the photon's energy is used to overcome the binding energy of the electron, and the rest changes into the kinetic energy of the electron:

$$K.E. = h\nu - E_b \quad (10)$$

where $h\nu$ is the incident photon energy and E_b is the binding energy of the electron. For the photoelectric effect to take place, the energy of the incident photon *must* be larger than the binding energy of the bound electron.

For example, suppose that the incident photon ejects an electron from the K-shell. After the photoelectron is ejected, a vacancy is created in the electron shell and the atom goes into an excited state. The vacancy left by the electron is filled by an outer shell electron (say, from an L-shell) with a lower binding energy. This leads to emission of characteristic X rays with energy $= E_K - E_L$. If the characteristic X rays have sufficient energy, they may also knock out electrons (called Auger electrons) from the surrounding shells. These Auger electrons leave behind more vacancies that in turn leads to generation of more X rays. The cycle is repeated until all of the photon energy is absorbed by the medium. In the tissue, the binding energies of electron shells are very small. Therefore, the characteristic X rays released in tissue are of very low energy and hence are locally absorbed. In the photoelectric energy range, all incident radiation on tissue is locally absorbed with no scatter radiation. In contrast, the characteristic radiation produced in high Z material (e.g., lead) is more energetic and is absorbed some distance away from the site of photoelectric effect.

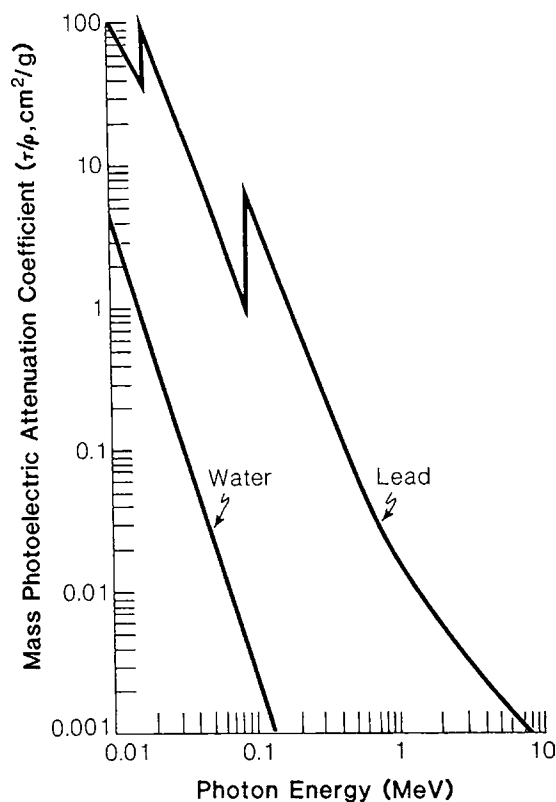


Figure 6. Variation of photoelectric mass attenuation coefficient with incident photon energy. Curves shown are for water ($Z = 7.82$) and lead ($Z = 82$).

The probability that a photon will undergo photoelectric effect depends strongly on its incident energy $E = h\nu$ and the atomic number Z of the absorbing material. In fact, the photoelectric mass attenuation coefficient τ/ρ varies directly as the cube of atomic number and inversely as the cube of photon energy (Z^3/E^3). Figure 6 shows a plot of τ/ρ versus energy for water ($Z = 7.5$) and lead ($Z = 82$). Due to the Z^3 dependence, a photon is 1000 times more likely to undergo photoelectric interaction in lead than in a water-like medium (e.g., tissue). Also, if it has sufficient energy to knock out an electron, a low-energy photon is more likely than a high-energy photon to participate in the photoelectric effect. It can be observed from the figure that as the photon energy increases, the probability of photoelectric interaction decreases rapidly (as $1/E^3$). The curve for lead also shows sharp discontinuities for incident photon energies of 15 keV and 88 keV. These peaks correspond to the binding energies of L- and K-shells, respectively, for lead and are called absorption edges. A photon having energy less than the electron's binding energy cannot undergo a photoelectric effect with electrons in that shell. However, as soon as the photon energy exceeds the binding energy of the electron, the probability of the photoelectric effect increases dramatically (like a resonance) and a sudden jump (or discontinuity) is observed in the plot of photoelectric attenuation coefficient versus energy. The discontinuity is greatest for the K-shell and becomes weaker for higher shells (L, M, N, etc.). Therefore, it is noticed that the most tightly bound

electrons have the greatest chance of undergoing photoelectric interaction. For water, the plot of τ/ρ again shows the same steady decline with energy, although the curve is much smoother. As the K-shell binding energy for a low Z medium is only about 0.5 keV, no discontinuities are observed.

The above behavior of the photoelectric effect has many important applications in medical physics:

1. Due to the Z^3 dependence, differential absorption of photons by bone, muscle, and fat is exaggerated and provides excellent X-ray film contrast in mammography and other diagnostic applications.
2. High Z materials, such as, BaSO_4 and Hypaque, are ideally suited for contrast enhancement in CT scanning.
3. As lead is a good absorber of low-energy photons, it is commonly used for shielding in diagnostic radiology procedures (for example, lead aprons).

The angular dependence of emitted photoelectron is a function of the photon's energy. For a low-energy photon, the electron has small kinetic energy and is emitted at 90° relative to the direction of the incident photon. As the photon energy increases, the photoelectrons also have higher energy and are emitted in a more forward direction.

COMPTON SCATTERING

As the incident photon energy increases (beyond 30–40 keV in tissue-like medium), the probability that it will undergo photoelectric effect decreases and the Compton effect becomes the dominant mode of interaction. This is the most important mode of interaction for tissue like materials.

In Compton scattering, the incident photon with energy $h\nu_0$ interacts with a loosely bound (or “free”) electron from an outer shell and transfers some of its energy to it. The photon is scattered at a lower energy ($h\nu'$) and scattering angle θ , and the recoil (or Compton) electron is ejected with an energy E at an angle ϕ relative to the incident photon's direction (Fig. 7). From conservation of energy and momentum, the following relations

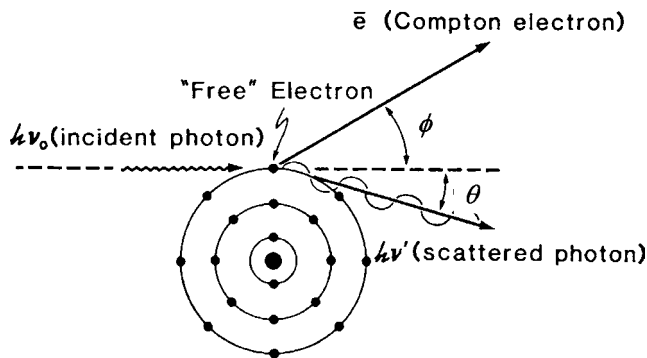


Figure 7. Compton effect: Incident photon scattering off a free electron.

can be obtained:

$$E = h\nu_0 \frac{\alpha(1 - \cos \theta)}{1 + \alpha(1 - \cos \theta)}$$

$$h\nu' = \frac{1}{1 + \alpha(1 - \cos \theta)}$$

$$\cot \phi = (1 + \alpha)\tan \theta/2$$
(11)

where $\alpha = h\nu_0/m_0c^2$ and $m_0c^2 = 0.511$ MeV is the rest mass energy of the electron.

In terms of a photon's wavelength, after undergoing a Compton interaction, the scattered photon has a longer wavelength λ' than that of the incident photon λ . The change in wavelength or “Compton shift” is independent of the incident photon energy and depends only on the scattering angle θ :

$$\Delta\lambda = \lambda' - \lambda = h/m_0c(1 - \cos \theta) = \lambda_C(1 - \cos \theta)$$
(12)

where $\lambda_C = h/m_0c = 0.02426 \text{ \AA} = 2.426 \times 10^{-10} \text{ m}$ is the Compton wavelength, or the wavelength of a photon whose energy is just equal to the rest mass energy of the electron.

The scattered photon's energy, and therefore how much energy is imparted to Compton electron, depends on the scattering angle and incident photon energy. Let us first consider the dependence of transferred energy on the photon's scattering angle θ (Fig. 8). If the photon undergoes a “head-on” collision with the electron, the photon is scattered backward ($\theta = 180$) and the electron moves forward ($\phi = 0$). In this case, the photon transfers most of its energy to the electron ($h\nu' = h\nu'_{\min}$ and $E = E_{\max}$). If the photon makes a “glancing” hit with the electron ($\theta = 0$), the electron receives very little energy and the scattered photon continues forward with the same energy as the incident photon ($h\nu' = h\nu$ and $E = 0$). In summary, when a photon is scattered at small angles ($\theta \rightarrow 0$), very little of its energy is transferred to the electron. However, as the photon scattering angle increases ($\theta = 0 \rightarrow 180$), a greater fraction of the incident energy is imparted to the electron.

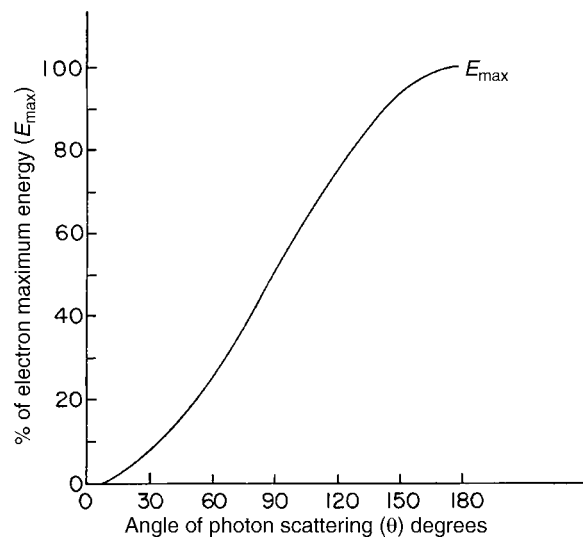


Figure 8. Compton scattering: Relationship between energy transferred to Compton electron and the scattering angle.

Unlike the photoelectric effect, however, the photon does not give all of its energy to the electron. The directional dependence of scattered electron and photon is governed by equation 11.

We next study the dependence of Compton effect on incident photon energy. For a low-energy photon, only a small portion of its energy is imparted to the electron. The scattered photon's energy is almost the same as the incident photon. Note that in the limiting case ($h\nu_0 \rightarrow 0$), this reduces to Rayleigh scattering where the incident photon suffers no loss of energy. As the photon energy increases, the fraction of energy transferred to the electron increases. For a very high-energy photon, the photon loses almost all of its energy to the electron and is emitted with a low energy.

It can be shown that a photon scattered at 90° can have energy of no more than 0.511 MeV regardless of the incident photon energy. If the photon is scattered backward, its maximum energy is only 0.255 MeV. These results are independent of the incident photon energy and have important consequences in the shielding design for treatment rooms in radiotherapy. For example, for side-scattered radiation (90° scatter), one needs to only shield for 0.511 MeV photons, and for backscattered radiation (180° scatter), the maximum required shielding is for 0.255 MeV photons. However, a photon scattered in the forward direction can have any energy up to that of the incident photon. As the incident photon energy increases, the recoil electron angle ϕ becomes smaller; i.e., the electron is more likely to be ejected in the forward direction.

As the Compton interaction involves a free electron, it is independent of the atomic number Z of the medium and depends only on the number of electrons per gram, which is constant for almost all materials (Table 1). Thus, the Compton mass attenuation coefficient σ/ρ is the same for all materials. In other words, gram for gram all materials will undergo the same Compton interaction. However, the linear attenuation coefficient σ will be larger for denser materials: In the Compton energy range, 1 cm of bone will attenuate more than 1 cm of tissue. The inherent contrast in the Compton range is due to density difference and not due to Z dependence as observed in the photoelectric effect. This fact, coupled with the presence of scattered photons, causes the mega-voltage X ray film quality to be inferior to that of kilovoltage films. In the soft tissue, the Compton effect is most important for photons with incident energy 0.1–10 MeV. As the photon energy increases, the probability of Compton interaction decreases.

PAIR PRODUCTION

A photon with energy more than 1.02 MeV may interact with the medium through pair production. In this reaction, the photon interacts with the field of the nucleus and disappears with the creation of a positive and negative electron (e^+/e^-) pair (Fig. 9). This reaction is an example of energy converting into mass. As the rest mass energy for electron or positron is $m_0c^2 = 0.511$ MeV, the threshold for pair production is 1.022 MeV. The total kinetic energy of the electron–positron pair is

$$h\nu - 1.022 = E^+ + E^- \quad (13)$$

The excess energy is generally shared equally between e^+ and e^- ; however, any ratio is possible. As the products have equal and opposite charge, the net charge is conserved in the reaction.

The above reaction can also take place in the presence of an electron: a process known as triplet production (e^+/e^- pair and the interacting electron). The threshold energy for triplet production is 2.04 MeV. However, the likelihood of triplet production is small compared with pair production.

The electron and positron formed in the pair- (or triplet-) production lose their energy in passing through matter via ionization and excitation of atoms, until they come to rest. Near the end of its track, a positron combines with an available electron to produce two annihilation photons, each with an energy of 0.511 MeV (Fig. 9). The photons are emitted in opposite directions (180 degrees apart) to conserve momentum. This reaction is opposite to the pair production in that matter is now converted into energy.

The likelihood of pair production is very small when the photon energy is about 1–2 MeV. However, it increases rapidly with energy and becomes the dominant interaction for photons with an incident energy above 10 MeV. This behavior is in contrast to the other photon interaction processes considered so far that decrease in likelihood with increasing photon energy. The pair production interaction also increases with Z of the atomic nucleus.

It varies as Z^2/atom or Z/g . This implies that high-energy X rays will be readily absorbed in high Z materials, leading to “beam-softening.” For this reason, lead is not recommended as a flattening filter in high-energy linear accelerators. For the same reason, HVL is not a useful concept in specifying beam quality of high-energy X rays. This can be explained by recalling that HVL is related to the attenuation coefficient as $\text{HVL} = 0.693/\mu$. In the high-energy range, μ increases (or HVL decreases) as the photon

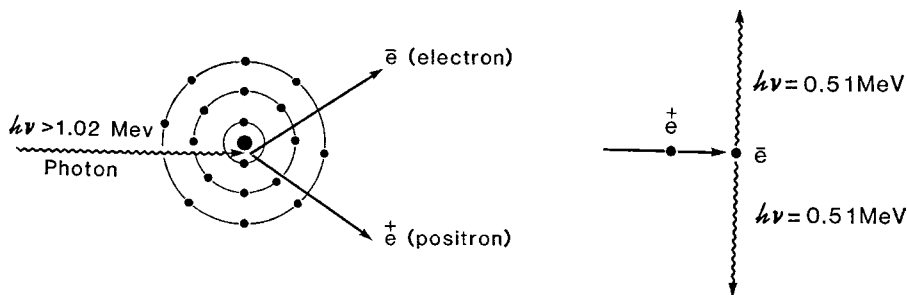


Figure 9. Illustration of pair production process and subsequent production of annihilation radiation.

energy increases. Therefore, HVL is not a meaningful quantity in describing a beam of high-energy photons.

PHOTODISINTEGRATION

A photon with still higher energy may penetrate the nucleus and knock out one of its constituents: a proton, neutron, or alpha particle along with more gamma rays. The incident photon is absorbed, and the nucleus is transformed into an unstable reaction product. The latter next returns back to a stable state via radioactive decay of a nuclear particle, for example, as in (γ, p) and (γ, n) reactions.

The threshold for the photodisintegration reaction is essentially the binding energy of nucleons in the nucleus. For low Z nuclei, this is above 10 MeV (except for Be: 2 MeV and ^2H : 1.5 MeV). For heavy nuclei, the nuclear binding energy is about 7 MeV. Due to Coulomb repulsion, the threshold for (γ, p) is lower than that for the (γ, n) reaction. Beyond the threshold energy, the probability for photodisintegration increases rapidly with increasing incident photon energy, reaches a maximum value, and then drops with further increase in energy. This peak is referred to as the *nuclear giant resonance* and is due to the electric dipole absorption of the incident photon.

Compared with the other reactions described above, the probability for photodisintegration is small and hence does not contribute significantly to photon attenuation. However, the interaction has great importance in shielding considerations for radiotherapy room design.

RELATIVE IMPORTANCE OF VARIOUS TYPES OF INTERACTIONS

Up to now we have considered how a monoenergetic beam of photons interacts with a medium having an atomic number Z . However, an X-ray beam from a radiotherapy machine is not monoenergetic, but it consists of a mixture of photons with various energies. Such a beam, in passing through an absorber, will undergo all of the above interactions to various degrees. The *total linear attenuation coefficient* may be written as the sum of the photoelectric, coherent, Compton, and pair production coefficients:

$$\mu_{\text{total}} = \tau + \sigma_R + \sigma_C + \kappa \quad (14)$$

Figure 10 shows a plot of μ_{total} versus photon energy along with contributions from individual component interactions.

The *total mass attenuation coefficient* may be written as

$$\frac{\mu_{\text{total}}}{\rho} = \frac{\tau}{\rho} + \frac{\sigma_R}{\rho} + \frac{\sigma_C}{\rho} + \frac{\kappa}{\rho} \quad (15)$$

The relative probability of various interactions depends on the incident photon energy and the Z of the material. In general, the photoelectric interaction is most common at low photon energies; the Compton effect dominates at intermediate energies and the pair production at high energies. As explained above, as coherent scattering is significant for low-energy photons (<10 keV) incident on high Z materials, it may be ignored.

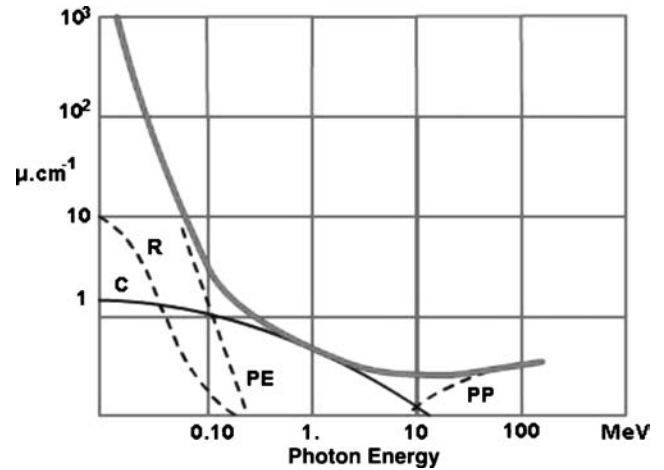


Figure 10. A plot of total linear attenuation coefficient that is composed of Rayleigh, photoelectric, Compton, and pair production processes.

Let us consider the plot of total mass attenuation coefficient as a function of photon energy for water and lead, which, respectively, represent a low Z and high Z material (Fig. 11). At low energies (≥ 10 keV), the photoelectric effect is dominant. The attenuation coefficient displays Z^3/E^3 behavior in this region. Therefore, μ is much higher for lead (10^3 times) compared with water and decreases rapidly with increasing photon energy.

As the photon energy exceeds electron binding energy (~ 100 keV and higher), the Compton effect takes over. As the Compton interaction is independent of Z of medium, lead and water have practically the same attenuation. As the photon energy is further increased, the attenuation coefficient decreases until the pair production becomes the dominant mode of interaction (for photon energies higher than 1 MeV).

In the pair production range, the attenuation coefficient increases with energy as $\log E$. Also, the likelihood of interaction is higher in lead than in water because of the Z dependence of pair production interaction. The above behavior of various modes of interactions is summarized in Table 2 for tissue-like material.

Another way to discuss the relative importance of various interactions is to look for them as regions of dominance in the plot of $h\nu$ versus Z (Fig. 12). The curves display points in the $h\nu$ and Z space for which the Compton effect equals the photoelectric effect and pair production. The Compton effect is most dominant between 1 and 5 MeV

Table 2. Relative Importance of Photoelectric, Compton, and Pair Production Processes in Water

Energy	Photoelectric	Compton	Pair Production
10 keV	95	5	0
25 keV	50	50	0
150 keV	0	100	0
4 MeV	0	95	5
24 MeV	0	50	50
100 MeV	0	16	84

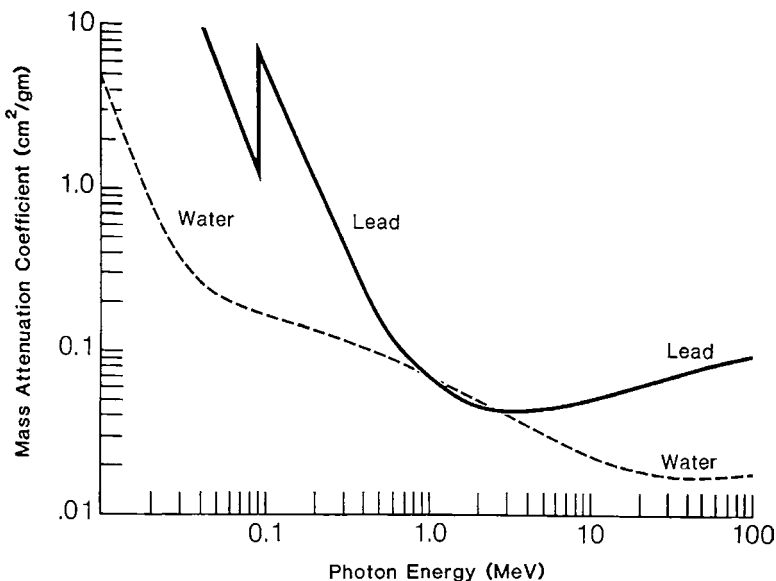


Figure 11. A plot of total mass attenuation coefficient as a function of incident photon energy for lead and water.

regardless of the type of material. According to this figure, for low Z (tissue-like materials), the Compton effect is the main interaction over a much wider energy range. For high Z materials, however, the photoelectric effect is dominant at low photon energies and pair production is the main interaction at high energies.

ENERGY TRANSFER AND ENERGY ABSORPTION COEFFICIENTS

When a beam of photons passes through an attenuator, either part or all of its energy is transferred to the medium (Table 3). The exact fraction of energy transferred depends on the incident photon energy and Z of the medium. If part of the incident energy is transferred, then the remaining photon energy is emitted as scattered photons. The scattered photons may further interact with the medium and lose some or all of their energy. Thus, a photon may undergo *multiple* interactions with electrons in the medium before it is absorbed or escapes out. Suppose $h\nu$ is the

incident photon energy, out of which, \bar{E}_{tr} is the average energy transferred to the medium. Let E_{scat} be the energy of scattered photons; then

$$h\nu = \bar{E}_{tr} + \bar{E}_{scat} \tag{16}$$

The fraction of photon energy transferred to electrons kinetic energy per unit absorber thickness can be written as

$$\mu_{tr} = \frac{\bar{E}_{tr}}{h\nu} \mu \tag{17}$$

where μ_{tr} is the linear energy transfer coefficient. The corresponding mass energy transfer coefficient is μ_{tr}/ρ .

Most of the energy transferred to the electrons is deposited in the medium at the site of interaction via ionization and excitation of atoms and is referred to as the absorbed energy E_{ab} (related to radiation dose delivered). However, a portion of the transferred energy is radiated away in the form of bremsstrahlung radiation (E_{rad}):

$$\bar{E}_{tr} = \bar{E}_{ab} + \bar{E}_{rad} \tag{18}$$

Therefore, the energy absorption coefficient may be written as

$$\mu_{ab} = \frac{\bar{E}_{ab}}{h\nu} \mu \tag{19}$$

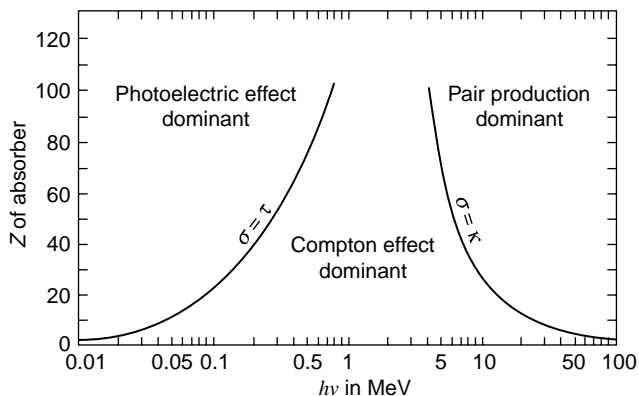


Figure 12. Relative importance of the three major types of X-ray interactions. The curves show the values of Z and $h\nu$ for which two neighboring effects are equal.

Table 3. Depth of Maximum Dose and Percent Depth Dose at 10 cm Depth for Megavoltage Photon Beams

Photon Beam Energy	Depth of Maximum Dose (cm)	Percent Depth Dose at 10 cm depth (%)
1.25 MV (⁶⁰ Co)	0.4	58.7
4 MV	1	63
6 MV	1.6	66.7
10 MV	2.5	73.2
18 MV	3.5	79.2
25 MV	5	84.5

X-RAY BEAM ENERGY PARAMETERS

As we have noted, the type of interaction an X-ray beam will undergo with a medium depends on its energy. This section presents some ways of defining energy of incident radiation.

The most comprehensive description of an X-ray beam is obtained via spectral energy distribution, i.e., a graph showing the relative population of different energy photons in the beam. The spectral distribution can be measured by several methods, including magnetic spectrometry, photoactivation, Cerenkov detection, and total-scintillation spectrometry, or it can be approximated by computational techniques based on bremsstrahlung interaction in a target. Although useful, the exact determination of the spectra is labor intensive. In addition, the spectral distribution consists of a vast amount of information that makes it difficult to compare two X-ray beams using this information. Instead, alternative simpler methods may be used to describe the quality of the X-ray beam and its spectra.

Beam Energy, Maximum Energy, or Peak Energy

Typical X-ray spectrum consists of photons with energy ranging from 0 to a maximum energy (E_{\max}). X-ray energy as nominally specified by the manufacturer is generally the energy of the peak intensity. This energy is the energy of the most probable electrons incident on the patient. The peak energy, however, is not a good indication of the X-ray spectrum, as two beams with same peak energy may have different spectrum.

Effective Energy

The X-ray beam is usually composed of photons of a mixture of energies. These will attenuate differently in matter. The effective energy of such a heterogeneous beam is defined as the energy of a monoenergetic beam that has the same HVL as the beam in question.

Half-Value Layer

HVL is the thickness of the absorber required to attenuate the beam intensity to half its original value.

Weighted Mean Energy

The weighted mean energy of a heterogeneous X ray beam is found from the mean mass-attenuation coefficient weighted by the energy fluence of the photons. The mean energy of the heterogeneous beam is approximately one third of the peak energy.

FURTHER READING

- Khan FM. *The Physics of Radiation Therapy*. 3rd ed. Philadelphia, PA: Lippincott Williams and Wilkins; 2003.
- Johns HE, Cunningham JR. *The Physics of Radiology*. 4th ed. Springfield, IL: Thomas; 1984.
- Evans RD. *The Atomic Nucleus*. Malabar, FL: Krueger; 1955.
- Cember H. *Introduction to Health Physics*. 3rd ed. New York: McGraw-Hill; 1996.

Ter-Pogossian MM. *The Physics Aspects of Diagnostic Radiology*. New York: Harper & Row; 1967.

Hendee WR, Ritenour ER. *Medical Imaging Physics*. 4th ed. New York: Wiley Science; 2002.

See also RADIATION THERAPY, QUALITY ASSURANCE; SCREEN-FILM SYSTEMS; THERMOGRAPHY.

X-RAYS, PRODUCTION OF

BRUCE HORN
Kaiser Permanente
Los Angeles, California

INTRODUCTION

Soon after Wilhelm Roentgen discovered X rays in 1895, scientists recognized their usefulness to visualize the internal anatomy of humans. For a number of years, X-ray tubes that radiologists used for this purpose were unreliable and produced low X-ray output. In 1913, William Coolidge invented the forerunner of the modern X-ray tube. Unlike earlier designs, this tube featured a heated tungsten filament as the source of electrons and utilized a high quality vacuum instead of the low gas pressure previously employed. The result was a reliable X-ray tube whose X-ray output could be reproducibly controlled over a wide range. Today's wide range of X-ray imaging techniques, from a simple chest radiograph to a multislice computed tomogram, still depend on an X-ray tube to produce X rays. Recent developments in X-ray tube design include improved heat handling capability, increased X-ray output, and reduced focal spot size.

PRODUCTION OF X RAYS

X rays, like radio waves, light, γ rays, and other types of electromagnetic radiation, are defined by their energy and source. X rays are produced by two methods involving the interaction of electrons with matter. In the first method, a stream of electrons directed at a target is decelerated by forces between the incident electrons and the atomic nuclei of the target material. Since a deceleration implies that kinetic energy is lost, one of the ways in which electrons lose energy is by creating photons of electromagnetic energy equal in energy to the kinetic loss. This process is called bremsstrahlung, or "braking radiation". The photons produced by this interaction are commonly called X rays. If a large number of electrons, all having the same initial kinetic energy, interact with a target material, the resulting bremsstrahlung consists of photons, or X rays, with a continuum of energies from zero to a maximum equal to the initial electron kinetic energy. This range of energies is called the X-ray spectrum. The shape of the spectrum remains the same regardless of the kinetic energy of the electrons. Although X rays are produced by this mechanism even if the target consists of a gas, the efficiency of X-ray production by bremsstrahlung is directly proportional to the product of the atomic number (Z) of the target

material and the kinetic energy of the electrons. Since the portion of the kinetic energy of the incident electrons that is not converted to X rays results in heating the target material, it is desirable to utilize high energy electrons and target materials of high atomic number to increase the efficiency of bremsstrahlung production as much as possible. However, in either therapeutic or diagnostic medical applications, X-ray beam penetration is controlled by selecting the initial kinetic energy of the electrons, which is controlled by the electrical potential difference between the electron source (cathode) and the target (anode). Although other considerations may dictate the choice of target material when designing X-ray tubes for some diagnostic applications, such as mammography (molybdenum or rhodium), the target material remains as the variable that can be optimized to provide the most efficient X-ray production. Due to its high atomic number, relative high melting point, good heat-transfer characteristics, and low cost, the best target material for use in diagnostic imaging applications is tungsten. At energies commonly employed in medical applications, approximately $\sim 1\%$ of the kinetic energy of the electrons is converted to bremsstrahlung by interaction with a tungsten target. The remaining 99% of the incident energy results in heating of the target. As will be discussed later, the primary challenge of X-ray tube design is how to most effectively handle heat generated by the X-ray production process.

In the second method, electrons impinging on the target also interact with the orbital electrons of the atoms in the target. If the incident electron has a kinetic energy greater than or equal to the binding energy of an orbital electron, the bound electron may be ejected from the atom as a result of the interaction. The resulting vacancy is then filled by one of the other orbital electrons falling into the potential well of the vacancy. This transition results in the release of a photon with an energy equal to the difference between the electronic states of the vacancy and the orbital electron that fills the vacancy. The released photon is called a characteristic X ray. Due to the quantized energy levels of orbital electrons, characteristic X rays have specific energies independent of the original kinetic energy of the accelerated electrons. For tungsten, the *K*-shell binding energy is 69 keV, and the *L*-shell binding energy is 11 keV. Characteristic X rays resulting from interactions with *L*-shell electrons are not important for medical applications since the resulting low energy photons are not useful and are severely attenuated by the encapsulating materials used in a practical X-ray tube. As a result, for practical purposes, the X-ray spectrum from a tungsten target produced by electrons with kinetic energies < 69 keV will not exhibit characteristic X-ray production. X-ray spectra produced by electrons with kinetic energies > 69 keV will have a series of narrow peaks of increased X ray production at photon energies corresponding to the energy differences for the possible *K*-shell orbital transitions superimposed on the bremsstrahlung spectrum. Figure 1 illustrates the typical shape of bremsstrahlung spectra, as well as the difference in characteristic X-ray production between two accelerating potentials. The existence of characteristic X rays in the spectra is relatively unimportant for most medical diagnostic imaging procedures. On the other hand,

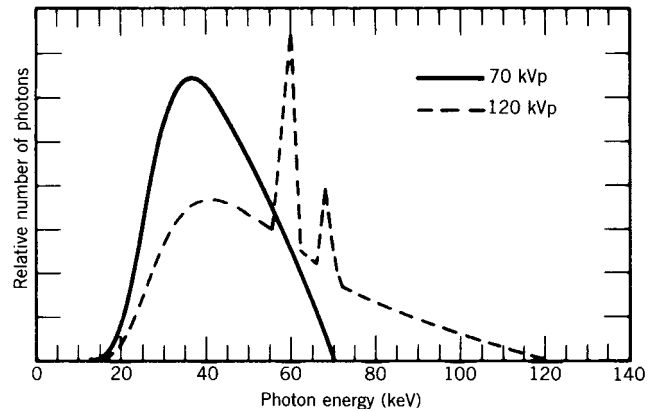


Figure 1. Diagnostic X-ray spectra.

the characteristic X rays produced by a molybdenum target X-ray tube are important to the process of image formation for film-screen mammography.

CONSTRUCTION OF X-RAY TUBES

An X-ray tube consists of a cylindrical envelope (tube insert) that contains the primary components necessary to produce X rays in a controlled manner. This envelope is contained within a tube housing that provides physical protection of the tube insert, stray radiation protection and a means of transferring heat generated during the X-ray production process from the insert to the surrounding environment. The insert contains an electron source (filament), a means of directing the electrons in the intended direction (focusing cup), and a target (anode). A typical X-ray tube insert is shown in Fig. 2. Unlike radiation emitted by radioactive material, X-ray production can be electrically switched on and off by controlling the voltage applied to the X-ray tube.

Envelope

The envelope housing the X-ray generating components is usually constructed of blown glass. However, some specialized X-ray tubes have metal or ceramic envelopes to improve the tube's heat handling capabilities. Air is evacuated from the envelope until an acceptable vacuum is achieved so that the electrons used in the X-ray generation process will not interact with gas molecules.

Filament

A small helical coil of tungsten wire, similar to the filament in an ordinary light bulb, is used to produce the electrons needed for the X-ray generation process. Passing an electric current through this tungsten filament, heats it to the point where free electrons are sufficiently energized to escape from the filament. This process is called thermionic emission. The freed electrons form a cloud, or space charge, around the filament. A separate filament is provided for each focal spot size, usually two per tube. During the X-ray generation process, a large negative electrical potential difference is applied to the filament relative to the anode

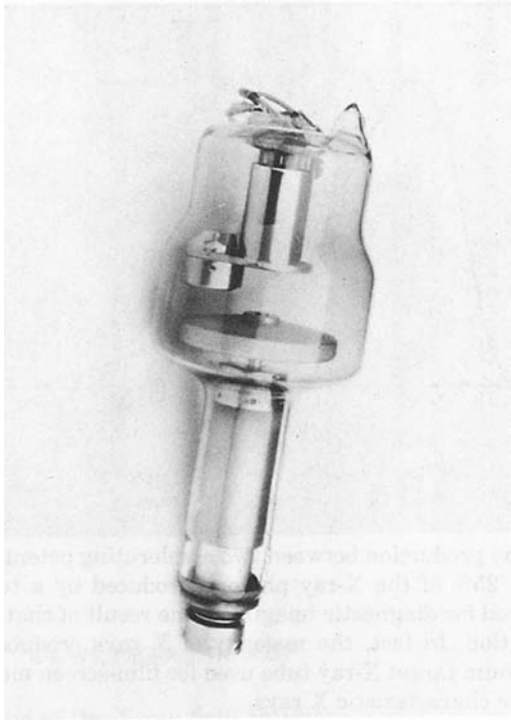


Figure 2. Typical medical X-ray tube insert.

so that electrons supplied by the cloud are repelled by the filament and attracted by the anode. If large quantities of electrons (high X-ray output) are needed at low accelerating potentials, the repulsive force of the space charge may inhibit thermionic emission and reduce the rate at which electrons are released.

Focusing Cup

A metal focusing cup (cathode) surrounds the filament on all sides, except that facing the anode. Figure 3 shows a typical focusing cup and filaments. During the X-ray generation process, the same large negative potential difference that is applied to the filament is also applied to the focusing cup. The surface of the focusing cup adjacent to

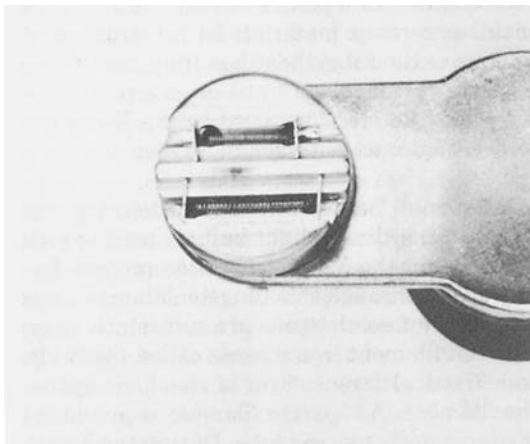


Figure 3. Focusing cup and filaments.

the filament is shaped in such a way as to focus the beam of accelerated electrons onto the anode. Therefore, the size of the filament and shape of the focusing cup control the size of the focal spot on the surface of the anode.

Anode

The anode (target) is the medium with which the electrons interact to produce X rays. In a typical X-ray tube used for medical applications, the anode is constructed of tungsten or an alloy of rhenium and tungsten. In the case of film-screen mammography, molybdenum or rhodium is often used for X-ray tube anodes because its characteristic X rays are better suited for breast imaging than the spectrum of X rays from tungsten. It is not necessary that the entire anode be constructed of the target material. Since the interactions of the electrons with the target take place near the surface, the target can be relatively thin and backed by a material that exhibits better heat-transfer characteristics, such as copper or graphite. During the X-ray generation process, a large positive electrical potential difference is applied to the anode to attract electrons emitted by the filament. As a result, the electrons are accelerated during their passage from the filament to the anode by the potential difference between the filament (cathode) and the anode, typically 20,000–150,000 V. Because the voltage supplied to the X-ray tube from a single-phase, full-wave-rectified source varies from zero to maximum and back to zero every 8.3 ms, this accelerating voltage is specified by its maximum value in units of peak kilovolts (kVp). Although three-phase or high frequency voltage sources exhibit considerably less voltage variation, their voltage waveforms are also described in terms of peak voltage. An X-ray tube with an anode-to-cathode potential difference of 100 kVp produces a spectrum of X rays with a maximum energy of 100 keV.

The surface of the anode is angled slightly from perpendicular with the electron beam. This angle is referred to as the target angle. The area on the anode where the electrons interact to produce X rays is called the focal spot or focus. The geometry of the X-ray tube components is designed so that the X-ray beam exits the tube at a 90° angle with respect to the electron beam inside the tube. The effective size of the focal spot is the size of the spot on the anode projected along the central axis of the X-ray beam. Since the width of the focal spot is the dimension along the axis perpendicular to both the electron beam and the X-ray beam central axis, only the length of the focal spot is affected by the target angle. The effective focal spot size is important because it determines an X-ray tube's ability to image small objects. The focal spot area in the plane of the surface of the anode is important because it limits the number of electrons per unit time that can impinge on the target without overheating. The relationship between the effective focal spot size and the actual size on the anode is given by the following formula:

$$f = F \sin \alpha$$

where f is the length of the focus effective size, F is the length of the actual focus on the anode, and α is the target angle. The concept of having a large focal area, over which

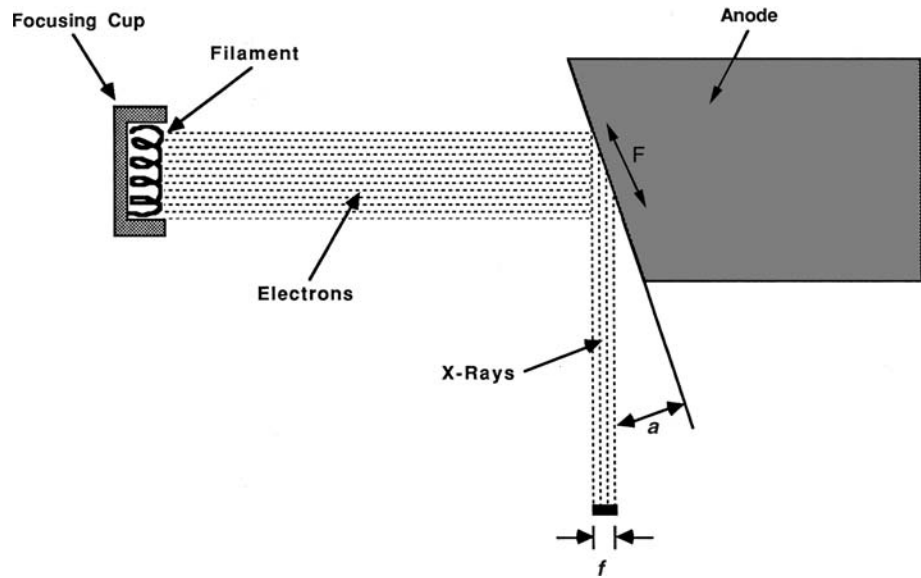


Figure 4. Line-focus principle.

to distribute heat while maintaining a small focal area for imaging by use of an angled target, is called the line-focus principle and is demonstrated in Fig. 4. For example, a 12° target angle with an effective focal spot of 1.0 × 1.0 mm has an actual focus size of 4.8 × 1.0 mm.

As illustrated in Fig. 5, the maximum X-ray field size is the largest field size that can be produced by the tube at a specified distance from the focal spot. Since X rays emitted by the focal spot at an angle greater than the target angle will be absorbed by the anode, the maximum field size (centered on the X-ray beam) along the dimension parallel to the anode–cathode axis of the X-ray tube is dependent on the target angle as follows:

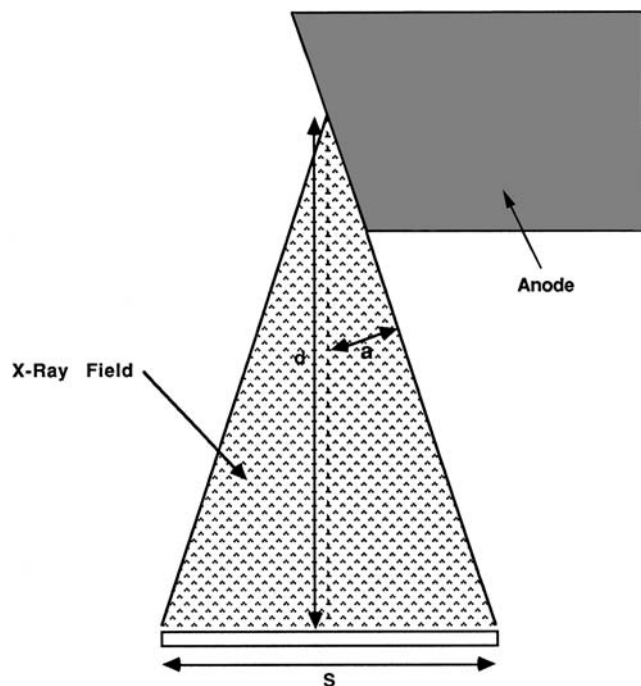


Figure 5. Maximum X-ray field size.

where S is the maximum size of the X-ray beam, d is the distance from the focal spot along the central axis of the X-ray beam, and α is the target angle. Therefore, although a larger target angle will result in a larger maximum field size, it will also result in lower heat loading capability. Conversely, a smaller target angle will result in a higher heat loading capability, but a smaller maximum field size. Simple algebra using the formula for maximum field size results in the conclusion that a target angle of at least 12° is required to cover the typical diagnostic imaging maximum field size of 43 cm at the commonly used distance of 100 cm. Diagnostic X-ray tube target angles typically are in the range of 12–15°, with a few special-purpose X-ray tubes having target angles as small as 9°. The intensity distribution of X rays along the direction parallel with the anode–cathode axis of the X-ray tube is not uniform. As shown in Fig. 6, comparison of the intensity at points equidistant from the central axis of the X-ray beam and equidistant from the focus indicates that the intensity of points toward the anode is less than that of corresponding points toward the cathode. This effect is called the heel effect. Since the X rays are generated from interactions of

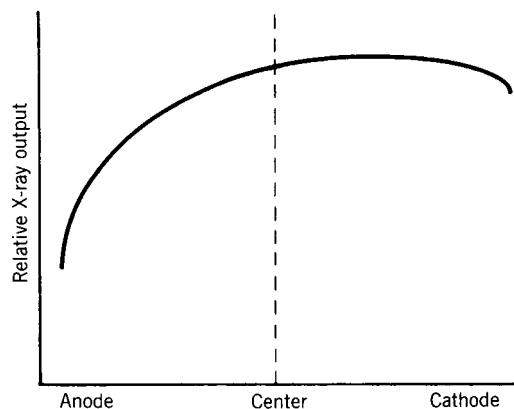


Figure 6. Heel effect.

the electrons slightly below the surface of the anode, X rays at the anode end of the field will have had to travel through additional anode material when compared with the cathode end. This attenuation of the X rays by the anode material produces a reduced intensity at the anode end. It also selectively reduces the number of low energy photons, which causes the spectrum of X-ray energies at the anode end of the X-ray field to have relatively more high energy photons and fewer low energy photons than the cathode end. As a result, the X rays at the anode end have a higher average penetrating power.

Due to the limiting effect of target angle on field size and the desire for higher heat loading capabilities, the rotating anode was developed. Rather than continuously directing the electron beam toward the same area on the anode throughout an X-ray exposure, directing the electron beam to different areas on the anode during the exposure will produce higher heat loading capability. By using a disk-shaped anode and rotating the disk about its central axis, the focus traces a circular path, or focus track, on the anode during an exposure. Since heat is distributed over a larger area, the X-ray tube's heat loading capability is greatly increased. For example, a rotating anode with a 10 cm diameter focus track and a 1.0 mm focal spot is capable of a heat loading >300 times that of the same anode held stationary during the exposure. Due to their exceptional heat loading advantage, essentially all X-ray tubes used for medical imaging have rotating anodes. On the other hand, fixed-anode tubes are used for dental radiography since the techniques employed do not produce enough heat to require the use of a rotating anode. Typical rotating-anode diameters are 7.5–10 cm, with some special purpose X-ray tubes employing anodes that have a diameter of 12.5 cm.

The tungsten disk in a rotating-anode X-ray tube is supported on its central axis by a molybdenum stem that is concentrically connected to a copper cylinder, or rotor. The rotor is supported internally on ball bearings to allow the entire anode/stem/rotor assembly to rotate. Figure 7 shows a typical tungsten rotating-anode assembly. Unlike a fixed-anode tube in which heat is transferred from the anode by conduction, the mounting of a rotating anode is designed to insulate the bearing from the heat contained in

the anode and, as a result, to prevent damage to the bearings. Heat transferred to the bearings decreases their lubrication, resulting in premature bearing failure. A rotating anode transfers heat to the surrounding tube housing by radiation, rather than by conduction.

Although the electrons are focused to a specific area on the anode, they may rebound from the focus with sufficient energy to interact at other positions on the anode to generate X rays. This off-focus radiation produces imaging artifacts. Generally, rotating-anode tubes generate more off-focus X rays than fixed-anode tubes because they have larger tungsten surfaces with which to interact. Although a significant amount of the output of an X-ray tube may be due to off-focus radiation, a well-designed collimator attached to the tube will reduce the affect of this radiation on clinical imaging.

Focus

As stated previously, a typical diagnostic X-ray tube has two focal spots, one approximately half the size of the other, with nominal sizes ranging from 0.3 to 2.0 mm. The larger focus is used for clinical applications that demand the shortest possible exposure time (consequently maximum heat generation). Conversely, the smaller focus is used for applications that require the capability to resolve the smallest anatomical features. X-ray tubes incorporating two focal spot sizes are usually built so that their foci are superimposed on the surface of the anode, even though two separate filaments are used to produce the two focal spots. This design alleviates alignment problems that would occur if the foci were not superimposed. However, X-ray tubes have been manufactured that employ a biangular anode so that the focus tracks of the two focal spots are on separate areas of the anode with different target angles. This design allows optimization of the target angle for each of the foci, but also causes some alignment problems.

When selecting an X-ray tube, one of the important factors to consider is the size of the focal spot. In order to have a standardized method of defining focal spot size, the National Electrical Manufacturers Association (NEMA) in the United States and the International Electrotechnical Commission (IEC) in Europe have each developed standards that address focal spot size specification and measurement. Since the X-ray tube manufacturing process is somewhat imprecise, the labeled focus size is a useful nominal value when comparing the relative sizes among different tubes. However, the actual focal spot size may vary considerably from the nominal size. In fact, measured focus length as much as 100% greater than the nominal size are considered to be within the tolerance specified by the standards. For example, a nominal 1.0 mm focus may have a measured length as large as 2.0 mm. Several methods of characterizing the focal spot have been developed: (1) slit method, (2) resolution or star pattern method, and (3) pinhole method.

The slit method employs a rectangular slit 10 μm wide and at least 5000 μm long used at enlargement factors of 1.0–3.0. The long axis of the slit is placed perpendicular to the axis of the focal spot whose dimension is desired. The resulting image on the film is a one-dimensional (1D)

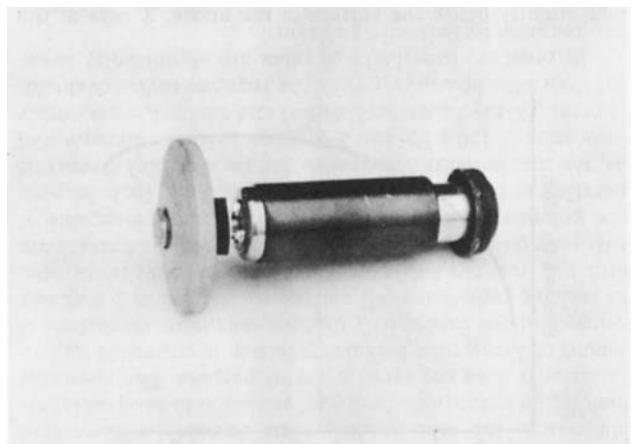


Figure 7. Rotating anode assembly.

distribution of the focal spot intensity along the chosen axis. Two images are therefore necessary to measure both the width and length of the focal spot. Both NEMA and the IEC specify the slit method as the method to be used for measuring the dimensions of focal spots. However, while the NEMA standard specifies measurement by eye using a 5–10X magnifying glass that has a built-in calibrated graticule, the IEC standard specifies the use of a scanning microdensitometer to obtain the line-spread function (LSF) of the width or length of the focal spot. The dimension of the focal spot is specified as the full width at 15% of maximum of the LSF.

The resolution method measures the ability of the focal spot to spatially resolve small objects and specifies the results in terms of spatial frequency (e.g., line pairs per mm), but does not provide a measurement of the dimensions of the focal spot. In this method, a line-pair resolution phantom is placed on the central axis of the X-ray beam between the focal spot and a single-emulsion film. The line-pair phantom consists of alternating lines of absorbing material (usually lead) and spaces, where adjacent lines and spaces have the same width. The width of the line pairs varies systematically over the face of the phantom to provide a range of line-pair sizes. The largest line-pair width, for which the line pair is completely blurred, is proportional to the focal spot size along the X-ray tube axis perpendicular to the line pairs. In order to facilitate the measurement of focal spot sizes by this method, a specialized phantom, commonly called a star pattern, which consists of alternating wedges of absorbing material and spaces in a circular configuration, was developed. Star pattern line-pair widths vary continuously from the outside edge to the center of the phantom. Using this phantom, both dimensions of the focal spot can be characterized from the same exposure. Since the phantom is circularly symmetric, the orientation of the phantom relative to the axes of the X-ray tube is unimportant. The image of the phantom on the film is analyzed by finding the points closest to the outside edge of the phantom for which the line pairs are completely blurred. The focal spot resolution can then be calculated using the following formula:

$$R = 180M/\pi\theta D$$

where R is the focal spot resolution in line pairs per millimeter, θ is the angle in degrees of a single wedge of absorbing material, D is the diameter of the blur on the film in millimeters, and M is the magnification factor of the phantom image. For the typical star pattern phantom with 2° wedges, this formula reduces to

$$R = 28.65M/D$$

The diameter corresponding to the resolution of the focal spot parallel with the anode–cathode axis of the X-ray tube (length) is the blur diameter that is perpendicular to this axis. Likewise, the diameter corresponding to the focal spot resolution perpendicular to the anode–cathode axis (width) is the blur diameter that is parallel with the anode–cathode axis. The principal advantage of the star pattern method of characterizing focal spots is its

ease of use. It does not require long exposure times and evaluates both the width and length of the focal spot from one exposure. The disadvantage of this method is that it does not provide an indication of the shape or intensity distribution of the focal spot.

The pinhole method utilizes a small circular pinhole 30–100 μm in diameter to image the focal spot onto a single-emulsion X-ray film for the purpose of determining its overall shape, orientation, intensity distribution and relative location. Although this method was used in the past to measure the dimensions of the focal spot, it is no longer considered suitable for doing so. The pinhole must be carefully aligned with the central axis of the X-ray beam. The distances between the focal spot and pinhole and between the pinhole and the film are adjusted so as to provide an enlargement factor of 1.0 or 2.0, depending on the nominal focus size. The film is exposed for a sufficient amount of time to produce an image of the focal spot with a density of 0.8–1.2 O.D. above base plus fog. Figure 8 illustrates the measurement geometry. The pinhole image is examined by eye using a magnifying glass, or by use of a scanning microdensitometer to quantitatively characterize the focal spot.

The focal spot size depends on the technique factor used during an exposure. Generally, the size of the focus increases as the tube current (mA) increases, and slightly decreases as the tube potential (kVp) increases. An increase in X-ray tube current means that more electrons travel from the filament to interact with the anode. The self-repulsion of the electrons causes the electron beam to slightly defocus, thereby interacting with a larger area on the anode. Since neighboring electrons in the beam are closer and more plentiful for larger tube currents, the effect of self-repulsion increases the focus size. An increase in

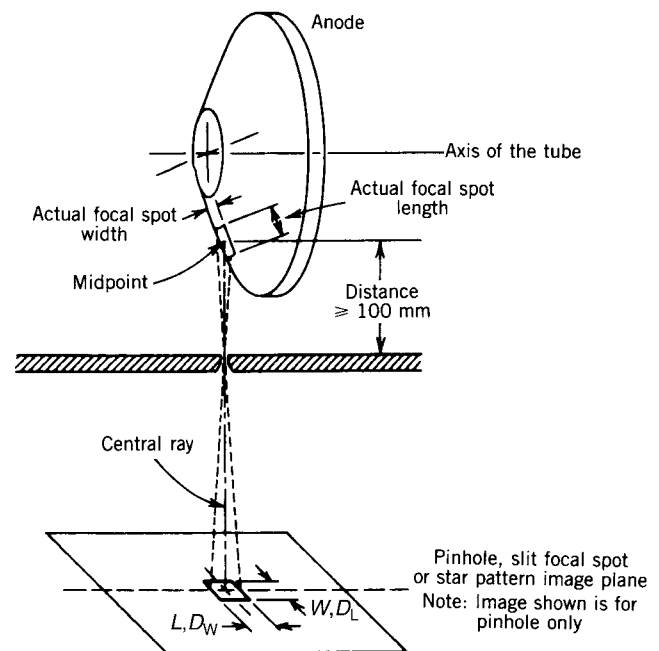


Figure 8. Focal spot size measurement geometry.

X-ray tube potential means that the electrons that travel from the filament to the anode are more energetic, and therefore have higher velocities. In addition, since neighboring electrons experience self-repulsion for a shorter period of time, the effect of self-repulsion is diminished and the focus size decreases. It is because of the effect of tube potential and current on focus size that both NEMA and the IEC specify the exposure factors for which focal spot sizes are measured. For a typical diagnostic X-ray tube, factors of 75 kVp, 50% of the maximum rated tube current at this tube potential, and 0.1 s exposure duration are employed.

Due to the line-focus principle, the projected focal spot size is dependent on the position in the X-ray field from which the actual focus on the anode is viewed. The observed target angle from a point at the cathode end of the X-ray field is larger, thereby producing a larger focal spot. Conversely, the observed focal spot size at the anode end is smaller. For example, the projected focus size at the cathode edge of a 40-cm field is 1.9 mm for an X-ray tube with a 12° target and a nominal focus size of 1.0 mm. All focal spot sizes are therefore specified relative to the central axis of the X-ray beam.

Stator

By surrounding the neck of the X-ray tube with a coil of wire in combination with a copper rotor attached to the anode inside of the tube forms an electric motor that rotates the anode. Exciting the coil with 60 Hz alternating current (ac) results in the anode being rotated at 3600 rpm, or one revolution every 16.7 ms, since it acts as a synchronous motor. However, this rotation speed is insufficient to avoid overheating the focal spot while meeting the clinical need for short, intense X-ray exposures of only a few milliseconds duration. For example, a 3 ms exposure would use only 18% of the available focal track. By applying 180 Hz ac to the stator windings so that the anode is accelerated to $\sim 10,000$ rpm within 1–2 s, heat will be spread over a greater percentage of the focal track (e.g., 50% of the track for a 3 ms exposure). The delay while the anode accelerates to the desired speed is called the “prep” time. During the prep time, the filament circuit is boosted so that the filament will have already reached its operating temperature by the time the X-ray exposure is initiated (when the anode has reached its desired speed). As soon as the exposure is completed, the accelerating voltage is removed from the stator. Due to mechanical resonances in the anode bearings, it is necessary to apply a braking voltage to the stator windings to decelerate the anode from 10,000 to 3,600 rpm within ~ 20 s. This action prevents destruction of the bearings by mechanical resonances. The anode is allowed to coast to a standstill once its speed is < 3600 rpm. Maintaining the anode speed at 10,000 rpm for long periods of time causes the bearings to wear out quickly.

Window. In most diagnostic X-ray tubes, the glass envelope of the tube insert serves as a window through which the X-ray photons pass. However, for X-ray tubes designed for applications requiring low energy photons, a

beryllium window is used because beryllium, unlike glass, does not significantly attenuate low energy photons.

Tube Housing

The X-ray tube insert must be contained within a metal tube housing, shown in Fig. 9, which provides a number of functions: mechanical protection against breakage of the glass envelope, an electrically grounded enclosure for safety, lead shielding against stray radiation emitted by the X-ray tube, a means to transfer heat from the anode to the outside environment, connections for the electrical cables from the high voltage power supply or generator, as well as a mechanical mounting for attachments. A glass or plastic window in the tube housing that is aligned with the central axis of the X-ray beam serves as a means for the useful X-ray beam to exit the tube. It also offers a means for directly viewing the anode to determine its physical condition. The volume between the insert and the tube housing is filled with oil to aid the transfer of heat deposited in the anode from the insert to the housing.

Although the X-ray photons emitted toward the window are desired, X-rays are also produced in many other directions. For this reason, the tube housing must contain sufficient lead shielding to reduce this “leakage” radiation to acceptable levels, since it does not serve a useful purpose. The allowable amount of X-ray tube leakage radiation is controlled by government regulations that specify that the maximum magnitude of the leakage radiation shall not exceed 100 mR in 1 h at any point 1 m from the focal spot when the X-ray tube is operated at its maximum continuous rated tube current and maximum

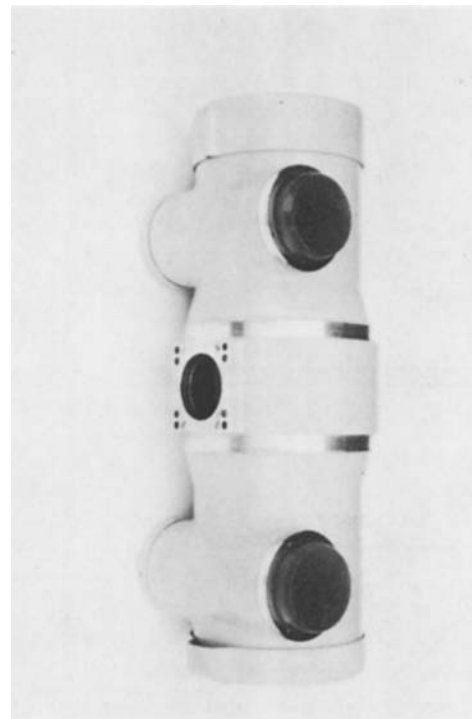


Figure 9. Typical medical X-ray tube housing.

tube potential. Although an X-ray tube may be capable of operating at tube currents as high as 1000 mA for brief periods, the maximum continuous rating for a typical diagnostic X-ray tube is 150 kVp and 3–6 mA. All manufacturers of diagnostic X-ray tubes, both dental and medical, must certify that their tube housings provide the required degree of shielding. In practice, tube leakage rates are significantly lower than required by regulations.

Most X-ray systems balance the tube potential difference between the anode and cathode so that the voltage difference between anode and ground is equal to the voltage difference between ground and the cathode. In other words, a 100 kVp tube potential means that the anode is 50 kVp positive with respect to ground, and the cathode is 50 kVp negative with respect to ground. For this reason, special high voltage connectors are required for both the anode and cathode. The standard high voltage connector contains three receptacles, each insulated from ground. In the anode connector, all three receptacles are electrically connected together, as well as to the anode. In the cathode connector, one receptacle is connected to the cathode, one receptacle is connected to the small focus filament, and one receptacle to the large focus filament. The receptacle connected to the cathode also acts as the common connection for the small and large filaments. In this way, the filaments are maintained at the same potential difference from ground as the cathode or focusing cup. The exceptions are X-ray tubes that utilize a biased focusing cup to shape the focal spot distribution, necessitating a cathode high voltage connector with four receptacles instead of three. The fourth receptacle is used to apply a bias voltage between the focusing cup and the filaments. Since the connectors are standardized, X-ray tubes from different manufacturers can be easily used with an X-ray system.

As discussed previously, X-ray photons emitted by the tube have a spectrum of energies. Since lower energy photons have little penetrating power, they tend to be absorbed near the skin surface of a patient undergoing a diagnostic exam and do not contribute to the X-ray image. For patient protection purposes, it is therefore desirable to eliminate or filter these low energy photons before they reach the patient. By reducing the low energy content of the beam, the materials used in the construction of the tube insert and tube housing that are in the path of the useful X-ray beam serve this purpose. Aluminum permanently mounted in the tube housing window provides additional filtration. All of the filtration that is permanently in the useful beam is referred to as the inherent filtration. Government regulations specify the minimum amount of inherent filtration that must be present in the X-ray beam as a function of the maximum tube potential at which the tube is operable.

HEAT DISSIPATION IN X RAY TUBES

Much of the design of an X-ray tube is predicated on the heat generated by the inefficiencies of the X-ray production process. Although the use of a rotating anode increases the ability of the focal spot to handle the instan-

taneous heat load during the short period of X-ray generation, there are limitations as to the total amount of heat that the anode and tube housing can each store. Basically, there are five parameters that describe the heat dissipation characteristics of an X-ray tube: (1) focal spot loading, (2) anode heat capacity, (3) anode cooling rate, (4) housing heat capacity, and (5) housing cooling rate.

These heat characteristics are described in terms of one of two units, heat units or watt-seconds. In the United States, heat parameters are specified in heat units (HU). The number of heat units generated by an X-ray exposure is given by the following formula:

$$HU = CVA t$$

where HU is the number of heat units, C is a waveform factor, V is the tube potential in peak kilovolts, A is the tube current in milliamperes, and t is the exposure duration in seconds. The waveform factor corrects for the differences in heating effect resulting from differences in average power of different high voltage waveforms. For a full-wave-rectified, single-phase waveform, the waveform factor is 1.0. For a 12-pulse, three-phase waveform, the waveform factor is 1.35. For example, an X-ray exposure of 100 kVp, 500 mA, 0.2 s using a single-phase generator would result in 10,000 HU. The same technique using a three-phase generator would result in 13,500 HU.

In Europe, heat parameters are expressed in watt-seconds (W·s) or joules (J). The heat generated by an X-ray exposure is given by the following formula:

$$E = KVA t$$

where E is the heat in watt-seconds or joules, K is a waveform factor, V is the tube potential in peak kilovolts, A is the tube current in milliamperes, and t is the exposure duration in seconds. In this case, the waveform factor for a full-wave-rectified, single-phase waveform is 0.74 and for a 12-pulse, three-phase waveform is 1.0. Therefore, the quantity of heat represented by one heat unit is 26% less than the quantity represented by 1 W·s or joule. Using the technique factors of the previous example (100 kVp, 500 mA, 0.2 s) results in 10,000 J of heat generated.

Focal Spot Loading

The heat handling characteristic of the X-ray tube focal spot is expressed in terms of kilowatts for an exposure duration of 0.1 s. The kilowatt loading of the focal spot is equal to the product of the tube potential in peak kilovolts and the tube current in milliamperes divided by 1000. It is the amount of power, or energy per second, deposited as heat in the anode. Typical X-ray tube inserts used in diagnostic radiology for high power procedures have focal spot loading capabilities of 30–40 kW for a small focus of 0.6 mm, and 100 kW for a large focus of 1.2 mm. The small focus of such a tube could not be used for the example technique of 100 kVp, 500 mA, because the focal spot loading would be exceeded. This demonstrates the trade-off between focal spot size and heat loading capability. A focal spot of 0.3 mm may have a maximum heat loading of as little as 9 kW. A tube insert with this rating could not be

used with a tube current >90 mA at 100 kVp, or >125 mA at 72 kVp, for an exposure duration of 0.1 s.

Anode Heat Storage Capacity

The maximum amount of heat that can be stored in the anode at any one time is its heat capacity. Depending on the manufacturer, the heat capacity is specified in terms of heat units or joules. The size and construction of the anode determine its heat capacity. Rotating anodes 3 in. (7.6 cm) in diameter have less heat storage capacity than anodes 4 in. (10.2 cm) in diameter. The typical "high power" diagnostic X-ray tube has a heat capacity of 300,000–400,000 HU. This is equivalent to the heat generated by 22–29 three-phase exposures of 100 kVp, 500 mA, 0.2 s each. Since the heat storage capacity is defined by the amount of energy deposited in the anode that raises its temperature from some reference temperature to the maximum operating temperature of the anode, the specified heat storage capacity will depend on the choice of reference temperature. For example, the storage capacity of an insert may be 400,000 HU when referenced to a "warm" anode, but is 600,000 heat units when referenced to ambient room temperature, that is, a "cold" anode. For this reason, it is important to be aware of the corresponding reference temperatures when comparing anode heat storage capacities of tube inserts of different manufacturers.

Anode Cooling Rate

Heat deposited in the anode is transferred to the tube housing. The rate at which the anode cools is dependent on the temperature difference between the anode and the tube housing. When the maximum rated amount of heat is stored in the anode, the cooling rate will be maximum; when the anode is cold, the cooling rate will be minimum. For this reason, the cooling rate of the anode is given as a cooling curve of heat stored in the anode as a function of time, with no additional heat deposited, as illustrated in Fig. 10. For an X-ray tube with a maximum anode cooling rate of $72,000 \text{ HU}\cdot\text{min}^{-1}$, at least 11 s would be required to transfer the heat generated by a 100 kVp, 500 mA, 0.2 s three-phase exposure from the anode to the tube housing. Since the maximum continuous heating rate is equal to the maximum cooling rate, the anode of this tube could be subjected to $1200 \text{ HU}\cdot\text{s}^{-1}$ continuously (e.g., 100 kVp, 8.9 mA).

Tube Housing Heat Capacity

Just as the anode is limited as to the amount of heat that can be stored in it at any one time, the heat capacity of the tube housing is also limited. Since the housing has a much larger mass than the anode, its heat capacity is significantly greater. The typical diagnostic X-ray tube housing has a heat storage capacity of 1,500,000 HU. This is equivalent to the heat generated by 111 three-phase exposures of 100 kVp, 500 mA, 0.2 s each.

Tube Housing Cooling Rate

Heat deposited in the tube housing is transferred to the surrounding ambient environment. The maximum housing

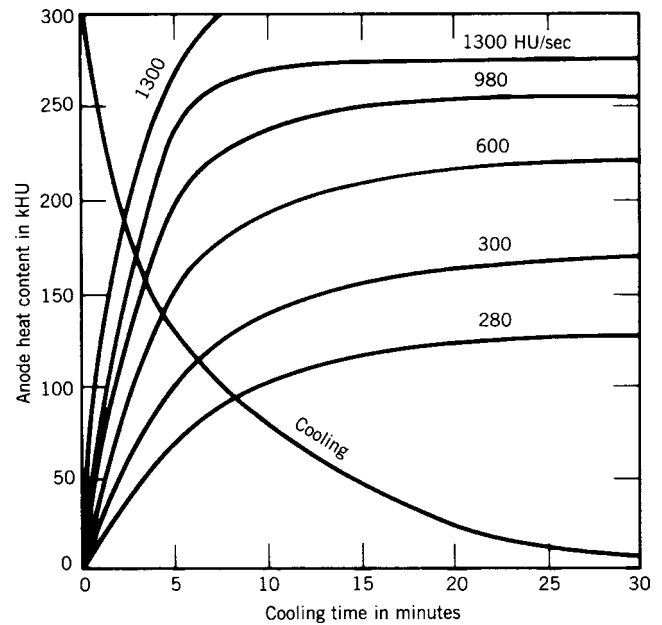


Figure 10. Anode cooling/heating curves.

cooling rate occurs when the housing is heated to its maximum capacity; the minimum cooling rate occurs when the housing contains little heat. The maximum cooling rate of a typical tube housing is $\sim 15,000 \text{ HU}\cdot\text{min}^{-1}$, unless additional measures have been taken to increase the cooling rate. A typical tube housing that has a fan to increase air flow over the housing has a maximum cooling rate of $\sim 24,000 \text{ HU}\cdot\text{min}^{-1}$. Another method of increasing the cooling rate is to constantly circulate the oil contained in the housing through an external heat exchanger. This method results in a maximum cooling rate of $\sim 50,000 \text{ HU}\cdot\text{min}^{-1}$. For high work load applications where X-ray exposures occur frequently enough that the heat generated by the exposures will not be totally dissipated from the X-ray tube before the next exposure, the cooling rate of the housing determines the rate at which heat can be generated in the tube. It is for this reason that manufacturers use the additional cooling measures discussed. For example, at least 54 s would be required to transfer the heat generated by a 100 kVp, 500 mA, 0.2 s three-phase exposure from a housing with a cooling rate of $15,000 \text{ HU}\cdot\text{min}^{-1}$. The addition of a fan would reduce this cooling time to 34 s.

Tube Rating Chart

To allow X-ray equipment users to easily determine whether a desired exposure technique exceeds the heat characteristics of an X-ray tube, all tube heat parameters are combined into one chart called a tube rating chart. A chart is provided by the manufacturer for each focal spot size and anode rotation speed. Each chart consists of a set of curves of tube potential versus exposure duration for different tube currents that defines the maximum single exposure capability of the X-ray tube when the tube is cold. An exposure technique that falls "below" the corresponding

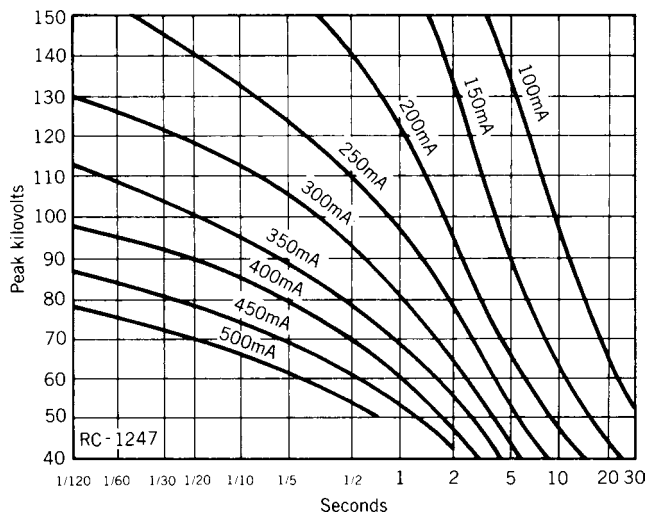


Figure 11. Tube rating chart.

curve on the chart is allowed; a technique that is “above” the corresponding curve is not allowed.

Using the tube rating chart shown in Fig. 11, the point corresponding to a technique of 100 kVp, 500 mA, 0.2 s is above the curve for 500 mA. Therefore, this technique exceeds the heat capabilities of this particular X-ray tube and is not allowed. Examination of this 500 mA rating curve reveals that a tube potential of 100 kVp is not allowed, regardless of the duration of the X-ray exposure. The point corresponding to 100 kVp and 0.2 s indicates that a maximum tube current of ~280 mA is obtainable for 100 kVp and 0.2 s exposure duration. Similarly, the point where the 500 mA rating curve intersects 0.2 s indicates that a maximum tube potential of 54 kVp is obtainable for a 500 mA and 0.2 s exposure technique. Since the amount of film blackening is proportional to the product of the tube current and exposure duration, the technique of 100 kVp, 500 mA, 0.2 s could be changed to 100 kVp, 250 mA, 0.4 s without affecting the resulting diagnostic image. The point on the tube rating chart for this latter technique is below the 250 mA rating curve and the technique is therefore allowed. Some manufacturers provide tube rating charts that provide rating curves of tube current versus exposure duration for different tube potentials, but these can be utilized in a similar manner.

Single-exposure tube rating charts are not useful for determining the feasibility of multiple radiographic exposures (or radiographic exposures combined with fluoroscopic exposures) that may come close to exceeding a tube’s heat capabilities. The appropriate cooling and heating curves, combined with direct calculation of heat units produced, must be used in these cases. The data sheets for X-ray tubes designed for serial exposure applications such as cineangiography often include tables of maximum allowable techniques for serial exposures.

X-RAY TUBE FAILURE

Like electric light bulbs, X-ray tubes have a limited length of life. Due to the electrical, mechanical, and heat stresses

that an X-ray tube experiences during its normal use, there are several types of X-ray tube failures.

Bearing Wear

Since the X-ray tube insert is a sealed envelope, the bearings of a rotating anode can only be lubricated at the time of manufacture. Although the insert is designed to prevent heat deposited in the anode from reaching the bearings, sufficient heating of the bearings takes place in the long term to gradually reduce their lubrication. Prolonged operation at high rotor speeds also rapidly decreases the lifetime of the bearings. Eventually, the bearings become noisy and, if the tube insert is not replaced, will seize, thus preventing the anode from rotating.

Filament Evaporation

By design, an electric current is used to heat the filament to produce electrons. If this current is excessive, the filament will melt, resulting in an open filament. In addition, evaporation of the filament material is a natural consequence of filament heating. This may result in a loss of filament material sufficient to cause failure.

Cracked Anode

If an X-ray exposure is made with the anode stationary, the anode may crack due to intense heat build-up at the focal spot. This may be caused by either bearing or stator failure. Even if the problem of the stationary anode is resolved (i.e., the stator repaired) the cracked anode presents a safety problem because it may shatter, implode the glass envelope, and pierce the tube housing. It is also possible to crack a rotating anode if the anode is subjected to a very high heat-generating technique when it is cold. For this reason, some manufacturers recommend that X-ray tubes be warmed up by making several low tube-current, long-duration exposures (70 kVp, 75 mA, 6.0 s) whenever the tube has not been used for 1 h or more.

Tungsten Deposition

The heating of the tungsten filament and tungsten anode during normal operation of the X-ray tube causes some of the tungsten to evaporate and to build up gradually as a deposit on the glass envelope of the tube insert. This deposit causes the glass envelope to be less able to repel electrons reflected from the anode because it reduces the electrical insulating properties of the glass envelope, thereby reducing the magnitude of the induced negative charge on the envelope. The electron bombardment of the glass releases trapped gases that cause the electrical breakdown (arcing) of the tube when the amount of gas released becomes great enough.

BIBLIOGRAPHY

Further Reading

Brecher R, Brecher E. *The Rays: A History of Radiology in the United States and Canada*. Baltimore (MD): Williams & Wilkins; 1969.

- Coulam CM, Erickson JJ, Rollo FD, James AE Jr. *The Physical Basis of Medical Imaging*. New York: Appleton-Century-Crofts; 1981.
- Hendee W. *Medical Radiation Physics—Roentgenology, Nuclear Medicine and Ultrasound*, 2nd ed. Chicago (IL): Year Book Medical Publishers; 1979.
- International Electrotechnical Commission. *X-Ray Tube Assemblies for Medical Diagnosis—Characteristics of Focal Spots*. Publ. No. 60336, Geneva: IEC, 2005.
- National Electrical Manufacturers Association. *Measurement of Dimensions and Properties of Focal Spots of Diagnostic X-Ray Tubes*. Stand. Publ. No. XR 5-1992 (R1999), Washington (DC): NEMA; 1999.
- Ter-Pogossian MM. *The Physical Aspects of Diagnostic Radiology*. New York: Harper & Row; 1969.

See also CODES AND REGULATIONS: RADIATION; SCREEN-FILM SYSTEMS; X-RAY EQUIPMENT DESIGN.

References for Radiotherapy, heavy ion:

BIBLIOGRAPHY

- Mould RF. *A Century of X-rays and Radioactivity in Medicine*: Institute of Physics Publishing; 1993.
- Attix FH. *Introduction to Radiological Physics and Radiation Dosimetry*: John Wiley & Sons; 1986.
- Khan FM. *The Physics of Radiation Therapy*, Second Edition: Williams & Wilkins; 1994.
- Hendee WR, Ibbott GS, Hendee EG. *Radiation Therapy Physics*: Wiley-Liss; 2005.
- Petti PL, Lennox AJ. Hadronic radiotherapy, *Ann Rev Nuclear Part Science* 1994;44:154–197.
- Scharf WH. *Biomedical Particle Accelerators*: AIP Press; 1994.
- Bewley DK. *The Physics and Radiobiology of Fast Neutron Beams*: Institute of Physics Publishing; 1992.
- Podgorsak EB, Metcalfe P, Van Dyk J. *Medical Accelerators*: Chapter 11 in Van Dyk J ed. *The Modern Technology of Radiation Oncology*, Medical Physics Publishing; 1999.
- Chu, WT, Ludewig BA, Renner TR. Instrumentation for treatment of cancer using proton and light-ion beams. *Rev Sci Instrum* 1993;64:2055–2122.
- Chen GTY. Radiotherapy, heavy ion; Verhey LJ, Proton beam radiotherapy. In: Webster JG, editor. *Encyclopedia of Medical Devices and Instrumentation*. New York: John Wiley & Sons; 1988.
- Castro JR. Results of heavy ion radiotherapy, *Radiation and Environmental Biophysics (Historical Archive)* 1995;34:45–48.
- Heavy charged particles in research and medicine. *Proceedings of a Symposium*. Berkeley, California, May 1–3, 1985, *Radiat Res Suppl.* 1985;8:SI-334.
- Kraft G. RBE and its interpretation, *Strahlentherapie und Onkologie* 175 Suppl. 1999;2:44–47.
- Matsufujil N, Fukumura A, Komori M, Kanai T, Kohno T. Influence of fragment reaction of relativistic heavy charged particles on heavy-ion radiotherapy, *Phys Med Biol* 2003;48:1605–1623.
- Pollock BE ed., *Contemporary Stereotactic Radiosurgery: Technique and Evaluation*: Futura Publishing Company; 2003.
- Bomford CK, Kunklery IH, Walter and Miller's *Textbook of Radiotherapy Radiation Physics, Therapy, and Oncology* 6th Ed. Churchill Livingstone; 2003.
- Purdy JA, Grant III W, Palta JR, Butler EB, Perez CA eds. *3-D Conformal and Intensity Modulated Radiation Therapy: Physics & Clinical Applications*: Advanced Medical Publishing, Inc., 2001.
- Webb S, *The Physics of Three-Dimensional Radiation Therapy: Conformal Radiotherapy, Radiosurgery and Treatment Planning*: Institute of Physics Publishing; 1993.
- Hongstrom Kr, Paciotti MA, Smith AR, Collier M, Comparison of static and dynamic treatment modes for the pion therapy beam at LAMPF. *Int J Radiat Oncol Biol Phys* 1980;6(12):1693–700.
- Sisterson J ed. *Particle Therapy Co-operative Group (PTCOG) Newsletter*, Mass. General Hospital, Harvard. MA. Jan. 2005.
- Rossi S, Amaldi U, The TERA programme: status and prospects. In: Larsson B, Crawford J, Wienreich R, eds. *Adv. in Neutron Capture Therapy*, Vol. I: Elsevier Science; 1997.
- Podgorsak EB, Podgorsak MB, *Special Techniques in Radiotherapy: Chapter 17*. In: Van Dyk J, ed., *The Modern Technology of Radiation Oncology*, Medical Physics Publishing; 1999.
- Podgorsak FB, Podgorsak MB, *Stereotactic Irradiation: Chapter 16*. In: Van Dyk J, ed., *The Modern Technology of Radiation Oncology*, Medical Physics Publishing; 1999.
- Washington CM, Leaver D, *Principles and Practice of Radiation Therapy*, Second Edition: Mosby; 2004.
- Wang CC, ed. *Clinical Radiation Oncology: Indications, Techniques, and Results* Second Edition: Wiley-Liss; 2000.
- Hall EJ, *Radiobiology for the Radiologist*, 4th Edition: J.B. Lippincott; 1994.
- Rossi HH, Zaider M, *Microdosimetry and Its Applications*: Springer; 1996.
- Greene D, Williams PC, *Linear Accelerators for Radiation Therapy*, 2nd Edition: Institute of Physics Publishing; 1997.
- Nias AHW, *An Introduction to Radiobiology*: John Wiley & Sons, 1990.
- Prasad KN, *Handbook of Radiobiology*, 2nd ed.: CRC Press; 1995.
- Tsipenyuk YM, *The Microtron: Development and Applications*: Taylor & Francis; 2002.
- Dobelbower RR, Abe M, *Intraoperative Radiation Therapy*: CRC Press; 1989.
- Chiu C, Fomytskyi M, Grigsby F, Raischel F, Downer MC, Tajima T, *Laser electron accelerators for radiation medicine: A feasibility study*. *Medical Physics* 2004;31:2041–2042.
- Bostick WH, Possible techniques in direct-electron-beam tumor therapy, *Phys Rev* 1950;77:564–565.
- Shih CC, High energy electron radiotherapy in a magnetic field. *Medical Physics* 1975;2:9–13.
- Whitmire DP, D.L. Bernard DL, Peterson MD, Magnetic modification of the electron-dose distribution in tissue and lung phantoms. *Medical Physics* 1978;5:409–417.
- Bielajew AF, The effect of strong longitudinal magnetic fields on dose deposition from electron and photon beams. *Medical Physics* 1993;20:1171–1179.
- Litzenberg DW, Fraass BA, McShan DL, O'Donnell TW, Roberts DA, Becchetti FD, Bielajew AF, Moran JM, An apparatus for applying strong longitudinal magnetic fields to clinical photon and electron beams. *Phys Med Biol* 2001;46:N105–N115.
- Becchetti FD, Sisterson JM, Hendee WR, Point/counterpoint: high energy electron beams shaped with applied magnetic fields could provide a competitive and cost-effective alternative to proton and heavy-ion radiotherapy. *Medical Physics* 2002;29:2435–2437.
- Chen Y, Bielajew AF, Litzenberg DW, Moran JM, Becchetti FD, Magnetic confinement of electron and photon radio-

- therapy dose—a Monte Carlo simulation with a non-uniform longitudinal magnetic field. *Med Phys* 2005;32:3810–3818.
41. Maughan RL, Yudelev M, Neutron Therapy: Chapter 21. In: Van Dyk J ed., *The Modern Technology of Radiation Oncology*. Medical Physics Publishing, 1999.
 42. Forman JD, Yudelev M, Bolton S, Tekyi-Mensah S, Maughan R, Fast neutron irradiation for prostate cancer, *Cancer and Metastasis Reviews* 2002;21:131–135.
 43. Zamenhof RG, Busse PM, Harling OK, Goorley JT, Boron Neutron Capture Therapy: Chapter 24. In: Van Dyk J ed., *The Modern Technology of Radiation Oncology*, Medical Physics Publishing; 1999.
 44. Larsson B, Crawford J, Wienreich R, eds. *Adv. in Neutron Capture Therapy*, Elsevier Science; 1997.
 45. Moyers MF, Proton Therapy: Chapter 20. In: Van Dyk J ed., *The Modern Technology of Radiation Oncology*, Medical Physics Publishing; 1999.
 46. Goitein M, Lomax AJ, Pedroni ES, Treating cancer with protons, *Physics Today*, Sept. 2002;45–50.
 47. Coutrakon G, Bauman M, Lesyna D, Miller D, Nausbaum J, Slater J, Johanning J, DeLuca PM, Siebers J, Ludewigt B, A prototype beam delivery system for the proton medical accelerator at Loma Linda. *Medical Physics* 1991;18:1093–1099.
 48. Litzenberg DW, Roberts DA, Lee MY, Pham K, Vander Molen AM, Ronningen R, Becchetti FD, On-line monitoring of radiotherapy beams: experimental results with proton beams. *Medical Physics* 1999;26:992–1006.
 49. Li Q, Kanai T, Kitagawa A, A model to evaluate the biological effect induced by the emitted particles from a beta-delayed particle decay beam. *Physics Med Biol* 2003;48:2971–2986.
 50. Jaekel O, Kraemer M, Karger CP and Debus J. Treatment planning for heavy-ion radiotherapy: clinical implementation and application. *Phys Med Biol* 2001;46:1101–1116.
 51. Li Q, Kitagawa A, Kanazawa M, Urakabe E, Kanai T, Tomitani T, Sato S, Wei Z. The production of ⁹C beam in the secondary beam line of the HIMAC facility and its potential application in cancer therapy, *Nuclear Physics A* 2004;746:288–2292.
 52. Miyamoto T, Yamamoto N, Nishimura H, Koto M, Tsujii H, Mizoe J, Kamada T, Kato H, Yamada S, Morita S, Yoshikawa K, Kandatsu S, Fujisawa T, The Working Group for Lung Cancer, Carbon ion radiotherapy for stage I non-small cell lung cancer. *Radiotherapy and Oncology* 2003;66:127–140.
 53. Hirao Y, Ogawa H, Yamada S, Sato Y, Yamada T, Sato K, Itano A, Kanazawa M, Noda K, Kawachi K, M. Endo M, Kanai T, Kohno T, Sudou M, Minohara S, Kitagawa A, Soga G, Takada E, Watanabe S, Endo K, Kumada M, Matsumoto S, Heavy ion synchrotron for medical use – HIMAC project at NIRS – JAPAN, *Nuclear Physics A* 1992;538:541–550.
 54. Krämer M, Jäkel O, Haberer T, Kraft G, Schardt D, Weber U, Treatment planning for heavy-ion radiotherapy: physical beam model and dose optimization. *Phys Med Biol* 2000;45/11:3299–3317.
 55. Kim J, Marti F, Blosser H, Study of a superconducting cyclotron for heavy-ion therapy, *Cyclotrons and their applications*, 2001 F. Marti ed., AIP Conference Proceedings Vol 2002;600-1:324–326.

INDEX

Italic numbers preceding page references indicate the volume numbers. Page numbers followed by f denote figures; page numbers followed by t denote tables.

- Abbott Cell-Dyn 4000 Hematology System, 2:414–415
- Abscesses, in implant-related infections, 1:116
- “ABCs” of CPR (airway, breathing, circulation), 2:36, 37, 45, 46, 53–55, 56, 57
- Abdominal aortic aneurysm (AAA), monitoring of, 2:8
- Abdominal electrocardiogram, fetal, 3:291–293
- Ab initio* (new fold) protein structure prediction, 1:221
- ABIOMED Pneumatic Ventricular Assist Device, 3:452
- Ablation. *See also* Cryoablation; Tissue ablation
- cardiac catheter, 6:369–372
 - chemical, 6:367–368
 - cornea, 6:378
 - direct ultrasound, 6:365
 - electrosurgical, 3:166–167
 - endometrial, 6:375
 - endovascular, 6:376–378
 - imaging during, 6:374–375
 - intervertebral disk, 6:378
 - laser, 6:365–366, 374, 378
 - microwave, 6:364–365, 371, 374, 375–376
 - prostate, 6:375–376
 - radio frequency, 6:363–364, 370–371, 373–374, 376, 378
 - thermal, 6:362–367
 - tumor, 6:372–375
 - ultrasound, 6:365, 376
- Ablation devices, manufacturers of, 6:377–378t
- Ablation modalities, frequency of use of, 6:363t
- Abnormal thermogram patterns, 6:348–349
- Abrading of skin, for electrodes, 1:136–137
- Abrasion–corrosion, 1:311
- Abrasive wear, 1:315
- Absorbable biomaterials, 1:255–267
- resorbable calcium ceramics, 1:260–262
 - resorbable composites, 1:262–264
 - resorbable implants, 1:255–257
 - resorbable polymers, 1:257–260
- Absorbable polymers, as biomaterials, 1:107
- Absorbed dose, 5:505
- Absorbed radiation dose, 5:466–467
- determining, 5:467–471
- Absorbed X-ray dose, determination of, 6:586
- Absorber canister, in anesthesia machines, 1:40
- Absorptiometry
- dual-energy, 1:552–553
 - dual-photon, 1:551–552
 - single-photon, 1:551
- ABX Pentra 120 Hematology System, 2:415–416
- Accelerators
- as neutron sources for BNCT, 1:577–578
 - pulsed-laser, 6:13
 - superconducting, 6:12
- Accelerator treatment couches, 5:591–593
- Accelerometer, 2:5, 6f
- Accident investigation/reporting, 6:118–119
- Accreditation Board for Engineering and Technology (ABET), 1:404, 405–407, 408
- Accu-Chek™ blood glucose meter, 1:23
- Accumulated difference of slopes (ADIOS), 1:74
- Accuspin tubes, separation of peripheral blood mononuclear cells using, 1:463–464
- Acellular dermis, in tissue engineering, 1:371
- AC impedance standards, 1:158. *See also* Alternating current (AC)
- Acoustic attenuation, 6:455
- Acoustic bubble response, 6:467
- Acoustic events, in the cardiac cycle, 4:163
- Acoustic fields, 6:458–459
- Acoustic imaging, 6:459–465
- Acoustic immittance
- dynamic, 1:100–101
 - static, 1:99–100
- Acoustic immittance measurement, 1:99–101
- Acoustic impedance, 3:3; 4:66t; 6:454–455
- Acoustic reflex measurement, 1:101
- Acoustic stimulation, in cochlear prostheses, 2:138
- Acoustooptic tunable filters (AOTFs), 4:499
- benefits of, 4:459–461
 - in confocal microscopy, 4:457–459
- Acquired brain injury (ABI), 6:71
- Acquired immune deficiency syndrome (AIDS), 1:115
- Acquired nystagmus, 5:140
- Acquisition, of adequate medical device technologies, 6:117–118
- ACR-NEMA digital image standard, 5:333
- Acrylic bone cement, 1:540–550
- compositions of, 1:541–542
 - fatigue test results for, 1:547
 - mechanical properties of, 1:546–547
 - polymerization heat of, 1:543–544
 - porosity, volumetric changes, and residual stress of, 1:545–546
 - setting and cementing technique for, 1:542–543
 - viscosity of, 1:543–544
- Acrylic materials, in dentures, 1:327. *See also* Acrylics
- Acrylic preformed laminate veneers, 6:98
- Acrylic resins, unfilled, 6:93–94
- Acrylics, 1:277. *See also* Acrylic materials
- AC small signal impedance standards, 1:160. *See also* Alternating current (AC)
- Actigraphy, 6:212
- Actinotherapy. *See* Ultraviolet radiation (UVR)
- Action potential(s), 3:110
- cardiac, 3:143–144
- Action potential systems modeling, 5:320–322
- Activated carbons, 1:300
- Activated charcoal, medical applications of, 1:301–302
- Activation energies, temperature dependence of, 1:364
- Activation energy barrier, 1:344
- Active pits, corrosion in, 1:310
- Activities of daily living assessment/training, virtual reality for, 6:76
- Activities of daily living (ADL) devices, 1:447–448
- Activity, concentration and, 1:123
- Activity sensor, 5:221–222
- Acute cellular rejection (ACR), after liver transplantation, 4:272
- Acute coronary events, 2:50–51
- Acute exceptional blood loss anemia, hyperbaric medicine and, 4:26
- Acute medical care ventilators, 6:505–514
- control scheme in, 6:507–508
 - operation of, 6:508–511
 - use of, 6:505–508
- Acute normovolemic hemodilution (ANH), 1:515
- Acute traumatic ischemias, hyperbaric medicine and, 4:25–26
- Adaptive blood pressure controller, 1:492–496
- Adaptive driving, for wheelchair users, 4:550–552
- Adaptive radiotherapy, 6:401
- Adaptive skiing, 4:550
- Adaptive threshold detection, in neonatal respiration monitoring, 5:21
- Addition polymerization, 1:330–331
- biomaterials via, 1:274

- Adenomatoid malformation, congenital cystic, 4:172–173
- Adherent cell monitoring, using impedance spectroscopy, 4:137–138
- Adhesion
of diamond-like carbon coatings, 1:318–319
in engineered tissue, 3:191
- Adhesive dental liners, 1:324–325
- Adhesive electrodes, solid conductive, 1:141
- Adhesives, dental, 1:324, 325
- Adhesive tape, with electrodes, 1:138, 139
- Adhesive wear, 1:315
- Adicol Project artificial pancreas, 5:228–229
- Adjustable pressure-limiting (APL) valve, 1:32, 33
- Administration, role in a total hospital safety program, 6:116
- Administration–information processing, for communication disorders, 2:211–213
- Adoptive immunotherapy. *See* Immunotherapy
- Adoptive T cell immunotherapy, of cancer in lymphopenic host, 4:115–116
- Adsorption, specific, 1:123–124
- Adult cardiopulmonary resuscitation (CPR), 2:53–55
- Adult mesenchymal stem cells (MSCs), in cartilage regeneration, 2:73
- Adults
electrogastrogram in, 3:91–92
high frequency ventilation in, 3:508–509
- Adult stem cells, in tissue engineering, 6:382–383. *See also* Adult mesenchymal stem cells (MSCs)
- Advamed, 4:316
- Advanced anesthesia monitoring, computer utility in, 1:43–44
- Advanced Cardiac Life Support (ACLS), 2:35, 48, 49–50
- Advanced functional imaging modalities, based on fluorescent microscopy, 4:498–502
- Advanced muscle assessment methods, 6:65–66
- Advanced radiation therapy, computed tomography simulation for, 2:273–275
- Adventitia layer, in arterial walls, 1:85
- Aeration components, in microbioreactors, 4:389
- Aerosols, nanoparticle, 5:3
- Affinity-based impedimetric biosensors, 4:143
- Affinity chromatography, 2:104–105. *See also* Chromatography
- Affymetrix GeneChips analysis, 1:223, 224
- Agarose, in tissue engineering, 1:370–371
- Age-dependent movements, 1:396
- Age-related macular degeneration (AMD), 5:293
clinical evaluation of, 5:293–294
- Aggregate display feedback schedule, 1:176–177
- Aging, myths about, 1:389
- AICD implantable defibrillator, 1:71
- Air
cobalt unit calibration in, 2:128–129
CPR and, 2:39–40
- Air conditioning, as a hospital problem, 6:112
- Air embolism, in hyperbaric medicine, 4:22
- Air embolism etiology, arterial and venous, 4:22t
- Air-filled capsule/vest, in neonatal respiratory monitoring, 5:16
- Air-filled sensors, 1:175
- Airflow, measurement of, 6:522
- Air kerma, determining, 5:467–471
- Air plethysmography, 5:239
- Airway obstruction, cardiopulmonary resuscitation and, 2:55–56
- Airway pressure, 6:516
- Airway pressure monitor, anesthesia machine, 1:39
- Airway pressure monitoring, during high frequency ventilation, 3:505
- Airy disk, in confocal microscopy, 4:462–463
- Alanine, as a chemical dosimeter, 5:474
- Alarming, based on hemodynamic parameters, 4:574
- Alarm panels, in gas systems, 3:380
- Alarm panels/monitoring, of vacuum systems, 3:383
- Alarms
anesthesia machine, 1:36–37
in blood gas analysis, 6:526–527
as a hospital problem, 6:111–112
human factors and, 3:540–541
- Alcaligen eutrophus*, 1:371
- Alcohol wipes, skin electrodes and, 1:136
- Alexa Fluor dyes, 4:467–468
- Alginate, in tissue engineering, 6:384
- Alginate scaffolds, 1:367, 370f
- Algorithms. *See also* Computer algorithms
arrhythmia morphological pattern recognition, 1:73–74
for automatic external defibrillators, 1:80
dual-chamber arrhythmia detection, 1:77–79
feature-based arrhythmia, 1:76–77
for implantable atrial defibrillators, 1:80
for implantable cardioverter defibrillator, 1:69–70, 71
for rate-based arrhythmia detection, 1:71–73
template-based arrhythmia analysis, 1:74–76
- Aliphatic carbon compounds, 1:296
- Alkaline hydroxides, as CO₂ absorbents, 1:32, 33
- Alkoxide precursors, of glass-ceramics, 1:288
- Alliance for Engineering in Medicine and Biology (AEMB), 4:312–313
- Alloderm, as a skin substitute, 6:177
- Allodynia, 6:440
- Allogenic cells, in engineered tissue, 3:192
- Allogenic tissue transplants, 1:355, 366
- Allografts, 3:443–444
regenerative, 1:290
- Allotropes, of carbon, 1:296–298, 300–301
- Alloy refining, of nickel–titanium shape memory alloys, 1:5
- Alloys
as prosthetic restorative materials, 1:325–326
shape memory, 1:1–12
- Alpha Eta Mu Beta, 4:316–317
- Alpha waves, EEG, 3:66
- Alternate image plane display, in computed tomography, 2:239
- Alternating current (AC), in cardiopulmonary resuscitation, 2:36–37. *See also* AC entries
- Alternating vision contact lenses, 2:327
- Alumina. *See also* Aluminum oxide
in dental implants, 1:328–329
in orthopedic prostheses, 1:317–318
reticulated, 1:355–356, 357, 358, 359, 360
- Aluminum oxide, as biomaterial, 1:108, 272
- Aluminum–titanium catalysts, 1:106
- Alveolar air/ventilation, CPR and, 2:40
- Alveolar minute ventilation, 6:104–105
- Alveolar ventilation, 2:42
- Alza Macroflux technology, 2:503
- Alzheimer's disease (AD)
cognitive training for persons with, 6:73–74
multidimensional study of, 1:396–397
- Amalgam, dental, 1:322–323
- Ambient light, pulse oximetry and, 5:212
- Ambient vibration, power generation with, 4:431–433
- Ambulatory blood glucose monitoring, 1:16–17
- Ambulatory blood pressure monitoring, 1:13–16, 489
basic techniques of, 1:14
clinical concepts related to, 1:14–15
indications of, 1:15
limitations of, 1:16
- Ambulatory blood pressure profile, interpretation of, 1:15
- Ambulatory event monitoring, 1:134
- Ambulatory glucose monitoring, significance of, 1:17. *See also* Ambulatory blood glucose monitoring
- Ambulatory monitoring, 1:12–18
with a Holter device, 1:12–13
- Ambulatory pump, drug infusion systems, 2:501–502
- Ambulatory tonometer, 6:407
- American Academy of Environmental Engineers, 4:317
- American Academy of Orthopaedic Surgeons, 4:317
- American Academy of Orthotists and Prosthetists, 4:317
- American Academy of Sleep Medicine guidelines, 6:212
- American Association of Engineering Societies, 4:317
- American Association of Physicists in Medicine, 4:317
- American Chemical Society, 4:317

- American College of Nuclear Physicians, 4:317
- American College of Physicians, 4:317
- American College of Radiology (ACR), 4:317–378; 6:560
- American College of Surgeons, 4:318
- American Congress of Rehabilitation Medicine, 4:318
- American Heart Association (AHA), 2:48, 49, 52
- American Institute for Medical and Biological Engineering (AIMBE), 4:314–316, 317
- American Institute of Biological Sciences, 4:318
- American Institute of Chemical Engineers, 4:318
- American Institute of Physics, 4:318
- American Institute of Ultrasound in Medicine, 4:318
- American Medical Informatics Association, 4:318
- American National Standards Institute (ANSI), 1:158
- anesthesia machine standards of, 1:31
- audiometry standards by, 1:92, 99
- American Physical Therapy Association (APTA), 1:168
- American Society for Artificial Internal Organs, 4:318
- American Society for Engineering Education, 4:318
- American Society for Healthcare Engineering of the American Hospital Association, 4:318
- American Society for Laser Medicine and Surgery, 4:319
- American Society for Testing and Materials (ASTM) designations, 1:104, 105
- American Society of Agricultural and Biological Engineers, 4:319
- American Society of Civil Engineers, 4:319
- American Society of Clinical Pathologists (ASCP), 1:455
- American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc., 4:319
- American Society of Mechanical Engineers, 4:319
- American Society of Neuroradiology, 4:319
- American Society of Safety Engineers, 4:319
- American Thoracic Society (ATS) standards, 5:369
- Amino acids, boron-containing, 1:573–574
- Amnesia, electroconvulsive therapy and, 3:56
- Amnioexchange, 4:176–178
- Amniopatch, 4:178
- Amplification, in linear variable differential transformers, 4:253
- Amplifier characteristics, EMG, 3:106–107
- Amplifier–recorder, in anorectal manometry, 1:63
- Amplifiers
- ECG, 1:175
- EEG, 1:172
- electrodermal biofeedback, 1:173
- EMG, 1:169
- heart rate, 1:175
- respiration, 1:174–175
- temperature, 1:170
- for use with strain gages, 6:286
- Amplitude and frequency analysis, in arrhythmia analysis, 1:76
- Amplitude distribution, in EEG analysis, 3:69
- Amplitude-integrated EEG (aEEG) monitor, 5:36
- Amplitude variability analysis (AVA), for automatic external defibrillators, 1:80
- Amsler grid, 6:529
- Amundson porous electrode, 1:153–154
- Anal pressure, resting, 1:63–64
- Anal sensation, assessment of, 1:65
- Analgesia, 3:25. *See also* Electroanalgesia
- Analog display feedback schedule, 1:176
- Analog-to-digital conversion (ADC), 5:112
- Analogue data, 1:396
- Analogue recorders, 6:49–51, 55–59
- Analysis of variance (ANOVA), 1:224
- Analyte 2000™ fiber-optic fluoroimmunoassay system, 4:378
- Analytical cytometry, of whole blood, 1:460
- Analytical methods, automated, 1:18–28
- Analytical reconstruction techniques, 2:247
- Analytic X-ray units, regulations related to, 2:175
- Analyzers. *See also* Oxygen analyzers
- automated, 1:19
- chemistry, 1:19–23
- critical care, 1:21–22
- general chemistry, 1:20–21
- immunoassay, 1:21
- Anaphylotoxins, 1:281
- Anatomical joint models, 4:216
- Anatomy, articular, 4:199–201
- Ancient DNA, 5:384–385
- Anderson loop circuit, 6:285–286
- Anemometry, thermal, 3:329–330
- Anesthesia
- automatic feedback control for, 1:49
- brain monitoring in, 4:558–560
- computer information technology in, 1:44–47
- computers in, 1:42–51
- diagnosis and decision assistance for, 1:50
- electrical, 3:27–30
- electroconvulsive therapy and, 3:57
- electropharmaceutical, 3:32
- future of computer technology in, 1:50–51
- as a procedure, 4:556
- as a process, 4:556–557
- system identification and adaptive control for, 1:49–50
- telemedicine in, 1:47
- typical process of delivering, 1:29
- Anesthesia control, computer-aided, 1:47–50
- Anesthesia control systems, computer-assisted, 1:48
- Anesthesia delivery systems, 1:29–31
- Anesthesia depth monitors, 1:44
- Anesthesia equipment, 1:43f. *See also* Anesthesia machines
- Anesthesia information system (AIS), 1:44–45
- Anesthesia machines, 1:28–42
- circle breathing circuit, 1:32–33
- components of, 1:34–37
- limitations of, 1:37–40
- new technologies in, 1:40–41
- overview of, 1:29–31
- Anesthesia monitoring, 4:555–565
- adequacy of, 4:557–563
- computers in, 1:43–44
- outside the operating theater, 4:563–564
- Anesthesia monitors, 1:43f; 5:35
- recently developed methods in, 4:560–562
- Anesthesia pump drug infusion systems, 2:501
- Anesthesia resident training, 1:45–46
- Anesthesia ventilator, 1:29
- in anesthesia delivery, 1:30–31
- Anesthesia ventilator technology, improvements in, 1:41–42
- Anesthesia workstations, 1:30f
- Anesthetic agents, delivering, 1:29
- Anesthetic depth
- indicators and measures of, 4:557–558
- neurophysiological variables with, 4:558
- parameters proposed for monitoring, 4:562–563
- Anesthetics, inhalational, 1:28; 3:377–378
- Aneurysms, vascular graft prostheses and, 6:493
- Anger, Hal O., 1:51–52
- Anger camera, 1:51–61
- collimation in, 1:56–58
- components of, 1:54f, 56f
- image generation in, 1:53f, 55–56, 57f
- improvements to, 1:58–60
- performance of, 1:60
- system description for, 1:52–58
- Angina, 2:50–51
- Angiodysplasia, 3:392
- Angiographic CAD, multivariate techniques to predict, 3:257–258
- Angiographic disease, predicting, 3:256
- Angiographic laboratories, 6:559
- Angiographic severity, analysis of studies predicting, 3:256
- Angiography. *See also* Digital angiography
- digital subtraction, 2:422–425
- non-catheter/noninvasive, 2:425–426
- in radiosurgery, 5:578
- Angiological thermatomes, 6:348–349
- Angioplasty, coronary, 2:349–360. *See also* Coronary angioplasty; Guidewire diagnostics
- Angioplasty catheters, diagnostics with, 2:354–355
- Animal gamma cameras, 5:104–105
- Animal imaging devices, 5:104–106
- hybrid, 5:106

- Animal models, of spine stabilization procedures, 3:581–582
- Animal PET imagers, 5:105–106
- Ankle Brachial Index (ABI) Test, 5:242
- Ankle joint, stability of, 4:221–222
- ANN-based blood pressure controller, 1:497
- Annual Conference on Engineering in Medicine and Biology (ACEMB), 4:312
- Anode, 1:465–466
in X-ray tubes, 6:601–603
- Anode cooling rate, in X-ray tubes, 6:607
- Anode cracking, in X-ray tubes, 6:608
- Anode heat storage capacity, in X-ray tubes, 6:607
- Anodic processes, in biomaterial degradation, 1:308–309
- Anodization, for porous biomaterial fabrication, 5:401
- Anorectal manometry, 1:62–69
clinical utility of and problems with, 1:66–67
equipment for, 1:62–63
instructions for patients undergoing, 1:63
medical terms associated with, 1:67–68
prolonged, 1:66
selecting appropriate test/maneuver for, 1:66–68
study protocol for, 1:63–65
- Anorectum
anatomy and physiology of, 1:62
structures of, 1:64f
- ANOVA (analysis of variance)
one-way, 6:254–256
two-way, 6:256
- Antenna arrays, in microwave ablation, 6:365
- Anterior cruciate ligament (ACL), reconstruction of, 1:256
- Anterior interbody cages, 6:235
- Anterior lumbar interbody fusion (ALIF) surgery, 6:235
- Anterior plating fusion techniques, 3:575–577
- Anterior spinal instrumentation, 3:579–580
- Antibiotics, biofilms and, 1:115
- Antibiotic susceptibility testing, 4:375
- Antibodies
in enzyme immunoassays, 1:21
fragmentation of, 4:599
heterogeneity of, 4:599
immune system and, 4:597–599
implants and, 1:112
structure and function of, 4:598–599
- Anticancer therapy, cyclodextrins in, 2:459
- Antigen recognition, 4:597–598
- Antigens, tumor-specific, 4:605
- Antimatter particles, 6:3–4
- Antimicrobial biomaterials, 1:118
- Antioxidants, lithotripsy-induced free radical injury and, 4:264
- Antiproliferative agents, 4:275
- Antiscatter grid, in mammography, 4:301–302
- Antitumor reactivity, of T cell subsets, 4:114–115
- Antitumor responses, eliciting, 4:606–607
- Anus. *See* Anorectal manometry; Anorectum
- Anxiety disorders, biofeedback clinical outcome literature related to, 1:178
- Aorta, measuring elasticity of, 1:86–87, 89, 90
- Aortic valves, porcine, 3:413
- Aortoenteric fistula, 3:390
- Aoyagi, Takuo, 1:471
- Apatite(s), 1:523–524
biological, 1:285
coral-derived, 1:374
elastic properties of, 1:527t
- Apatite-wollastonite (A/W) glass-ceramics, 1:288
- Apertured NSOM, 4:439–441
- Aperture impedance, signal detection in, 2:413
- Aperture impedance WBC counting technique, 2:411
- Apertureless NSOM, 4:441–443
- Aphasia, 2:211
- Apligraf, as a skin substitute, 6:176
- Apnea detection. *See* Ventilatory monitoring
- Apparatus. *See also* Equipment; Instrumentation
cryosurgery, 2:363–366
for electrophoresis, 3:138
intraoperative radiotherapy, 6:13–15
manual muscle testing, 6:65
- Applanation tonometry (AT), 5:236–237
- Application-based neurological monitors, 5:35
- Application Service Provider (ASP), 5:350
teleradiology model, 6:308
- Application Specific Integrated Circuits (ASICs), 1:420
- Applicator cooling, in tissue ablation, 6:366
- Applicator dependent perturbations, 1:609–611
- Apraxia, 2:210
- Aqueous electrolyte sensor, oxygen analyzers, 5:202–204
- Arbo-prep[®] cream, 1:136
- Archival strategies, 5:348–349
- Ariel Computerized Exercise System, 1:398–399
- Ariel Performance Analysis System, 1:394, 397
- Armature mass, in linear variable differential transformers, 4:255
- ARMAX (autoregressive moving average with inputs) model, 1:492, 493, 494
- Arm ergometry, 3:251
- Aromatic carbon compounds, 1:296
- Array design, in ultrasound imaging, 6:457–458
- Arraying robots, 4:367–369
- Array platforms, hybridization and fluidics stations in, 4:369
- Arrays, DNA, 2:433
- Arrhenius rate process, burns modeled as, 1:192
- Arrhythmia, 1:69. *See also* Automated arrhythmia analysis; Defibrillators; Pacemakers
terminal, 2:47
unipolar electrograms of, 1:69f
- Arrhythmia detection
dual-chamber, 1:77–79
rate-based analysis for, 1:71–73
- Arrhythmia monitoring, 4:568
- Arterial blood
oxygenation of, 6:518–519
oxygen in, 5:210–212, 212–213
- Arterial blood pressure, 1:490–491; 4:568–569; 6:402. *See also* Arterial tonometry
measurement of, 5:235
- Arterial elasticity
mathematical expression of, 1:87–88
measurement of, 1:85–87
pressure–diameter relations for, 1:87–88
strain energy equations for, 1:88
- Arterial hemodynamics, assessing, 3:490–492
- Arterial input function (AIF), 6:435–436
methods for estimating, 6:435–436
- Arterial sample collection, 2:15
- Arterial system, 3:483–492
as a windkessel model, 3:485–486
- Arterial tonometry, 1:488–489; 6:402–410
applications of, 6:407–408
considerations related to, 6:406–407
measurement accuracy in, 6:408
operational principles of, 6:402–404
- Arterial tree, wave propagation and reflection in, 3:487–490
- Arterial-venous oxygen difference (AVO₂ difference), 2:21
- Arterial walls
structure and basic characteristics of, 1:85
uniaxial tensile behavior of, 1:87
- Arteries
elastic properties of, 1:85–91
elastic properties of diseased, 1:89–91
elastic properties of normal, 1:88–89
- Arteriovenous oxygen difference, 2:13, 14
- Arthroplasty
interposition, 3:514
low friction, 3:515
- Arthroscopy, 3:181
- Articular anatomy, 4:199–201
- Articular cartilage, 4:202–204
biomaterial scaffolds necessary for engineering, 2:74
cells capable of generating, 2:73
composition and structure of, 2:63–65
functional tissue engineering of, 2:74–75
tissue engineering of, 2:73–75
- Articular chondrocytes, in cartilage regeneration, 2:73
- Articular disk, 6:415–417
- Artifacts
in computed tomography, 2:253–257
EEG, 3:68
imaging using deconvolution, 4:522

- misinterpreted, 6:577
- Artificial blood, 1:511–523
hemoglobin solutions, 1:516
perfluorochemical emulsions, 1:513–515
- Artificial eye, 1:453
- Artificial heart, 3:449–459
clinical need for, 3:449
design considerations for, 3:450–451
electric, 3:457–458
future directions of research on, 3:458
pneumatic, 3:457
- Artificial hip joint bearing surfaces, tribology of, 3:517–518
- Artificial hip joints, 3:514–525
applications of, 3:518–521
bearing surfaces of, 3:515–516
future developments in, 3:521–522
historical development of, 3:514–515
nomenclature related to, 3:522–523
tribological and biological methodology for, 3:516–518
- Artificial intelligence, in exercise training, 1:393
- Artificial intervertebral disk, 3:587–588
- Artificial larynx, 4:230–231
- Artificial neural networks (ANN), 1:497
- Artificial pancreas, 5:224–234
clinical studies of, 5:230, 231t
components of, 5:226
cost of, 5:230
historical background of, 5:225
indications for, 5:230
outlook for, 5:230
physiological considerations related to, 5:225–226
prototypes of, 5:227–230
types of, 5:221
- Artisan lenses, optical quality of, 4:237–238
- Aseptic loosening, 1:314
- As low as reasonably achievable (ALARA) principle, 6:472, 573
- Aspiration, through micropipettes, 1:507
- Assessment, of communication disorders, 2:214–217
- Assisted circulation, 3:459–461
- Assisted reproductive technologies, microfluidic systems as, 4:394
- Assistive devices
benefits and limitations of, 2:223–224
for communication disorders, 2:221–224
hearing, 2:222–223
language, 2:221–222
new directions in, 2:225
- Assistive technologies. *See also* Mandated web accessibility
for the blind and visually impaired, 1:443–455
future possibilities for, 1:453–454
GPS navigational aids, 1:451–453
independent living aids, 1:447–448
indoor navigational aids, 1:453
intelligent electronic travel aids, 1:450–451
low vision reading aids, 1:444–445
mobility aids, 1:448, 449t
- Association for Applied Psychophysiology and Biofeedback (AAPB), 1:168, 172
- Association for Computing Machinery, 4:319
- Association for the Advancement of Medical Instrumentation (AAMI), 1:489; 4:319
electrode standards, 1:158–161
- Association of Biomedical Communications Directors, 4:319
- Association of Environmental Engineering and Science Professors, 4:319
- Asthma, 3:507
- ASTM F-136 alloy, 1:105
- Asymmetry analysis, 6:351–352
- Asystole, 2:47
- Atherosclerosis
laser Doppler flowmetry for, 2:382
mechanical properties of arteries with, 1:90–91
vascular graft prostheses and, 6:491–493
- Atherosclerotic disease, 2:50–51
- Athletic performance, factors underlying, 1:392
- Atomic absorption spectrometry
equipment for, 3:319–321
medical applications of, 3:321
theoretical basis for, 3:318–319
- Atomic attenuation coefficient, 6:593
- Atomic emission detector (AED), 4:323
- Atomic force microscope (AFM), 4:503–504.
See also Atomic force microscopy (AFM)
imaging, 4:504–505
theory and experimental approach of, 4:506–513
- Atomic force microscopy (AFM)
for characterizing surfaces, 1:351–352
contact mode (static mode) in, 4:507–508
DNA sequencing with, 2:435
vibration mode (dynamic mode) in, 4:508–510
- Atraumatic Restorative Treatment (ART), 1:324
- Atrial defibrillators, implantable, 1:80
- Atrial fibrillation (AF), 1:69; 6:370
- Atrial flutter (AFI), 1:69
- Atrial sensing, 5:221
- Atrioventricular valves, 2:43
- Attention assessment/training, virtual reality for, 6:74–75
- Attention deficit disorder (ADD), EEG biofeedback treatment for, 1:172
- Attention deficit hyperactivity disorder (ADHD), 6:75
- Attenuation
acoustic, 6:455
photon, 6:591–593
- Audio conferencing, in office automation systems, 5:158
- Audiograms, 1:95–96
- Audiologists, 1:91–92
- Audiology, 1:91
- Audiometers, 1:92
“speech mode” setting of, 1:96
types of, 1:92–93
- Audiometric calibration, 1:93
- Audiometric standards, 1:93
- Audiometric threshold prediction/estimation, otoacoustic emissions and, 1:102
- Audiometry, 1:91–103
acoustic immittance measurement, 1:99–101
defined, 1:91–92
electrophysiologic, 1:97–99
origins of, 1:92–93
otoacoustic emissions, 1:101–102
psychophysical methods of, 1:93–94
pure tone, 1:93–96, 97
speech, 1:96–97
- Audio recordings, for communication disorders, 2:213
- Audio-visual system, in high-dose-rate remote afterloaders, 1:595
- Auditory brainstem response (ABR), 1:98–99
- Auditory evoked fields, biomagnetic measurements and, 1:244
- Auditory evoked potentials, 1:97–98
auditory threshold estimation/prediction with, 1:98–99
classification of, 1:98
- Auditory feedback, from EEG biofeedback instrumentation, 1:171
- Auditory implants. *See* Cochlear prostheses
- Auditory mechanism, 1:94f
- Auditory periphery, 2:134–135
- Auditory processing disorder, 2:211
- Auditory steady-state response (ASSR), 1:99
- Auditory system, 2:133–134
sound pathways of, 1:94–95
- Auditory threshold, estimation/prediction of, 1:98–99
- Augmentative and alternative communication (AAC) systems, 2:202–210
assessment of, 2:207–208
characteristics of, 2:203–207
needs served by, 2:202–203
privacy issues related to, 2:210
training individuals for, 2:208–209
vocabulary selection for, 2:209–210
- Augmentative communication system.
See Communication devices
- Auscultation, 1:14
- Auscultatory technique, 1:486
- Austria, infrared imaging in, 6:353
- Authenticity, in teleradiology, 6:307–308
- Autism spectrum disorders, 2:211
“AutoAnalyzer,” 1:23
- Autobalance oxygen analyzers, 5:201
- Autoclaving, 6:275
- AutoCrane positioning device, 6:398
- Autograft, 1:283
cultured epithelial, 6:190–191
- Autografting, for organ function loss, 6:182
- Autologous cells, in engineered tissue, 3:192
- Autologous tissue transplants, 1:355, 366

- Automated analytical methods, 1:18–28.
See also Analyzers; Automated analyzers
 clinical laboratory, 1:23–24
 patient preparation, specimen collection, and handling in, 1:19
- Automated analyzers, 1:19. *See also* Analyzers
 manufacturers of, 1:20t
- Automated arrhythmia analysis, 1:69–84
 early work in, 1:70–71
 feature-based algorithms in, 1:76–77
 template-based algorithms for, 1:74–76
- Automated arrhythmia detection, devices that use, 1:79–80
- Automated blood cell counters, 2:88–89
- Automated checkout feature, anesthesia machines, 1:40, 42
- Automated cytology, 2:388–405
 cells in, 2:389–390
 cellular parameter measurement in, 2:392–400
 current usage of, 2:403–405
 cytochemical probes in, 2:390–392
 data acquisition, processing, and modeling in, 2:400–403
 devices used in, 2:397–400
 future prospects for, 2:405
- Automated differential counts, 2:410–421
 measurable properties of white blood cells, 2:410–411
 measurement techniques in, 2:411–412
 measurement technique versus result accuracy and laboratory efficiency, 2:419–420
 sample preparation in, 2:413
 sample stability in, 2:412–413
 signal detection in, 2:413–414
 signal generation in, 2:413
- Automated external defibrillators (AEDs), 2:406–408
- Automated hematology analyzer, 2:410
- Automated perimetry, 6:529
- Automated syringe-filling drug infusion system, 2:502
- Automatic exposure control (AEC), 6:569
 in X-ray equipment, 6:558
- Automatic external defibrillators (AEDs), 1:79–80; 2:56–57
 standards for, 1:160–161
- Automatic gain control, in neonatal respiration monitoring, 5:21
- Automatic pattern recognition WBC counting technique, 2:412
- Automation, human factors and, 3:541
- Automation systems, office, 5:149–160
- Autoregressive (AR) model, 3:71
 power spectrum estimation using, 3:77–79
- Autoregressive moving average (ARMA) model, 3:71–72
- Autoregulation, in whole-body models, 5:305–306
- Auto-titration, in continuous positive airway pressure, 2:333–334
- Average corrosion rate, 1:311
- Axial CT scan, 2:237–238
- Axial flow ventricular assist devices, 3:455–456
- Axial rotation
 of the middle and lower cervical spine, 3:556–557
 of the occipital-atlantoaxial complex, 3:554–555
- Axon guidance, in neurons, 1:414
- BAC clones, in genome analysis, 1:222
- Back-etch-silicon-on-insulator (BESOI) structures, 2:4
- Back focal plane interferometry (BFPI), 5:178, 179
- Backing materials, for electrodes, 1:140–141
- Back pain
 biofeedback clinical outcome literature related to, 1:179
 non-fusion solutions for, 6:237
- Backprojection operation, 2:249
- Backscatter factor, 2:130
- Back surface topography, in scoliosis, 6:127–128
- Bacteremia, implant-related, 1:116
- Bacteria
 in biofilms, 1:113, 115
 electron microscopic diagnostic criteria for, 4:484
 Gram-negative, 1:319–320, 371
 Gram-positive, 1:319–320
- Bacterial artificial chromosomes (BACs), 1:222
- Bacterial detection systems. *See* Microbial detection systems
- Bacterial infection
 biological response to, 1:113
 biomaterial surfaces and, 1:113–118
- Bag-ventilator selector switch, 1:33
- Bakken, Earl, 1:432
- Ball-and-socket bearing wear geometry, 1:316
- Ball milling, nanoparticle fabrication via, 5:2
- Balloon catheter, 2:350–352
- Balloon kyphoplasty, 6:234
- Balloon pump, intraaortic, 4:162–171
- Bandwidths, in biofeedback, 1:169, 170
- BANG gel dosimeter, 5:491, 492f
- Banked blood, risks of, 1:512–513
- Bar-coded drug infusion system, 2:503
- Barotrauma. *See* Hyperbaric medicine
- “Barrel Method,” 2:35–36
- Basal layer of epidermis, 1:131
- Baseline temperature, 2:29
- Baseline wander, in electrodes, 1:159
- Basement membrane, 6:180
 morphology of, 6:188
- Basic data method, of formulating tissue substitutes, 5:256
- Basic Local Alignment Search Tool (BLAST), 1:219, 222
- Basketball, wheelchair, 4:548–549
- Basophils, 2:82, 411
- Batch sensor fabrication, 2:1
- Baths, paraffin wax, 3:467–468
- Batteries
 in high-dose-rate remote afterloaders, 1:596
 micromachined, 4:430
- Battery-powered pacemaker, 1:151
- Bayer ADVIA 70 Hematology System, 2:414–415
- Bayer ADVIA 2120 Hematology System, 2:416–418
- Bayer ADVIA Centaur immunoassay analyzer, 1:21
- Bayesian approaches, in kinetic parameter estimation, 6:434–435
- Bayesian inference, 1:240–241
- Bayesian networks, 1:226–227
- Beam alignment devices, in intraoperative radiotherapy, 6:22–23
- Beam design, in a computed tomography simulator, 2:271–273
- Beam determination, in three-dimensional conformal radiotherapy, 6:33–34
- Beam filtration, in CT scanners, 2:235
- Beamformers, 1:241
- Beam gantries, 6:12
- Beam hardening artifact, in computed tomography, 2:255–256
- Beam optimization, as a neutron source for BNCT, 1:578–579
- Beam quality specifiers, 5:475–476
- Bearing rotation, in X-ray equipment, 6:563
- Bearing wear, in X-ray tubes, 6:608
- Beckman Coulter IDS system, 1:24, 25f
- Beckman Coulter LH 750 Hematology System, 2:418
- Bedside hemodynamic monitoring, 4:567–568
- Behavior, reinforcement of, 1:166–167
- Bekesy audiometric tracing, 1:94f
- Bellows assembly, anesthesia machine, 1:38
- Benefit assessment, in radiotherapy treatment planning optimization, 6:38–39
- Benign prostatic hypertrophy, treatment of, 4:542
- Benzoyl peroxide (BPO), 1:541–542
- Berlin Heart Excior, 3:452
- BetaCath delivery device, 1:605
- β -hemolytic streptococci, 1:114
- Beta waves, EEG, 3:66–67
- Bias current tolerance standards, 1:159
- Bicarbonate, in blood CO₂ transport, 1:468
- BiCMOS receiver circuitry, 1:425, 426
- Bicycle ergometer, 3:251
- Bidirectional intracardiac shunt, 2:16–17
- Bilateral cochlear implants, 2:139
- Bilayers, 1:348
- Bilevel positive airway pressure, 2:332
- Biliary complications, after liver transplantation, 4:272
- Bilirubin measurement, optical sensors in, 5:171–172
- Bilirubin monitoring, neonatal, 5:28–29
- Bimetallic corrosion, biomaterial failure from, 1:278
- Bimetallic thermometers, 6:313
- Bin area method (BAM), 1:74–75

- Binary display feedback schedule, 1:176
- Binary leaf collimator, 6:397
- Binding energy, X-ray photon spectroscopy and, 1:349
- Binomial distribution, 6:244–245
- Binomial variables, binomial test for, 6:249–250
- Bioactive ceramic materials, 1:272
use in medical devices, 1:314
- Bioactive fixation, 1:285, 289
- Bioactive glass foams, sol–gel derived, 1:292–293
- Bioactive glasses, 1:284, 285
bioactivity mechanism of, 1:285–286
melt-derived, 1:287
porous melt-derived, 1:291–292
sol–gel-derived, 1:287–289
- Bioactive materials, 1:104, 284
Class A and B, 1:289
genetic control by, 1:290
- Bioactive photoresist, 1:411
- Bioactive skin substitutes, role of, 6:172.
See also Skin substitutes
- Bioactivity, of glasses, 1:286–287
- Bioactivity test, of biomaterials, 1:360–362
- Bioartificial livers, 4:393
- Bio-barcode technology, nanoparticle-based, 4:379–380
- Biobrane, as a skin substitute, 6:175
- Biorburden, 6:274
- Bioceramics, 1:283–296, 355
challenge for, 1:284–285
as medical devices, 1:284
nearly inert, 1:284
in regenerative medicine, 1:290
resorbable, 1:285
- Bioceramic scaffolds, 1:374–375
types of, 1:291
- Biochemical measurement, neonatal, 4:590–591
- Biochemical precursors, 1:574–575
- Biochemical probes, with DNA specificity, 2:391t
- Biocompatibility, 1:8, 130
of biomaterials, 1:281–282
of carbon biomaterials, 1:301
of engineered tissue, 3:201
of implants, 1:256
of materials, 1:104–120
of nanoporous membranes, 2:451
in orthopedic devices, 5:191–192
surface modification of scaffolds for, 1:379
- Biocompatibility assessment, of polymeric biomaterials, 1:341
- Biocomposites, 1:289
- Biodegradable polymeric drug delivery systems, 2:504
- Biodegradable polyurethane, in tissue engineering, 1:373–374
- Biodegradable synthetic polymers, 1:339–340
- Bioeffects, ultrasound-related, 6:471–473
- Bioelectric potential (BEP), 1:559–560
- Bioelectrodes, 1:120–166
designing, 1:120–121, 137–158
electrical properties of electrode–skin interface, 1:122–137
history of, 1:137–139
modern designs for, 1:146–149
skin and, 1:131–137
skin preparation for, 1:136–137
standards for, 1:158–162
- Bioengineering, defined, 4:312
- Bioengineering Definition Council (NIH), 1:403
- Biofeedback, 1:166–188
applied clinical examples of, 1:176–178
cardiopulmonary, 1:174–176
clinical outcome literature related to, 1:178–183
defined, 1:166
electrode placement in, 1:177
future directions of, 1:183–185
operant conditioning and, 1:167
sensitivity (gain) of, 1:177
- Biofeedback Certification Institute of America (BCIA), 1:168
- Biofeedback instrumentation
EEG, 1:171–173
electrodermal, 1:173–174
temperature, 1:170–171
- Biofeedback modalities/instrumentation, 1:168–169
- Biofeedback professionals, credentials for, 1:168
- Biofeedback schedules, in clinical applications, 1:176–177
- Biofeedback sessions, length and outline of, 1:177–178
- Biofeedback training, 1:166
theories underlying, 1:167–168
- Biofilm, as a feature of implant-related infection, 1:113, 115
- Bioglass, 1:314
- Bio-heat equation, 6:348
- Bioheat transfer, 1:188–197
blood flow and, 1:191
effects of blood perfusion on, 1:189–190
subzero effects and, 1:192
therapeutic applications of, 1:190
thermal conductivity and thermal diffusivity measurements, 1:192–196
thermal injury and, 1:192
thermoregulation and, 1:190–192
- Bioimpedance
characteristics of, 4:124–126
in cardiovascular medicine, 1:197–216
intrathoracic, 1:204–208
measured, 1:199
measurement of, 4:122–124
reactive applications of, 1:208–213
resistive applications of, 1:199–208
transthoracic, 1:199–204
- Bioimpedance theory, 1:198–199
- Bioinert ceramics, 1:272
- Bioinformatics, 1:216–230
computational modeling and biological network analysis, 1:224–227
genome analysis, 1:221–223
microarray analysis, 1:223–224, 225f
phylogenetic trees and, 1:220
protein folding, simulation, and structure prediction via, 1:220–221
sequence alignment in, 1:217–220
- Biological apatite, 1:285. *See also* Hydroxyapatite (HA)
- Biological applications, porous materials for, 5:392–406
- Biological assay multicolor optical coding, nanoparticles in, 5:6
- Biological drugs, cyclodextrins as carrier for, 2:459–460
- Biological effectiveness factors, in boron neutron capture therapy, 1:572
- Biological effects
of systemic hyperthermia, 4:46–51
of ultraviolet radiation, 6:474–476
- Biological interactions, biomaterial failure related to, 1:279–281
- Biological network analysis, 1:224–227
- Biological polymers, 1:329–330
- Biological-probabilistic approaches, to treatment plan optimization, 6:44–46
- Biological responses
to biomaterials, 1:108–113
microbioreactors for understanding, 4:394–395
to sol–gel-derived bioactive glasses, 1:293–294
- Biological samples
chemistry, conformation, and conductivity of, 4:519–521
identification of compounds in, 3:345
- Biological tags, nanoparticles as, 5:5
- Biological tissues, propagation of ultrasound in, 4:66–67
- Biologic devices, sterilization employed on, 6:278t
- Biologic scaffold materials. *See also* Biomaterial scaffolds
ethylene oxide sterilization of, 6:276–277
heat sterilization of, 6:275–276
ionizing radiation sterilization of, 6:277–278
sterilization of, 6:273–282
- Biology, scanning tunneling microscopy in, 4:517–519
- Biomagnetic instrumentation, 1:231–242
- Biomagnetic liver susceptometry (BLS), 1:247
- Biomagnetic measurements, applications of, 1:242–248
- Biomagnetism, 1:230–255
future directions of, 1:248–249
- Biomagnetometer signal interpretation, 1:238–242
- Biomagnetometer systems, 1:236–238
- Biomarkers, in exercise stress testing, 3:255–256
- Biomaterials, 1:104, 267–283. *See also* Absorbable biomaterials; Bioceramics
antimicrobial, 1:118
biocompatibility of, 1:281–282
biological response to, 1:108–113
in bone tissue engineering, 6:390–391
carbon, 1:296–308
in cardiovascular tissue engineering, 6:392–393, 393–394

- Biomaterials (*Continued*)
 chemical bonding of bone to, 1:109–110
 classification of, 1:269
 common uses for, 1:269t
 composite, 1:108
 corrosion and wear of, 1:308–322
 covalent modification of, 6:388–389
 defined, 1:267
 for dentistry, 1:322–329
 factors in failure of, 1:278–281
 future directions for, 1:282
 government regulation of, 1:268–270
 healthcare treatments requiring, 1:270t
 history of, 1:267–268
 journals related to, 1:269t
 market size and applications related to, 1:268
 normal local tissue response to, 1:108–110
 polymeric, 1:329–342
 in restorative dentistry, 1:313
 standards for, 1:281–282
 structural properties of, 1:362–364
 for tissue engineering, 1:367–375; 6:383–387
 types of, 1:270–278
 in vascular graft prostheses, 6:498
- Biomaterials Access Assurance Act, 4:315
- Biomaterials scaffolds, for engineering
 articular cartilage and meniscus, 2:74.
See also Biologic scaffold materials
- Biomaterials testing, 1:354–365
 materials and methods related to, 1:355–357
- Biomaterial surfaces, 1:341, 342–354
 adsorption of proteins at, 1:344–345
 ambient techniques for characterizing, 1:351–352
 analysis of, 1:348–351
 cell behavior at, 1:345
 modification of, 1:345–348
 properties of, 1:343–345
- Biomaterial–tissue interface, degeneration of, 1:110–111
- Biomechanical models, of mastication, 6:417–424
- Biomechanical research, strain gages in, 6:287
- Biomechanics
 of diabetic skin ulceration, 6:206–207
 of engineered tissue, 3:200–201
 of exercise fitness, 1:384–403
 as a hospital problem, 6:114
 occipital–atlantoaxial complex, 3:554–555
 of scoliosis, 6:122–137
 of scoliosis progression during growth, 6:129
 skin, 6:202–208
 of spine stabilization procedures, 3:568–591
- Biomedical applications
 capacitive microsensors for, 2:1–12
 of fluorescence measurements, 3:345–346
 of microfluidics, 4:426–427
 of polymers, 1:333–341
- Biomedical devices, surface modification
 used in, 1:346t
- Biomedical engineering
 defined, 4:372
 thermocouples in, 6:345–346
- Biomedical engineering applications,
 polymers in, 5:388–391
- Biomedical engineering associations/
 societies, in the United States,
 4:316–321
- Biomedical engineering education,
 1:403–409
 career preparation in, 1:404–405
 core curriculum subjects in, 1:406t
 course requirements in, 1:405–406
 history of, 1:403–404
 professional skills and, 1:406–408
 undergraduate curriculum in, 1:405–408
- Biomedical Engineering Society (BMES),
 1:403; 4:316, 319–320
- Biomedical engineers, 1:403
- Biomedical equipment maintenance,
 3:223–229
 documentation of, 3:227–228
 environmental rounds in, 3:228
 evaluation of preventive maintenance
 program, 3:225–227
 historical background of, 3:223–224
 inclusion criteria for, 3:224
 inventory in, 3:224
 preventive maintenance procedures in,
 3:225
 selecting the preventive maintenance
 interval, 3:224–225
 staffing requirements for, 3:228
- Biomedical laboratory, data in, 2:308–310
- Biomedical laboratory computer system,
 2:306–321
 architecture of, 2:310–313
 basics of, 2:313–314
 emerging and future developments for,
 2:316–320
 software in, 2:314–316
- Biomedical laboratory research, historical
 origins of, 2:307–308
- Biomedical programs (BMDP), 6:263
- Biomedicine, photomicrography in,
 5:296–297
- Biometric systems, in management of
 medical records, 4:358–359
- Biomimetic materials, 1:277
- Biomolecule manipulation, nanoparticles
 in, 5:6–7
- Biomolecule separation/purification,
 nanoparticles in, 5:5
- Biophysical Society, 4:320
- Biopolymers, 1:329–342
- Biopotentials, in biofeedback
 instrumentation, 1:168
- Bioprosthetic heart valves, 3:413–415,
 429–430, 443–445
 dynamics of, 3:421–422
 versus mechanical heart valves,
 3:446–447
- Biopsy, stereotatic, 6:269
- Bio-Pump, 3:454
- Bioreaction kinetics, 4:387
- Bioreactors
 in bone tissue engineering, 6:391
 in cardiovascular tissue engineering,
 6:393, 394
 design of, 6:388
 in tissue engineering, 6:387–388
- Bioreactor technology, engineered tissue
 and, 3:200
- Bioresist, 1:411
- Bioresorbable ceramics, 1:272
- Biosensors, 4:376
- Biosensor techniques, impedimetric, 4:143
- Biosignal monitoring electrodes
 external, 1:137–139
 implanted and external, 1:130
 standards for, 1:158–160
- Biostator artificial pancreas, 5:227
- Biosurface engineering, 1:409–417
 patterning methods in, 1:409–414
- Biosusceptometers, 1:238
- Biosusceptometry, 1:231
 biomagnetic measurements and, 1:247
- Biotelemetry, 1:417–429
- Biotelemetry systems, 1:418–423
 diagnostic applications of, 1:423
 interface electronics in, 1:418
 packaging and encapsulation of,
 1:422–423
 power sources for, 1:422
 rehabilitative, 1:424–427
 therapeutic applications for, 1:424
 transducers in, 1:418
 wireless communication in, 1:418–422
- Bioterrorism, monoclonal antibodies and,
 4:607
- Biotribology, 1:314
- Biphasic theory, 2:69
- Biphasic waveforms, 1:128
- Bipolar electrodes, 1:151
 configuration of, 1:198–199; 3:106
- Bipolar forceps electrode field, in
 electrosurgery, 3:175–176
- Bipolar lead systems, in exercise stress
 testing, 3:248
- Bipolar recording configuration, 3:114–115
- Bipolar recordings, with EEG biofeedback
 instrumentation, 1:171
- Bipolar umbilical cord occlusion, 4:179
- Birdcage electrode design, 1:154
- Bispectral index (BIS), 2:506
 neurological monitor, 5:38–39
- Bispectral Index Score (BIS), 4:560
- Bispectrum analysis, EEG, 3:71
- Bisphenol A polycarbonate, 1:338
- “Bisping” transvenous screw-in electrode,
 1:152
- Bite block, 5:590
- Black body, 6:361
- Bladder, overactive, 1:441
- Bladder applications, for porous
 biomaterials, 5:403
- Bladder dysfunction, neurostimulation of,
 1:429–443
- Bladder dysfunction hardware, future of,
 1:441–442
- Bladder function, visceral neural signals in
 control of, 3:128–129

- Bladder stimulation, 1:430
- Bleeding. *See also* Blood
 lower GI, 3:391–392
 upper GI, 3:385–390
- Blindness, 6:532–533
 defined, 1:443–444
 prevalence of, 1:444t
- Blind persons
 assistive technology for, 1:443–455
 readings aids for, 1:445–447
- Blind source separation, 1:242
- Blogs, in office automation systems, 5:156
- Blood. *See also* Artificial blood; Bleeding;
 Hemo-entries
 banked, 1:512–513
 mass and thermal transport via, 1:188
 measurement of oxygen in, 6:522–524
 measurement of pH and carbon dioxide
 in, 6:525–526
 thermal properties of, 2:28
- Blood-cell cancers, 4:606
- Blood cell counters, 2:81–90. *See also*
 Complete blood cell count
 automated, 2:88–89
 electronic, 2:84–85
 history and principles for, 2:82–86
 rationale for cell count, 2:82
 types of, 2:82–86
- Blood cell counting, impedance
 plethysmography and, 4:130
- Blood cells. *See also* Red blood cell entries;
 White blood cell entries
 characteristics of, 2:81t
 measuring rheological properties of,
 1:506–507
 nature of, 2:81–82
 biomaterial failure and, 1:280
- Blood collection, tube guide for, 1:457t
- Blood collection/processing, 1:455–465
 safety in, 1:456
 specimen storage in, 1:460t
 standard changes in, 1:455–456
- Blood collection system/equipment,
 1:457–458
- Blood compatible glassy carbons, 1:301
- Blood component home health care devices,
 3:531–532
- Blood contact, biomaterial failure related
 to, 1:280
- Blood flow. *See also* Blood rheology;
 Cutaneous blood flow
 character of, 3:21–22
 cooling and, 3:466
 effective, 2:16–17
 heat transfer and, 1:191
 maldistribution of, 6:164
 pulmonary, 2:16–18
 smooth muscles and, 1:190
 systemic, 2:16–18
 thermotherapy and, 3:464–465
- Blood flow conductivity, based on
 erythrocyte orientation, 1:208
- Blood gas analysis
 considerations in, 6:526–527
 impact of, 2:113
- Blood gas analysis apparatus, 2:112
 three-function, 2:113
- Blood gas catheters, intravascular optical,
 5:168–169
- Blood gas diffusion, through the skin,
 1:476–477
- Blood gas electrodes, 1:465–467
 principles of, 1:466
- Blood gas measurement(s), 1:465–483
 basic concepts in, 1:465
 capnometry and capnography, 1:478–
 483
 continuous intravascular blood gas
 monitoring, 1:474–476
 ear oximetry, 1:471
 intrapartum fetal pulse oximetry,
 1:475–476
 neonatal, 4:590–591
 in neonatal monitoring, 5:22–25
 optical sensors in, 5:168
 oximetry in, 1:469–471
 pulse oximetry, 1:471–474
 transcutaneous blood gas monitoring,
 1:476–478
- Blood gas monitoring
 continuous intravascular, 1:474–476
 fetal, 3:298–299
 transcutaneous, 1:476–478
- Blood gas physiology, 1:467–469
- Blood glucose
 controlling, 3:402–403
 prediction of, 3:402
- Blood glucose monitoring, ambulatory,
 1:16–17
- Blood glucose sensors, 1:16–17
- Blood loss anemia, hyperbaric medicine
 and, 4:26
- Blood oximeters, 6:523
- Blood-oxygen-level-dependent (BOLD)
 imaging, 5:246
- Blood oxygen monitoring
 methods development in, 2:113–114
 transcutaneous, 2:113
- Blood $p\text{CO}_2$, methods of measuring, 2:109–
 110
- Blood perfusion, 1:188
 effects on heat transfer, 1:189–190
- Blood platelet aggregation, collagen-
 induced, 1:107
- Blood pressure (BP). *See also* Arterial
 tonometry; Continuous BP monitoring
 automatic control of, 1:490–500
 central and peripheral, 3:491–492
 control schemes for, 1:491–498
 estimation of, 1:488
 home health care devices, 3:526–527
 monitoring equipment for, 4:569–570
- Blood pressure load, 1:14–15
- Blood pressure measurement, 1:485–490
 algorithmic components of, 1:487–488
 arterial, 5:235
 continuous, 5:235–237
 direct techniques of, 1:485–486
 in exercise stress testing, 3:247–248
 in shock patients, 6:165
 instantaneous, 5:235
 neonatal, 5:26–27
 noninvasive (indirect) techniques of,
 1:485, 486–489
- sources of inaccuracy in, 1:488
- Blood pressure measurement devices,
 1:489
 accuracy of, 1:489–490
- Blood pressure measurement techniques,
 alternative, 1:488–489
- Blood pressure monitoring, 2:8–9;
 4:568–569
 ambulatory, 1:13–16, 489
 equipment for, 4:569–570
 finger, 1:489
 semiautomatic, 1:489
 static calibration in, 4:570
 wrist, 1:489
- Blood pressure profile, interpretation of,
 1:15
- Blood rheology, 1:500–511. *See also* Blood
 flow
 clinical conditions and, 1:503–509
 rheological properties of blood,
 1:500–502
- Blood specimen processing, 1:459–464
- Blood specimens, collecting, 1:19
- Blood vessels
 anatomy of, 6:491
 in coronary artery replacement, 6:394
 engineered, 3:204
- Blood viscosity, 1:500–502
- Blotting techniques, 4:603
- Blue light phototherapy, 6:486
- B lymphocytes, implants and, 1:112, 113
- Bode plot, 1:124–125; 5:375
- Bodily systems, age-related alterations in,
 1:390
- Body
 oxygen transport in, 5:209–210
 temperature regulation of, 6:378–319
- Body composition, impedance
 plethysmography and, 4:129–130
- Body fat home health care devices, 3:530–531
- Body fluid analysis, reasons for, 1:18
- Body mold systems, 5:589–590
- Body-section tomographic equipment,
 quality control of, 6:574
- BODY Simulation, 5:303, 304–305
- Body surface area (BSA), 2:15, 18
- Body surface potential mapping, 3:49–50
- Body temperature
 defined, 6:312
 effects of elevated, 4:46–48
- Body temperature home health care
 devices, 3:529–530
- Body tissues, values of resistivity of, 4:122t
- Body weight home health care devices,
 3:532–533
- Bohr effect, 2:41
- Bohr equation, for physiological dead
 space, 5:431–432
- Bokros, Jack, 1:302
- Boltzmann transport equation, 5:534
- Bolus, 5:596
- Bolus infusion test, 4:4
- Bond angles, in protein structure
 prediction, 1:220
- Bonding. *See also* Bone bonding
 fusion, 2:3, 4f
 hydrophobic, 2:3

- Bone. *See also* Osteo- entries
 adaptive properties of, 1:534–535
 anisotropic properties of, 1:529
 bioceramics and, 1:284–285
 cancellous, 1:524, 533–534
 ceramic biomaterials and, 1:272–273
 chemical bonding to biomaterials, 1:109–110
 as a composite material, 1:532–533
 elastic anisotropy of, 1:526t
 elasticity of, 1:525
 elastic properties of, 1:526t
 electrical properties of, 1:535–536
 electric character of, 1:558–560
 fatigue strength of, 1:531
 feedback mechanisms of, 1:535
 fracture in, 1:542
 mechanical properties of, 1:523–534; 6:420
 properties of, 1:523–540
 stiffness of, 1:525t, 529
 strength of, 1:530
 stress concentrations of, 1:531
 structure of, 1:523–524
 synthetic hydroxyapatite and, 1:285
 treatments invasive (implanted) electric, 1:562–563
 viscoelasticity of, 1:531–532
 yielding and plastic deformation of, 1:530–531
- Bone bonding, 1:109–110
- Bone cement(s). *See also* Acrylic bone cement
 commercial, 1:541t
 components of, 5:193t
 creep of, 1:546–547
 PMMA, 1:336
- Bone cement polymerization process, phases of, 1:542t
- Bone defects, treatment efficacies for, 1:568
- Bone deformities, developmental, 6:231–232
- Bone degeneration, spinal, 6:232
- Bone densitometry
 radiography-based, 1:550–553
 single-energy, 1:551
- Bone density analysis, computed tomography and, 2:243
- Bone density measurement, 1:550–558
 radiography-based densitometry, 1:550–553
- Bone disease, implants for, 6:234
- Bone fixation, 1:256
- Bone grafts, for spinal fusion, 6:236
- Bone health, strain gages in, 6:287
- Bone imaging, using single photon emission computed tomography, 2:283
- Bone implants, tissue response to, 1:109–110
- Bone ingrowth, 1:109
- Bone measurement
 QUS parameters in, 1:554–555
 speed of sound or ultrasonic wave propagating velocity for, 1:555
- Bone mineral density (BMD), 1:529–530
 assessment of, 1:550
- Bone quality, motivation to assess, 1:555–556
- Bone remodeling, 1:534–535
 cellular and biochemical aspects of, 1:535
- Bone scans, with Anger camera, 1:53f
- Bone status measurement methods, 1:555–556
- Bone stimulators, 6:237
- Bone strain, *in vivo*, 1:531
- Bone substitutes, 1:355
- Bone tissue, engineered, 3:203
- Bone tissue engineering
 combination approaches to, 6:391–392
 inductive approaches to, 6:391
 strategies for, 6:389–392
- Bone treatment, noninvasive electric, 1:563–564
- Bone tumors, cryosurgical treatment of, 2:375
- Bone ununited fracture, 1:558
- Books/reports. *See* Medical books/reports
- Boolean networks, 1:226
- Bootstrap method, 6:260
- Boranes, polyhedral, 1:573–574
- Borderline hypertension, 1:15
- Borg scale, 3:252
- Borg scales of perceived exertion, 3:254t
- Boron-containing agents, optimizing delivery of, 1:576–577
- Boron-containing amino acids, 1:573–574
- Boron-containing porphyrins, 1:575
- Boron delivery agents, 1:573
 high molecular weight, 1:576
 low molecular weight, 1:573–575
- Boron neutron capture therapy (BNCT), 1:571–58; 6:8, 9
 clinical studies of, 1:579–581
 clinical trials related to, 1:579–582
 neutron sources for, 1:577–579
 radiobiological considerations related to, 1:572–573
- Boronophenylalanine (BPA), 1:572
 clinical trials of, 1:581
- Boston brace scoliosis treatment, 6:131
- Boston Scientific ablation system, 6:373
- Boundary lubrication, 1:315; 6:417
- Bovine plexiform bone, 1:524, 525f
- Bovine spongiform encephalopathy (BSE), 1:513, 517
- Bowel disease, inflammatory, 3:392
- Boyle, Henry, 1:40
- Boyle's law, 2:14; 4:19
- Brace function, biomechanical evaluation of, 6:130
- Braces
 orthopedic, 6:231
 use in scoliosis, 6:130
- Brachytherapy, 5:482. *See also* High-dosage-rate brachytherapy; Intravascular brachytherapy (IVB)
 gels in, 5:492–493
 high dose rate, 6:26–27
 regulations related to, 2:162, 163t
- Brachytherapy formalism, standardized, 5:423
- Bracing, computer modeling of, 6:130–131
- Braille Institute, 1:446
- Braille readers/displays, 1:445–446
- Brain. *See also* Magnetoencephalography (MEG)
 musculoskeletal system and, 1:385
 stroke and, 2:52–53
- Brain dysfunctions, EEG and, 3:68
- Brain function, MEG studies of, 1:244
- Brain-generated noise, 1:234f
- Brain imaging, using single photon emission computed tomography, 2:283
- Brain injury
 acquired, 6:71
 biofeedback clinical outcome literature related to, 1:182–183
 cognitive training for persons with, 6:71–72
 language disorders associated with, 2:211
- Brain monitoring, in anesthesia, 4:558–560
- Brain pathology, MEG applications and, 1:245–246
- Brain perfusion, impedance plethysmography and, 4:128
- Brain perfusion monitoring, in shock assessment, 6:167
- Brain stimulation, electromagnetic, 3:60.
See also Electroconvulsive therapy (ECT)
- Brain tumors
 clinical studies of BNCT for, 1:579–581
 recent and ongoing clinical trials related to, 1:579–581
- Branched polymers, as biomaterials, 1:274
- Branching, in thermoplastics, 1:332
- Breast bridge, 5:590
- Breast cancer, high intensity focus ultrasound for, 4:78–79
- Breast cancer detection, IR imaging in, 6:349
- Breast magnetic resonance imaging, 4:294
- Breathable layers, for electrodes, 1:140–141
- Breath control devices, 5:593
- Breathing
 sleep-disordered, 6:211–212
 work of, 6:517–518
- Breathing circuit(s), 1:31–32, 40
- Breathing frequency
 measurement of, 6:520–522
 monitoring, 6:515
- Breath sounds, 5:378
 in neonatal respiratory monitoring, 5:17
- Bridges, dental, 1:325–329
- British Hypertension Society (BHS), 1:489, 490
- Broadband ultrasound attenuation (BUA), 1:554
 in trabecular bone measurement, 1:555
- Bromocresol green (BCG), 1:20
- Bronchial tumors, cryosurgical treatment of, 2:374
- Brookhaven studies, of BNCT tolerance, 1:580–581
- Brush electrode systems, 1:156
- Bryant–Cardan angles, 4:210–211
- Bubble equilibration methods, 2:109–110
- Buckminsterfullerene, 1:298

- Bucky balls, 1:298, 305
 Bucky tray, 6:569
 Bulk micromachining technologies, 2:3, 4f
 Bulletins, Nuclear Regulatory Commission, 2:171
 Burger, Rudolph, 1:138
 Burn injury, 6:170–172
 Burns, 1:192
 bioactive skin substitutes for, 6:169–179
 deep second degree (deep partial thickness), 6:170–171
 etiology and prognosis relative to depth of, 6:171t
 laser Doppler flowmetry for, 2:382
 mean survival rate after, 6:171t
 scars and pain in, 6:171–172
 size and depth of, 6:170
 superficial second degree, 6:170
 third degree (full thickness), 6:171
 Bursae, 4:199–201
 Bursitis, tissue regeneration and, 1:109
 Burst and analyzing methods. *See* EEG burst and analyzing methods
 Burst detection
 reasons for, 3:73–74
 methods of, 3:79–80
 Burst-episodes, analyzing, 3:74
 Bursting, signal-power changes during, 3:74–79
 Bursts, mechanisms of, 3:72–73
 Butler–Volmer equation, 1:124, 129
- Cabling, in electroneurography, 3:116
 Cadaver models, osteoligamentous, 3:573
 Cadaver skin, as a skin substitute, 6:173–174
 CADE schemes, 2:298
 Cadmium telluride (CdTe) detectors, 5:518
 CAD schemes. *See also* Computer-assisted detection/diagnosis (CAD)
 evaluation of, 2:298–302
 scoring criteria for, 2:299–300
 CAD server, 2:303
 CADx schemes, 2:297–298
 Cages, in spine stabilization, 3:571–572
 Calcineurin inhibitors, 4:275
 Calcium alginate scaffolds, 1:367
 Calcium ceramics, resorbable, 1:260–262
 Calcium concentration, in bioactivity testing, 1:360–361, 364–365
 Calcium-containing materials, bone bonding to, 1:109–110
 Calcium ions, bioactive glasses and, 1:286
 Calcium oxides, as biomaterials, 1:273
 Calcium phosphate materials, 1:261–262
 Calcium phosphates, as biomaterials, 1:108
 Calcium scoring, computed tomography and, 2:243
 Calcium sulfate materials, 1:260–261
 Calendars, in office automation systems, 5:154
 Calibration
 audiometric, 1:93
 in effective thermal property measurements, 1:196
 in intraoperative radiotherapy, 6:23
 resistance, 1:194, 195
 Caloric testing, vision-related, 5:140
 Calorimetry, in determining absorbed dose and air kerma, 5:469–470
 Cameras, fundus, 5:291
 Canada, infrared imaging in, 6:353
 Cancellous bone, 1:524
 mechanical properties of, 1:533–534
 Cancer(s). *See also* Carcinogenicity; Oncology; Tumors
 adoptive T cell immunotherapy of, 4:115–116
 blood-cell, 4:606
 cryosurgical treatment of, 2:373–374
 positron emission tomography in, 5:413–414
 systemic hyperthermia in, 4:56–59
 Cancer chemotherapy effects, biofeedback clinical outcome literature related to, 1:180
 Cancer immunotherapy, approaches to, 4:605
 Cancer risks, associated with low-dose X rays, 2:259–260
 Cancer therapy. *See also* Cancer treatment
 heavy-ion radiotherapy in, 6:10–11
 nanoparticles in, 5:6
 Cancer treatment, prostate, 6:376
 Cantilever calibration, in force spectroscopy, 4:510
 Capacitance
 double-layer, 1:123
 at electrode-electrolyte interface, 1:125
 Capacitance strain gages, 6:283
 Capacitance-to-frequency converter, 2:8
 Capacitive coulometry oxygen analysis, 5:204
 Capacitive coupling (CC), 1:563–564
 Capacitive electronic interfaces, 2:7–10
 Capacitive microsensors, 2:1–12
 fabrication technologies for, 2:2–5
 medical field applications of, 2:2
 operation issues of, 2:5–6
 sensitivity of, 2:5–6
 Capacitive pressure sensors, 2:2, 5
 Capacitive transducers, 2:2
 Capillary array electrophoresis, 2:432
 Capillary electrometer, 1:137
 Capillary electrophoresis, 2:432–433
 in microdialysis sample analysis, 4:409–410
 Capillary gas chromatography, 2:104
 Capillary isoelectric focusing (CIEF), 2:106
 Capnography. *See also* Capnometry
 clinical uses of, 1:481
 limitations of, 1:481
 in neonatal monitoring, 5:14
 phases of, 1:480–481
 role in CPR, 1:483
 Capnometers
 CO₂ sampling techniques using, 1:479–480
 components and operational principle of, 1:482
 Capnometry, 1:478–480
 measurement techniques in, 1:478–479
 role in CPR, 1:483
 sublingual, 1:481–482
 Carbamino compounds, in blood CO₂ transport, 1:468, 469
 Carboflo vascular graft, 1:276f
 Carbon, as a biomaterial, 1:273. *See also* Carbons
 Carbon biomaterials, 1:296–308
 biocompatibility of, 1:301
 medical applications of, 1:301–306
 properties of, 1:300–301
 Carbon dioxide. *See also* CO₂ entries; End-tidal CO₂; pCO₂ electrode
 in anesthesia delivery, 1:31, 32
 in blood, 1:465; 6:525–526
 in circle breathing circuit, 1:33
 CPR and, 2:39–40
 measurement in gas, 6:526
 Carbon dioxide content (ctCO₂), 2:19–20
 Carbon dioxide electrodes. *See* CO₂ electrodes
 Carbon dioxide monitor, 1:43f
 Carbon dioxide transport
 in blood gas physiology, 1:468–469
 cardiopulmonary resuscitation and, 2:40–41
 Carbon fibers, 1:299
 Carbon-filled silicone rubber electrode, 1:148f
 Carbonic acid, in blood CO₂ transport, 1:468
 Carbon monoxide, in anesthesia delivery, 1:33
 Carbon monoxide poisoning, hyperbaric medicine and, 4:26
 Carbons. *See also* Carbon
 activated, 1:300, 301–302
 amorphous, 1:297
 glassy, 1:299
 naturally occurring, 1:296–298
 pyrolytic, 1:302–306
 structure of, 1:296
 synthetic, 1:299–300
 Carcinogenicity
 implants and, 1:112–113
 of metallic biomaterials, 1:104–105
 Cardiac action potential, 3:143–144
 Cardiac arrest dysrhythmias, 2:44–48
 Cardiac Arrest Survival Act, 2:49
 Cardiac arrhythmia(s), 6:370
 biofeedback clinical outcome literature related to, 1:180–181
 Cardiac autoregulation, in whole-body models, 5:307
 Cardiac catheter ablation, 6:369–372
 Cardiac catheterization, in shock assessment, 6:167
 Cardiac cells, mathematical modeling of, 3:144–145
 Cardiac compression–cardiac flow hypothesis, 2:37–38
 Cardiac contractility, 3:478–480
 Cardiac cycle, 1:485; 3:477; 4:163
 Cardiac cycle event detection, 1:201–202
 Cardiac death, sudden, 1:69
 Cardiac electrical activity, 2:43–44; 3:40f
 Cardiac electrodes, 1:149–150

- Cardiac Event Monitoring (CEM), 1:13
- Cardiac function, assessing, 3:480–482
- Cardiac gated CT, 2:238
- Cardiac index (CI), 2:18
- Cardiac magnetic resonance imaging, 4:292–294
- Cardiac mapping and imaging, 6:371
- Cardiac monitors, combined with transthoracic impedance, 5:22
- Cardiac muscle, 2:43; 6:393
- Cardiac neonatal monitoring, 5:13
- Cardiac operation, minimally invasive, 4:525
- Cardiac output (CO), 2:12–13, 15; 3:21.
See also Heart entries
determination of, 4:573–574
Fick technique for, 2:12–21
indicator dilution measurement of, 2:21–25
noninvasive measurement of, 1:200–201
thermodilution measurement of, 2:25–35
- Cardiac output computers, 2:31
- Cardiac pacing, 1:150–151, 151–152
- Cardiac performance, age-associated changes in, 1:390
- Cardiac perfusion imaging, using single photon emission computed tomography, 2:282
- Cardiac physiology, 2:42–44
- Cardiac rhythms, 1:174. *See also* Cardiac cycle
- Cardiac structures, physical modeling of, 5:287–288
- Cardiac surgery
high frequency jet ventilation and, 3:507
history of, 3:459–461
- Cardiogenic artifact, 5:19
- Cardiogenic artifact rejection, in neonatal respiration monitoring, 5:22
- Cardiogenic shock, 6:164
management of, 6:168
- Cardiology, use of endoscopes in, 3:181–183. *See also* Heart entries
- Cardiomyocytes (CMs), 6:393
- Cardiopulmonary biofeedback, 1:174–176
- Cardiopulmonary bypass (CPB), 3:459–460
- Cardiopulmonary bypass circuit, 3:461–462
- Cardiopulmonary resuscitation (CPR), 2:35–63. *See also* Cardiac physiology; Community CPR; Pulmonary physiology
adult, 2:53–55
airway obstruction and, 2:55–56
automatic external defibrillator and, 2:56–57
cardiac arrest dysrhythmias and, 2:44–48
cardiovascular disease and, 2:50–51
cerebrovascular disease and, 2:51–53
certification in, 1:456
historical events in, 2:59
historical perspective on, 2:35–38
oxygen and carbon dioxide transport and, 2:40–41
pediatric and infant, 2:57–58
pulmonary circulation and, 2:41–42
- role of capnometry–capnography in, 1:483
- Cardiopulmonary stress testing, 5:440–441
- Cardiotocography, fetal, 3:296–298
- Cardiovascular applications, polymers in, 1:276
- Cardiovascular–circulation systems modeling, 5:307–310
- Cardiovascular devices, biomaterial surfaces of, 1:343
- Cardiovascular disease, 2:50–51
- Cardiovascular events, hypertensive patients with high risk of, 1:15–16
- Cardiovascular medicine, bioimpedance in, 1:197–216
- Cardiovascular monitoring, thermistors in, 6:335
- Cardiovascular reactivity
biofeedback clinical outcome literature related to, 1:180–181, 182
- Cardiovascular regulation systems modeling, 5:310–313
- Cardiovascular system
high intensity focus ultrasound in, 4:79–81
role in respiratory mechanics, 6:105–106
- Cardiovascular tissue engineering, strategies for, 6:392–394
- Cardioverter defibrillators, implantable, 1:69–70, 71–79
- Career preparation, for biomedical engineering, 1:404–405
- CareSuite system, 1:44–45
- Carol, Mark, 6:397
- Carpometacarpal (CMC) thumb joint replacements, 1:304
- Cartilage. *See also* Collagen(s)
articular, 4:202–204
biomaterial scaffolds necessary for engineering, 2:74
cells capable of generating, 2:73
composition and structure of, 2:63–66
engineered, 3:203–204
functional tissue engineering of, 2:74–75
hyaline and articular, 2:63–65
mechanical properties of, 2:66–71
properties of, 2:63–80
regeneration of, 2:73
repair strategies of, 2:71–73
in tissue engineering, 1:366; 2:73–75
wear and degeneration of, 2:71
- Cartilage applications, for porous biomaterials, 5:403
- Cartilaginous joints, 4:199, 202f
- Cartilaginous tissue, masticatory system, 6:420–421
- CARTO system, 6:372
- Cassette-type ECG recorder, 1:13
- Cash flows, negative and positive, 1:26–27
- Cassen, Benedict, 1:51
- Casson fluid, blood as, 1:501, 502
- Catheterization, umbilical artery/vein, 4:589–590
- Catheters, 2:2
balloon, 2:350–352
calibrating, 2:31
distal, 4:12
- guiding, 2:349
with heating elements, 2:27, 28
intravascular optical blood gas, 5:168–169
in microsurgery, 4:532
mixed-venous fiber optic, 5:165–166
proximal, 4:9–10
pulmonary artery, 2:29–30; 4:572–573
Swan–Ganz, 2:26, 30
for thermal dilution measurement, 2:24f
umbilical artery, 4:594–595
umbilical artery/vein, 4:590
umbilical venous, 4:595
volume conductance, 1:206–207
water-perfused, 1:62
- Catheter-tip transducer systems, 4:579
- Cathode, 1:465–466
- CEDIA enzyme immunoassay, 1:21
- Cell adhesion, protein-mediated, 5:396–397
- Cell-based drug screening, impedance spectroscopy as a transducer in, 4:140–142
- Cell behavior
growth factors necessary for modulating, 2:74
at surfaces, 1:345
- Cell counters, blood, 2:81–90
- Cell culture, in medical microbiology, 4:375
- Cell Culture Analogs (CCAs), micro, 4:393
- Cell elasticity measurements, 4:512
- Cell imaging, impedance plethysmography and, 4:130–131
- Cell manipulation, nanoparticles in, 5:6–7
- Cell-mediated immune response, 1:113
- Cell metabolism
cooling and, 3:466
temperature and, 3:464
- Cell patterning, potential applications for, 1:413–414
- Cell Preparation Tube (CPT), separation of peripheral blood mononuclear cells using, 1:462–464
- Cells
in automated cytology, 2:389–390
in biosurface engineering, 1:409
controlling attachment, morphology, and differentiation of, 1:414
instrumentation for defining, enumerating, and isolating, 4:603–604
monitoring attachment and spreading in, 4:139–140
in tissue engineering, 1:366
tissue regeneration and, 1:108–109
- Cell seeding, in tissue engineering, 6:387
- Cell separation, in microbio reactors, 4:390–391
- Cell separation/purification, nanoparticles in, 5:5
- Cell sorting, 2:397–399
- Cell sources
in bone tissue engineering, 6:390
in cardiovascular tissue engineering, 6:392, 393, 394
- Cell-to-cell interactions, in engineered tissue, 3:200
- Cell types, in engineered tissue, 3:192–193

- Cellular cardiomyoplasty (CCM), 6:393
- Cellular imaging, 2:90–101
fluorescence and, 2:91
Raman and CARS microscopy, 2:98–99
- Cellular parameter measurement
in automated cytology, 2:392–400
devices used in, 2:397–400
- Cellular patterning techniques, 1:409–414
- Cellular processes, in engineered tissue,
3:190–194
- Cellular solid porous biomaterial
fabrication, 5:400
- Cellular thermal damage, 4:48
- Celsius temperature scale, 6:312
- Cemented prosthesis fixation, 5:194
- Cementless prosthesis fixation, 5:194–195
- Cements
calcium phosphate, 1:262
dental, 1:324–325
- Censoring, 6:262–263
- Centigrade temperature scale, 6:312
- Central auditory processing disorder, 2:211
- Central auditory system, 1:95
- Central limit theorem, 5:534
- Central nervous system applications, for
porous biomaterials, 5:404
- Centrifugal flow ventricular assist device,
3:454–455
- Centrifuges, separation of peripheral blood
mononuclear cells using, 1:462
- Cerabone, 1:287
- Ceramic materials, use in medical devices,
1:313–314
- Ceramic-on-ceramic hip joints, 3:521
- Ceramic-on-metal hip joint bearing,
3:521–522
- Ceramics
as biomaterials, 1:107–108, 272–273
as prosthetic restorative materials,
1:326–327
resorbable calcium, 1:260–262
- Cerebral autoregulation, in whole-body
models, 5:306–307
- Cerebral cortex, organization of, 3:63
- Cerebral function monitor, 5:35–36
- Cerebral injury, EEG and, 3:67
- Cerebral metabolism, EEG and, 3:67
- Cerebral State Monitor (CSM), 4:562
- Cerebral systems modeling, 5:313–316
- Cerebral topography, EEG and, 3:67
- Cerebrospinal fluid (CSF), 4:576–578
circulation of, 4:1–2
pulsations of, 4:2
resistance to reabsorption of,
4:3–5
- Cerebrospinal fluid drainage valves,
4:10–12
- Cerebrovascular disease, 2:51–53
vascular graft prostheses and, 6:493
- Cervical cap, 2:340–341
- Cervical dilatation, continuous monitoring
of, 3:300
- Cervical region, surgical procedures and,
3:566–568
- Cervical spine, anatomy of, 3:547–551
- Cervical spine region, degeneration–
trauma in, 3:562–564
- Cervical spine stabilization/fusion,
3:573–579
- CFR–CFR_g correlations, 2:357–359
- “Chain of Survival,” 2:49, 51, 52, 56
- Chain polymerization, 1:330–331
- Change
anchor-based methods of determining,
5:445t
clinically significant, 5:447–449t
distribution-based methods of
determining, 5:446t
group versus individual, 5:451
- Char, 1:299
- Charcoal, medical applications of, 1:301–302
- Chardack electrode, 1:151
- Charge-coupled device cameras, 4:492–493
- Charge-coupled devices (CCDs),
5:296–297; 6:557
- Charged particle patient totals, 6:2t
- Charge transfer mechanism, at electrode–
patient interface, 1:120
- Charles’ law, 2:14; 4:19
- Charnley, John, 1:255, 540, 541
- Chart abscissa generation, in graphic
recorders, 6:51
- Chat systems, in office automation
systems, 5:157
- Checkmate delivery device, 1:604
- Chemical ablation, 6:367–368
- Chemical colorimetric airway detector,
1:482–483
- Chemical composition, of thermoplastics,
1:332
- Chemical dosimetry, in determining
absorbed dose and air kerma,
5:470–471
- Chemically modified FETs (CHEMFETs),
4:190
- Chemicals, as a hospital problem, 6:111.
See also Compounds
- Chemical shifts, X-ray photon spectroscopy
and, 1:349–350
- Chemical sterilants, 6:279t
- Chemical stimulation, in developing visual
prostheses, 6:534
- Chemical vapor deposited (CVD) carbons,
1:299–300
- Chemical vapor deposition, for porous
biomaterial fabrication, 5:401
- Chemistry, of biological samples, 4:519–521
- Chemistry analyzers, 1:19–23
- Chemotherapy
in conjunction with hyperthermia, 4:68
systemic hyperthermia and, 4:56–58
- Chest electrodes, 1:138
- Chest pain, 2:50
- Chicago scintillation camera, 1:52
- Child Language Analysis (CLAN), 2:216
- Children
EGG in, 3:93–94
high frequency ventilation in, 3:508–509
indications for liver transplantation in,
4:269t
- χ^2 test, to compare unpaired samples,
6:250–251
- χ^2 value, 6:249
- China, infrared imaging in, 6:353
- Chin vs. St. Barnabos Medical Center,*
3:538–539
- Chitosan, in tissue engineering, 1:370
- Chloride electrode test, 2:387
- Chondrocytes, 2:63. *See also* Cartilage
entries
- Chromatographic systems, 2:102
- Chromatography, 2:101–109
general theory of, 2:102–103
types of, 2:103–108
- Chromium, from implants, 1:111–113
- Chromosome analysis, automated cytology
for, 2:403
- Chronic implantable system, 1:434–435
surgical technique for, 1:435–436
- Chronic pain, morphine analgesia for, 3:32
- Chronic regional pain syndromes, 6:229
- Chronic rejection, following liver
transplantation, 4:272–275
- Circadian blood pressure variability,
dipping and, 1:15
- Circle breathing circuits, 1:32f. *See also*
Semiclosed circle system
anesthesia machine, 1:38
virtues and limitations of, 1:33–34
- Circuits
in integrated-circuit temperature
sensor, 4:159–160
ISFET, 4:191–193
- Circular collimators, radiosurgery based
on, 5:578
- Circular electrodes, current density under,
1:145–146
- Circulation
assisted, 3:459–461
of cerebrospinal fluid, 4:1–2
coronary, 3:482–483
pulmonary, 2:41–42; 4:567
systemic, 4:567
in whole-body models, 5:305
- Circumferential cuff electrodes, 3:123–124
- Clarion hi-focus electrode system, 1:154
- Clark-type sensor, 1:477
- Class A and B bioactive materials, 1:289
- Classical conditioning, 1:166
- Classification
in automated cytology, 2:403
in computer-assisted detection/
diagnosis, 2:296
- Clausius–Clapeyron equation, 1:5
- Cleanliness, as a hospital problem, 6:114
- Cleft palate–lip, 2:211
- Clemson Advisory Board, 1:267
- Clinical and Laboratory Standards
Institute (CLSI), 1:455–456
- Clinical applications, strain gages in, 6:288
- Clinical biofeedback training, 1:167–168
- Clinical dosimetry, X-ray, 6:586–589
- Clinical engineering, medical device safety
program and, 6:117
- Clinical inspection, of EEG chart records,
3:69
- Clinical laboratory automation, 1:23–24
automation options and system design
in, 1:23–24
available automation systems for, 1:24
NCCLS guidelines for, 1:24

- Clinical literature, intracranial pressure monitoring, 4:580–582
- Clinically significant change, studies for determining, 5:447–449t
- Clinical management software, for communication disorders, 2:212
- Clinical significance. *See also* Clinically significant change
 checklist for assessing, 5:450t
 determining and interpreting, 5:444, 445–451
 recent developments related to, 5:451
- Clinical studies
 of BNCT for brain tumors, 1:579–581
 CAD, 2:301
 codes and regulations for, 2:147
- Clinical target volume (CTV), 5:545, 546
- Clinical testing questionnaire, 3:221t
- Clinical trials, 6:262–263
 codes and regulations related to, 2:143–144
 double blind, 6:262
 hyperthermia and chemotherapy, 4:70t
 hyperthermia and radiation, 4:69t
 within patient studies versus across patient studies, 6:262
- Clitoral vacuum, 6:154
- Clitstim, 6:154
- Closed-loop system control, 2:504
- Closed scavenger systems, anesthesia machine, 1:39
- Clotting. *See also* Coagulation; Thromb-entries
 biomaterial failure and, 1:280
 in pulmonary artery catheters, 2:30
- CLUSTAL-W, 1:219
- Clustering, in tracer kinetics, 6:435
- Clusters of orthologous groups (COG) genes, 1:223
- CMOS MEMS interface electronics, 1:418
- CMOS technology, 1:420, 421, 422, 423, 426, 427; 2:10
- CO₂, transcutaneous, 2:114–116. *See also* Carbon dioxide entries
- CO₂ absorbents, 1:32–33
- CO₂ electrodes, 2:109–120. *See also* Blood gas entries
 accuracy of, 2:117
 applications of transcutaneous technology, 2:116–117
 design details of, 2:111–113
 history of, 2:110–111
 limitations of, 2:117
- CO₂ sampling techniques, 1:479–480
- Coagulase negative staphylococci, 1:114
- Coagulation, of blood, 1:503. *See also* Clotting
- Coating properties, of plasma polymerization monomers, 1:346t
- Coatings, development of biologically responsive, 1:347–348
- Coating techniques, 1:355–365
- Cobalt
 from implants, 1:112–113
 properties of, 2:122t
- Cobalt-60 units, 2:120–133
 activation physics related to, 2:122–124
 calibration of, 2:128–130
 design of, 2:125
 first clinical applications of, 2:132
 geometric penumbra and, 2:131
 head design in, 2:125–126
 history of, 2:121–122
 isodose charts and, 2:131–132
 mounting, 2:126–127
 radiation beam characteristics in, 2:127–128
 relative dose functions and, 2:130–131
 source strength specification and, 2:124–125
- Cobalt alloys, as biomaterials, 1:270, 271
- Cobalt–chromium alloys, 1:312
 as biomaterials, 1:105
 in metal-on-metal prostheses, 1:317
- Cobalt–chromium–nickel alloys, in dental prosthetics, 1:325
- COBAS AMPLICOR™ analyzer, 4:376
- Cochlear implant electrodes, 1:154, 155
- Cochlear prostheses, 2:133–141
 acoustic and electrical stimulation in, 2:138
 auditory periphery and, 2:134–135
 auditory system and, 2:133–134
 benefits and risks of, 2:137
 bilateral implants of, 2:139
 candidates for implants, 2:133s
 conditioning pulses and, 2:138–139
 evaluation of, 2:137
 fine structure and, 2:139
 future of, 2:138
 high density electrode arrays and, 2:138
 history of, 2:133
 implantation cost of, 2:137–138
 operation of, 2:135–137
- Coded excitation, 6:469–471
- Coded harmonic excitation, 6:469–471
- Codes, medical device, 2:141–153. *See also* Radiation codes/regulations
- Cofactors, in enzyme activity measurement, 2:195
- Cognition, evoked potentials and, 3:236–237
- Cognitive assessment, for augmentative and alternative communication systems, 2:208
- Cognitive rehabilitation, computers in, 6:71–79
- Cognitive task analysis, effect on human factors and medical devices, 3:540
- Cognitive training
 for persons with brain injury, 6:71–72
 for persons with dementia, 6:73–74
 for persons with psychiatric disorders, 6:72–73
 for students with learning disabilities, 6:72
 virtual reality for, 6:74
- Cohen, David, 1:230
- Coherence function, in separating ventricular fibrillation from tachycardia, 1:79
- Coherent anti-Stokes Raman (CARS) spectroscopy, 2:98–99
- Coherent scattering, 6:593–594
- Coke, 1:299
- Cold indicator solution, 2:30–31
- Cold packs, 1:190
- Cold therapy. *See* Heat and cold therapy
- Cole, Kenneth S., 1:197–198
- Cole-Cole plot, 1:209–210
- Colitis, ischemic, 3:392
- Collaborative writing systems, in office automation systems, 5:157
- Collagen(s), 1:340–341, 523–524. *See also* Cartilage
 in arterial walls, 1:85
 as biomaterial, 1:107
 cooling and, 3:466
 in skin, 6:203–204
 synthetic hydroxyapatite and, 1:285
 in tendons and ligaments, 4:241–242
 thermotherapy and, 3:464
 in tissue engineering, 1:367–370; 6:384
- Collagen-induced blood platelet aggregation, 1:107
- Collagen matrices, naturally derived, 6:196–198
- Collagen scaffolds, 1:378
- College of American Pathologists (CAP), 1:455
 point of care testing guidelines by, 1:22–23
- Collimation
 Anger camera, 1:56–58
 in CT scanners, 2:235
- Collimator assembly, in X-ray equipment, 6:565
- Collimators, 6:554–555
 multileaf, 5:576
- Colloidal chemistry, nanoparticle fabrication via, 5:3
- Colloidal drug delivery devices, 2:464–486
 characterization of, 2:466
 classification of, 2:465
- Colloids, oxygen-carrying, 1:515–516
- Colonography, computed tomography, 2:261–263
- Colonoscopy, virtual, 2:261–263
- Colon targeting, cyclodextrins in, 2:460
- Color Doppler, 3:1–2, 9, 10
- Color Doppler flow imaging (CDFI), 5:245
- Color Doppler signal processing, 3:11f
- Color flow Doppler, 6:464–465
- Color flow imaging (CFI), 5:245
- Colorimetric airway detector, 1:482–483
- Colorimetry, 1:469; 2:187–197; 4:323–324
 clinical applications of, 2:191–196
 future developments in, 2:196
 instrumentation for, 2:189–190
- Colposcope, 2:198
- Colposcopy, 2:197–202
 accessory instruments to, 2:200–202
 digital, 2:201–202
 findings in, 2:198
 indications for 198, 2:199
 technique of 199, 2:200
- Combined blood gas analysis apparatus, 2:112

- Combined gas law, 2:14
 Combined magnetic field (CMF), 1:567
 Committed effective dose, 5:505
 Committed equivalent dose, 5:505
 Committees, in a total hospital safety program, 6:115
 Common peroneal nerve stimulation, 1:430
 Communication devices, 2:202–210. *See also* Augmentative and alternative communication (AAC) systems
 Communication disorders
 assessment of, 2:214–217
 assistive devices for, 2:221–224
 computer administration–information processing and, 2:211–213
 computer applications for, 2:210–229
 future directions of, 2:224–225
 intervention in, 2:217–221
 normal function analysis and, 2:213–214
 Communication, in a total hospital safety program, 6:115
 Communication networks, in teleradiology, 6:305–306
 Communicative competence, training for, 2:208–209
 Community CPR, mechanisms for, 2:48–50
 Compact bone
 as a composite material, 1:532–533
 electrical properties of, 1:536
 mechanical properties of, 1:523–534
 Comparative genomics, 1:222–223
 Comparative protein modeling techniques, 1:221
 Compartment modeling, 6:431–434
 three or more compartments in, 6:433
 Compartment syndrome, hyperbaric medicine and, 4:25–26
 Compensating filter, 5:597
 Complement system, implants and, 1:112
 Complete blood count (CBC), 1:459, 460; 2:86–88
 Complex formation
 in colorimetry, 2:192
 methods used to detect, 2:454
 Complex impedance plot, 1:124
 Complexity-based neurological monitor, 5:39
 Complexity measurements, in separating ventricular fibrillation from tachycardia, 1:79
 Complex systems, niosomes in, 2:475
 Compliance, in anorectal manometry, 1:67
 Compliance issues, in continuous positive airway pressure, 2:335–336
 Composite biomaterials, 1:108. *See also* Composites
 Composite resins. *See* Resin-based composites
 Composites. *See also* Composite biomaterials
 bioactive, 1:289
 resin-based, 6:93–99
 resorbable, 1:262–264
 Compound biological effectiveness (CBE), 1:572
 Compound microscope, 4:523
 Compounds. *See also* Chemicals
 fluorescent, 3:345
 identification in biological samples, 3:345
Comprehensive Accreditation Manual for Hospitals (CAMH), 6:115
 Compressed gas drive mechanism, 6:506
 Compressed gas inlets, in anesthesia machines, 1:34–35
 Compressed Gas Association (CGA) gas system standards, 3:381
 Compressive properties, of cartilage and meniscus, 2:67–69
 Compton scattering, 6:595–596
 Computation, in exercise training, 1:395
 Computational modeling, 1:224–227
 Computational models
 dissemination of, 3:147–148
 impact on ventricular cells, 3:148
 of murine ventricular action potentials, 3:146–147
 visual prostheses and, 6:540
 Computation times, in computer-assisted detection/diagnosis, 2:296–297
 Computed radiography (CR), 5:337
 comparison with digital radiography, 5:345–346
 for mammography, 4:302
 Computed radiography systems, 5:337–338
 available, 5:339
 technological advances in, 5:339–340
 Computed tomography (CT), 2:230–258. *See also* Computed tomography screening; Computerized tomography entries; CT entries; Single photon emission computed tomography (SPECT)
 artifacts in, 2:253–257
 basic principles of, 2:230–233
 cardiac gated, 2:238
 cone beam, 2:250
 in diagnosing implant-related infection, 1:116
 display techniques in, 2:239–242
 electron beam, 2:232–233
 evolution of, 2:231–233
 fast X-ray, 5:246
 helical, 2:233
 image quality in, 2:250–253
 multidetector, 2:233
 noise levels in, 6:575–576
 numbers in, 2:238–239
 phantom materials in, 5:266–267
 quality control in, 6:574–578
 quantitative analysis in, 2:243
 radiation dose in, 2:244–246
 in radiosurgery treatment planning, 5:577
 scan pitch in, 2:237
 special clinical functions of, 2:242–244
 stereotaxis based on, 6:267
 techniques in, 2:237–238
 Computed tomography angiography (CTA), 2:242–243
 Computed tomography colonography, 2:261–263
 Computed tomography screening, 2:258–266
 cancer risks associated with, 2:259–260
 for early stage lung cancer, 2:263–264
 full-body, 2:26
 radiation doses from, 2:260–261
 Computed tomography simulation
 for advanced radiation therapy, 2:273–275
 for intensity modulated radiation therapy, 2:273–274
 process of, 2:269–273
 for stereotactic radiosurgery, 2:275
 for tomotherapy, 2:274–275
 Computed tomography simulator, 2:266–277; 5:529–530. *See also* Computed tomography simulation
 future of, 2:275–276
 patient data acquisition using, 2:268
 patient positioning and immobilization for, 2:268–269
 Computer administration–information processing, for communication disorders, 2:211–213
 Computer-aided anesthesia control, 1:47–50
 Computer aided design (CAD), in orthopedic device designs, 5:190
 Computer-aided diagnosis (CAD), 6:351. *See also* Computer-assisted detection/diagnosis (CAD)
 Computer-aided radiation dose planning, 5:455–463
 dose calculation in, 5:458–460
 dose display and plan evaluation in, 5:462–463
 image registration in, 5:457–458
 treatment plan optimization in, 5:460–462
 virtual simulation in, 5:455–457
 Computer algorithms, in pulse generators, 5:222. *See also* Algorithms
 Computer analysis
 of polysomnographic data, 6:214–219
 of sleep microstructure, 6:219
 of sleep studies, 6:213–224
 Computer animation, in EEG biofeedback instrumentation, 1:171
 Computer applications, for communication disorders, 2:210–229
 Computer-assisted anesthesia control systems, 1:48, 49f
 Computer-assisted cognitive retraining (CACR), 6:71–72
 Computer-assisted detection/diagnosis (CAD), 2:284–306. *See also* CAD entries
 advanced applications of, 2:302–303
 clinical studies of, 2:301–302
 defined, 2:285
 future studies of, 2:303
 need for, 2:287–288
 workings of, 2:285–287
 Computer-assisted detection/diagnosis algorithms, 2:288–297
 evaluation of, 2:297–302
 Computer-assisted instruction (CAI), 6:72

- Computer-assisted sleep staging (CASS), 6:215–219
- Computer-based biofeedback instrumentation, 1:168–169
- Computer-based heart beat detection, 3:50–51
- Computer-based instrument systems, software in, 2:314–316
- Computer-based interactive programs, for Alzheimer's disease, 6:73
- Computer-based medical record systems, 4:351–361. *See also* Computer-based patient records (CPRs)
- Computer-based monitoring ECG systems, 3:50–51
- Computer-based patient records (CPRs), 4:352–353
exclusive features of, 4:353–354
features of, 4:355–358
roadblocks in implementing, 4:354–355
- Computer-generated technology, for rehabilitation, 6:74
- Computer information technology, in anesthesia, 1:44–47
- Computerization, in exercise stress testing, 3:249
- Computerized decision support, in hemodynamic monitoring, 4:575
- Computerized injectors, in anesthesia machines, 1:41
- Computerized medical device management systems, 6:118
- Computerized Profiling (CP), 2:216
- Computerized tomography, in cryosurgery, 2:372. *See also* Computed tomography (CT)
- Computerized tomography books/reports, 4:344
- Computerized variable bypass vaporizers, in anesthesia machines, 1:41
- Computer modeling, of bracing, 6:130–131
- Computer networking, 5:350–353
- Computer networks, large-area, 1:47
- Computers. *See also* Personal computers; Virtual reality
in anesthesia, 1:42–51
in biomechanics, 1:385, 386–387
in the biomedical laboratory, 2:306–321
books/reports on, 4:349
cardiac output, 2:31
in cognitive rehabilitation, 6:71–79
with EEG instrumentation, 1:172
in electrocardiography, 3:34–53
in exercise training, 1:393–401
in hybrid SPECT/CT and PET/CT systems, 5:119–121
image processing, analysis, and display using, 5:117–118
information technology and, 5:118–119
integration into polysomnography, 6:214
limitations in sleep analysis, 6:221–223
in medical education, 4:307–311
in NM imaging, 5:108–114
in nuclear medicine, 5:106–124
in tomographic reconstruction, 5:114–117
patient records and, 4:310–311
talking, 1:445, 446–447
utility in advanced anesthesia monitoring, 1:43–44
- Computer simulation, 1:45–46
applications of, 5:121–122
- Computing, statistical, 6:263–264
- Comroe, Julius, 5:430
- Concentration, activity and, 1:123
- Concentration recovery, in microdialysis sampling, 4:412
- Concentration–time curve, in pharmacokinetics, 5:270–272
- Concentric contraction, 1:392
- Concept validation, in IR-based breast screening, 6:352–353
- Condensation polymeric biomaterials, 1:274
- Condensation polymerization, 1:330
- Condenser lens, in the transmission electron microscope, 4:481
- Conditioned response (CR), 1:166
- Conditioned stimulus (CS), 1:166
- Conditioning, 1:166
- Conditioning film, 1:115
- Condoms. *See also* Contraceptive devices
female, 2:339–340
male, 2:338–339
- Conductance. *See* Skin conductance activity (SCA)
- Conductance catheter systems, 1:206–207
- Conduction-heat devices, 3:467
- Conductive electrodes, 1:146–148
current density under, 1:145–146, 148
- Conductive keratoplasty, 6:378
- Conductive rubber electrodes, 1:148
- Conductivity, of biological samples, 4:519–521
- Conductors, 1:466
- Cone beam CT, 2:250
- Cone beam errors, in computed tomography, 2:257
- Cone-plate visometer, 1:505f
- Conference of Radiation Control Program Directors (CRCPD), 2:156–157, 172. *See also* CRCPD model regulations
- Confocal fluorescence microscopy, 4:493–494
- Confocal microscope light detectors, 4:455–457
- Confocal microscopy, 4:449–477. *See also* Confocal fluorescence microscopy
acoustooptic tunable filters in, 4:457–459
advantages and disadvantages of, 4:453–455
Airy disk and lateral resolution in, 4:462–464
Alexa Fluor dyes in, 4:467–468
benefits of acoustooptic tunable filters in, 4:459–461
cyanine dyes in, 4:468
fiber-based, 3:309–310
fluorescent dyes in, 4:466–467
fluorescent environmental probes in, 4:468–469
fluorescent proteins and, 4:471–472
fluorophores for, 4:464–465
laser and arc-discharge spectral lines in, 4:466t
laser scanning, 2:92–93
laser scanning configuration in, 4:452–433
organelle probes in, 4:469–470
principles of, 4:450–452
quantum dots in, 4:470–471
quenching and photobleaching in, 4:472–474
resolution and contrast in, 4:461–462
- Conformal particle therapy, 6:34
- Conformation, of biological samples, 4:519–521
- Conformity–Gradient Index (conformal) (CGIc), 5:583
- Conformity–Gradient Index score (CGI_s), 5:582–583
- Confrontation visual field exam, 6:528–529
- Congenital cystic adenomatoid malformation, 4:172–173
- Congenital diaphragmatic hernia, 3:506
- Congenital heartstagmus, 5:140
- Congestive heart failure (CHF), 1:205–206
- Conical metal disk electrodes, 1:139
- Connection errors, in thermocouples, 6:344
- Consent, for electroconvulsive therapy, 3:55
- Constant compliance regime, 1:351
- Constant flow infusion method, 4:4–5
- Constant phase angle impedance, at electrode–electrolyte interface, 1:125
- Constant pressure infusion method, 4:4
- Constant temperature heating technique, 1:193–194
- Constipation, manometric features of, 1:65–66
- Constipation pain, biofeedback clinical outcome literature related to, 1:180
- Constitutive laws, 1:88
- Contact impedance, at electrode–skin interface, 1:121
- Contacting motion sensors, in neonatal respiratory monitoring, 5:16
- Contact lenses, 1:277; 2:321–329
astigmatic designs in, 2:326
care of, 2:324
corneal physiology and response to, 2:324–325
designs for presbyopia and monovision, 2:326–327
fitting of soft contact lenses, 2:323
fitting philosophy for rigid gas permeable contact lenses, 2:323
history of, 2:321–322
optics and design of, 2:322
for orthokeratology, 2:327
poly(methyl methacrylate)-rigid gas permeable design for, 2:322–323
safety of, 2:325–326
soft, 2:323
wear schedules for, 2:323–324
- Contact potential, at electrode–skin interface, 1:121
- Contact units, for radiation therapy, 6:583
- Container electrode, 1:143

- Contamination, of polymerase chain reaction, 5:381–382
- Content-Addressed Storage (CAS), 5:350
- Content-based image retrieval (CBIR), 4:359
- Contingent negative variation (CNV), 3:237
- Continuous BP monitoring, 5:235–237.
See also Blood pressure entries
- Continuous flow ventricular assist devices, 3:454
- Continuous intraarterial pO_2 measurement, 5:212–213
- Continuous intravascular blood gas monitoring (CIBM), 1:474–476
clinical uses for, 1:475
limitations and complications of, 1:475
- Continuous intravascular neonatal blood gas/biochemical sensors, 4:591–592
- Continuous phase composite (CPC), 1:263
- Continuous positive airway pressure (CPAP), 1:38; 2:329–336; 6:508, 511–512
advantages of, 2:336
comfort/compliance issues in, 2:335–336
definitions related to, 2:330t
indications for use of, 2:334–335
leak circuit modification in, 2:329–330
monitoring/titration issues related to, 2:332–334
in obstructive sleep apnea/hypopnea syndrome (OSAHS), 2:329–332
variations in, 2:330–332
- Continuous variables
parametrical hypothesis testing on, 6:251
probability density function for, 6:244
- Continuous wave Doppler (CW), 3:1
- Continuous wave Doppler ultrasound (US), in peripheral vascular noninvasive measurements, 5:241–242
- Contraception, research and development related to, 2:347–348
- Contraceptive devices, 2:336–349
bilateral tubal sterilization, 2:346–347
cervical cap, 2:340–341
efficacy of, 2:337–338
female condom, 2:339–340
informed consent concerning, 2:338
intrauterine devices, 2:345–346
Lea's shield, 2:341
male condom, 2:338–339
transdermal contraceptive patch, 2:343–344
vaginal spermicides, 2:340
- Contraceptive diaphragm, 2:342–343
- Contraceptive implants, 2:344–345
- Contraceptive patch, transdermal, 2:343–344
- Contraceptive sponge, 2:341–342
- Contraceptive vaginal ring, 2:344
- Contractility, of the heart, 4:567
- Contraction, of muscles, 1:391–392
- Contrast, in computed tomography, 2:252–253
- Contrast bath hydrotherapy, 3:464
- Contrast dye echocardiography, 2:38
- Contrast enhancement, in computed radiography, 5:340
- Contrast imaging, 6:465–471
in echocardiography, 3:23
- Contrast perturbation, 1:610
- Contrast ratio, in X-ray systems, 6:572
- Contrast/resolution
in computed radiography systems, 5:339
in confocal microscopy, 4:461–462
- Control advance moving average controller (CAMAC), 1:493
- Control interface, in augmentative and alternative communication systems, 2:204
- Controlled drug delivery
principles of, 2:437–440
various approaches to, 2:438–440
- Controllers, for near-field scanning optical microscopy, 4:438–439
- Control mechanisms, for movement, 1:385–386
- Control panel, in X-ray equipment, 6:569
- Convection hydrotherapy, 3:468
- Conversion gain, in X-ray systems, 6:572
- Convolutions, in CT reconstruction
methods, 2:247–248
- Convolution–superposition radiation dose calculation, 5:459
- Convolution theorem, 2:248–249
- Cooled thermal detectors, 6:349–350
- Coordinated movement, 1:385–386
- Coordinate systems, in joint biomechanics, 4:209
- Cooximetry, versus oximetry, 1:470–471
- Copolymers, 1:332; 5:388
- Copper–nickel–titanium alloys, in dental prosthetics, 1:325
- Coral, as biomaterial, 1:272–273
- Coral-derived apatite, in tissue engineering, 1:374
- Cordis Checkmate system, 1:602–604
- Cornea, physiology and response to contact lens wear, 2:324–325
- Cornea ablation, 6:378
- Corneal reflection recognition, 3:270
- Corneocytes, 1:131
- Coronary angioplasty, 2:349–360
percutaneous transluminal, 4:542
- Coronary artery bypass
minimally invasive, 4:541–542
predicting improved survival with, 3:258
- Coronary artery disease, 2:50–51
- Coronary artery replacement, 6:394
- Coronary atherosclerosis, vascular graft prostheses and, 6:491–493
- Coronary autoregulation, in whole-body models, 5:307
- Coronary circulation, 3:482
- Coronary interventions, percutaneous, 3:258
- Coronary micro-syringe, 2:503
- Coronary systems modeling, 5:316
- Corpectomy models, 3:573
- Correction-based radiation dose calculation, 5:458–459
- Correlation analysis, EEG, 3:69–70
- Correlations, 6:256–257
- Correlation waveform analysis (CWA), 1:70
in separating ventricular fibrillation from tachycardia, 1:79
template matching by, 1:76
- Correlation waveform analysis, 1:74
- Corrosion
biomaterial failure from, 1:278
of biomaterials, 1:308–314
- Corrosion fatigue (CF), 1:311
- Corrosion rate, of metallic biomaterials, 1:309, 311
- Corrosion resistance
of metallic biomaterials, 1:104–105
of nickel–titanium shape memory alloys, 1:7–8
- Corrosive wear, 1:315
- Cortical approach, to developing visual prostheses, 6:535–536
- Cortical bone, 1:524
- Cortical injury monitor, 5:35
- Corticosteroids, 4:273–274
- CORVUS system, 6:397
- Costs, of high-dosage-rate brachytherapy, 1:599–600
- Cough reflex, 1:64
- “Coulter Counter,” 1:23, 24
- Counseling, in speech intervention, 2:218
- Counterpulsation principle, 4:164
- Count rate performance, Anger camera, 1:60
- Coupled motion, in scoliosis, 6:126
- Coupling efficiency, in illumination fibers, 3:306
- Coupling medium, in lithotripsy, 4:260
- Cournand, Andre Frederick, 5:430
- Covalent modification, of biomaterials, 6:388–389
- Craniofacial research, strain gages in, 6:287–288
- Craniospinal system, hydrodynamics of, 4:3
- Craniotomy, stereotactic, 6:269
- Crawling, by mammalian cells, 1:345
- CRCPD model regulations, 2:172–173
- Creep, of bone cement, 1:546–547
- Creutzfeldt–Jakob disease (vCJD), 1:513, 517
- Crevice corrosion cells, 1:310
- Critical Assessment of Fully Automated Structure Prediction (CAFASP), 1:221
- Critical Assessment of techniques for protein Structure Prediction (CASP), 1:221
- Critical care analyzers, 1:21–22
- Critical care instruments, 1:22t
- Critical cell path length, in tissue regeneration, 6:185–186
- Critical stress intensity factor (K_{1c}), of bone, 1:530
- Crossed cylinder wear geometry, 1:316
- Cross-linked polyethylene hip joints, 3:521
- Cross-linked polymers, as biomaterials, 1:274
- Cross-linking
in elastomers, 1:332–333
in thermosets, 1:332

- Cross-linking (*Continued*)
 in UHMWPE, 1:334–335
 of ultrahigh molecular weight polyethylene, 1:316–317
- Cross-spectral analysis, EEG, 3:70–71
- Crosstalk minimization, in ocular motor recording, 5:146–147
- Crowns, dental, 1:325–329
- Crush injury, hyperbaric medicine and, 4:25–26
- Cryoablation (cryotherapy, cryosurgery), 6:366–367, 371, 374, 376
- Cryobiology, 1:188–189, 192
- Cryogenics, 1:236
- Cryopreservation, 1:192
 of engineered tissue, 3:201–202
- Cryoprotective agents (CPAs), 1:192
- Cryosurgery, 2:360–378
 apparatus associated with, 2:363–366
 clinical uses for, 2:372–376
 comparison with other treatment methods, 2:376
 effect on tissue, 2:362–363
 future directions in, 2:376–377
 historical developments in, 2:361–362
 monitoring technique for, 2:368–372
 techniques for, 2:366–368
- Cryosurgical device, 2:365f
- Cryotherapy, 3:474–475
 history of, 3:463–464
- Crystalline carbons, 1:296–297, 298f
- Crystalline polymers, 1:314
- Crystallinity, polymer degradation via, 1:257–258
- Crystal structures, of nickel–titanium shape memory alloys, 1:3–4
- CsCl-type structure, for Ni–Ti shape memory alloy, 1:3–4
- CT image slice thickness, 6:577
- CT imaging system, Hi-ART II, 6:399–401
- CT perfusion imaging, 2:242–243
- CT reconstruction techniques, 2:246–250
- CT scan. *See also* Computed tomography (CT)
 axial, 2:237–238
 dynamic, 2:238
 helical, 2:238
- CT scanners
 components of, 2:234–237
 primary technical parameters of, 2:269t
- CT simulators, 5:527, 586–587
- Cuff electrodes, 1:157
- Cuff oscillometry, 1:14
- Cultured epithelial autograft, 6:190–191
- Curable prosthetic intervertebral nucleus (PIN), *in situ*, 3:586
- Current. *See also* Currents
 at electrode-electrolyte interface, 1:126–127
 endogenic ionic, 1:199
 high frequency, 3:157
- Current-clamp technique, 3:142–143
- Current density, distribution under an electrode, 1:144–146, 148
- Current density hotspots, 1:121
- Currents. *See also* Current
 brain, 1:238–239
 extracellular, 3:110–111
 of linearity, 1:128–129
 Current sources, in temperature measurement electronics, 6:327
- Cushion form hip joint bearings, 3:522
- Cutaneous afferent feedback
 for correction of dropfoot, 3:126–127
 for restoration of hand grasp, 3:127
- Cutaneous blood flow, Doppler
 measurement of, 2:378–384. *See also* Laser Doppler flowmetry
- Cutting processes, electrosurgical, 3:169–170
- CW Doppler assessment, 5:243
- Cyanide poisoning, hyperbaric medicine and, 4:27
- Cyanine dyes, 4:468
- CyberKnife robotic radiosurgery unit, 5:578
- Cycle length (CL) values, in rate-based arrhythmia detection, 1:71
- Cyclodextrins (CDs), 2:452–454
 as drug delivery systems, 2:454–460
 elimination of, 2:454
 toxicological profile of, 2:454
- Cyclotron radionuclide production, 5:92
- Cyclotrons, regulations related to, 2:175
- Cylindrical model, of a vessel segment, 1:199–200
- Cylindrical tubes, blood rheology studies in, 1:504–505
- CYPHER stent, 1:278
- Cystic fibrosis sweat test, 2:384–388. *See also* Quantitative Gibson–Cooke Pilocarpine Iontophoresis Sweat Test (GCST/QPIT)
 history of, 2:385
 state of the art of, 2:386–387
- Cystoscopy, 3:183
- Cytochemical probes, 2:390–392
- Cytochemistry, 2:411
- Cytokine-release system, for tissue engineering, 1:375–377
- Cytokines, in whole blood, 1:460
- Cytology. *See* Automated cytology; Cell entries
- Cytoplasmic granularity, 2:411
- Dacron, 1:338
 in cardiovascular applications, 1:276
 vascular graft prostheses, 6:494
- Daily activity and sleep home health care devices, 3:534
- Dalton's law, 1:465; 2:14; 4:19
- Danazol, 6:154
- Danmeter AEP Monitor/2, 4:561–562
- Data, in the biomedical laboratory, 2:308–310
- Data acquisition
 in automated cytology, 2:400–403
 in single photon emission computed tomography, 2:278–280
- Data acquisition system
 in CT scanners, 2:235–237
 in neurological monitors, 5:34
- Data analysis
 for DNA microarrays, 4:366
 in joint biomechanics, 4:209–211
 for prosthetic heart valve testing, 3:431–432
- Database management system (DBMS), 4:355, 356–357
- Databases, for automated arrhythmia detection, 1:81
- Data collection
 in exercise training, 1:394–395
 in kinematic analysis, 4:207–209
- Data conferencing, in office automation systems, 5:158–159
- Data knife, 4:529
- Data manipulation software, for communication disorders, 2:213–214
- Data migration, 5:348–349
- Data presentation/analysis, for impedance spectroscopy, 4:134–135
- Data reformatting, in teleradiology, 6:305
- Data restoration, using deconvolution, 4:522
- Data samples
 in statistical methods, 6:240–241
 types of, 6:241
- Data storage, in phonocardiography, 5:287
- Data transformation, in exercise training, 1:395–396
- Davey, Humphrey, 5:430
- DaVinci, Leonardo, 5:429
- Daytime sleepiness, 6:212
- d-Be (deuteron-beryllium) reaction, in neutron production, 5:58–59
- DC offset voltage standards, 1:158, 161
- dc power, in linear variable differential transformers, 4:254–255. *See also* Direct current entries
- d-D (deuteron-deuterium) reaction, in neutron production, 5:60
- Decision assistance, anesthesia-related, 1:50
- Decision support, computerized, 4:575
- Decision Support System (DSS), 5:151
- Decompression illness (DCI), 4:23
- Deconvolution, imaging artifacts and data restoration using, 4:522
- Deep brain stimulation (DBS), 3:27
- Deep heating devices, 3:470
- Deep reactive ion etching (DRIE), 2:5
- Deep second degree (deep partial thickness) burns, 6:170–171
- Deep zone, of articular cartilage, 2:64, 65
- Defecation, 1:67
 rectoanal pressure changes during, 1:64
- Defender, 1:77
- Defibrillation, 2:49. *See also* Defibrillators
 cardiopulmonary resuscitation via, 2:36–37
 mechanisms of, 2:409
 open chest, 2:37
 terminal arrhythmia and, 2:47
- Defibrillation electrode, 1:144f
- Defibrillation overload recovery standards, 1:159
- Defibrillation recovery standards, 1:160–161
- Defibrillators, 2:406–410
 atrial, 1:80

- automatic external, 1:79–80; 2:56–57
 external, 2:406–407
 implantable, 2:407–409
 implantable cardioverter, 1:69–70, 71–79
 publicly available, 2:49
 standards for, 1:160–161
 Deformity, orthotics and, 6:80
 Degenerate polymerase chain reaction, 5:384
 Degeneration–trauma, spinal motion changes due to, 3:562–566
 Degenerative bone disease, implants for, 6:234
 Degenerative disk disease, 6:232–233
 implants for, 6:234–237
 Degradable polymeric biomaterials, 1:279
 Degradation
 biomaterial failure from, 1:279
 of biomaterials, 1:308–309
 of polymers, 1:314
 Degradation byproducts, of porous biomaterials, 5:398–399
 Degree of polymerization (DP), 1:331
 Degrees of freedom (DOF), 6:246, 255–256
 of joints, 4:205–207
 Dehydration rehydration vesicles (DRVs), 2:469
 Delivery catheter
 for Guidant Galileo IVB system, 1:606–607
 in IVB devices, 1:603
 for Novoste BetaCath, 1:604
 Delrin, 1:338
 Delta waves, EEG, 3:66
 Dementia, 2:211
 cognitive training for persons with, 6:73–74
 Demineralized bone particles, in tissue engineering, 1:374–375
 Demodulation, in linear variable differential transformers, 4:253–254
 Density, of bone, 1:529–530
 Dental amalgam, 1:322–323
 Dental arches, 6:412f
 Dental bridges, 6:425
 Dental ceramic materials, classification of, 1:326t
 Dental electrosurgery, 3:160
 Dental fillings, 1:322–325
 Dental implants, 1:327–329; 6:426–427
 Dental research, strain gages in, 6:287–288
 Dental restorative filling materials, resin-based composite, 6:93–99
 Dentin, 1:324
 Dentistry
 adhesives, cements, and liners in, 1:324–325
 biomaterials for, 1:322–329
 prosthetic restorative materials in, 1:325–329
 Dentition, 6:411–414
 Denture base materials, 1:327
 Denture materials, 1:325–329
 Dentures
 fixed partial, 6:425
 removable, 6:425–426
 Deoxyhemoglobin, 2:40
 Department of Defense (DOD), 2:155
 Department of Energy (DOE), 2:155
 Department of Transportation (DOT), 2:155
 Departments, in a total hospital safety program, 6:116
 Depolarization width, in feature-based arrhythmia algorithms, 1:76
 Depth monitors, in anesthesia, 1:44
 Derivative area method (DAM), 1:74–75
 Derivatization, 2:193
 Derjaguin approximation, 1:351
 Dermal drug delivery, cyclodextrins in, 2:455–456
 Dermal regeneration template (DRT), 6:193–196
 Dermal replacement, living, 6:191
 Dermis, 1:131
 morphology and function of, 6:188–189
 properties of, 6:169–170
 Dermis components, involved in healing and skin substitutes, 6:172t
 Descriptive statistics, 6:241–243
 Desflurane vaporizers, anesthesia machine, 1:36
 Desktop publishing, in office automation systems, 5:154
 Desorption, of polymers, 1:314
 Detection system, polymerase chain reaction as, 5:384. *See also* Computer-assisted detection/diagnosis (CAD)
 Detective quantum efficiency (DQE), 5:338; 6:556
 Detector arrays, 5:482
 Detector calibration, in X-ray therapy equipment, 6:585–586
 Detectors, for three-dimensional dosimetry, 5:482–484
 Detrusor stimulation, 1:431, 432
 Developmental bone deformities, spinal, 6:231–232
 Developmental language disorders, 2:211
 Developmental spine deformities, implants for, 6:234
 Deviant sexual arousal, 6:159–161
 assessment of, 6:159–161
 Device calibration, in microdialysis sampling, 4:405–409
 Device–device interaction, as a hospital problem, 6:113
 Device operators, as a hospital problem, 6:112
 Dewars, 1:236, 237
 Diabetes, acute coronary events and, 2:50.
 See also Glucose sensors
 Diabetes mellitus, laser Doppler flowmetry for, 2:382
 Diabetic skin ulceration, biomechanics of, 6:206–207
 Diagnosis. *See also* Computer-assisted detection/ diagnosis (CAD); Diagnostic entries
 fluorescence techniques in, 6:480–481
 of implant-related infection, 1:116
 uses of ultraviolet radiation in, 6:480–482
 Diagnosis assistance, anesthesia-related, 1:50
 Diagnostic applications, of optical sensors, 5:172–173
 Diagnostic audiometer, 1:93f
 Diagnostic biotelemetry microsystems, 1:423
 Diagnostic computer-based ECG systems, 3:41–50
 Diagnostic electron microscopy, 4:484–486
 Diagnostic imaging, phantom materials in, 5:266–267
 Diagnostic probes, monoclonal antibodies as, 4:601–604
 Diagnostic radiological physics books/ reports, 4:341–343
 Diagnostic recordings, in neonatal monitoring, 5:30
 Diagnostics
 with angioplasty catheters, 2:354–355
 biomaterial surfaces and, 1:343
 use of microarrays in, 4:370–371
 Diagnostic X-ray units, requirements for, 2:176t
 Dialysates
 fluorescence for, 4:411
 mass spectrometry for, 4:411
 Dialysis, 1:212–213
 Diamond, 1:296–297
 Diamond detectors, 5:477
 Diamond-like carbon (DLC), 1:300, 305, 313–314
 Diamond-like carbon coatings
 alloyed with metals, 1:319–320
 on orthopedic prostheses, 1:318–320
 Diamond-like carbon encapsulation, 1:157
 Diaphragm, contraceptive, 2:342–343
 Diaphragmatic hernia, congenital, 3:506
 Diarthrodial joints, 2:63
 motion and forces on, 4:212
 Diaspirin Cross-Linked Hemoglobin (DCLHb), 1:517, 518
 clinical trial of, 1:519
 Diastolic pressure, 1:485
 Diathermy
 microwave, 3:472
 pulsed short-wave, 3:471–472
 short-wave, 3:470–472
 Dichroic mirror, 3:346
 DICOM message, 5:555t. *See also* Digital Imaging and Communications in Medicine (DICOM) standard
 Dieulafoy vascular malformation, 3:390
 Differential amplifier, EEG, 1:172
 Differential counts. *See* Automated differential counts
 Differential current density (DCD) electrode, 1:154
 Differential equations, in biological network analysis, 1:226
 Differential leukocyte count (DLC), 2:87
 Differential lysis, in white blood cells, 2:411
 Differential preamplifier, 3:113–114
 Differential pressure oxygen analyzers, 5:201–202

- Differential scanning calorimetry (DSC) measurement, of SMA transition temperature, 1:7
- Differentiated thyroid cancer, radioiodine therapy dosimetry for, 5:569
- Diffraction lenses
in intraocular lenses, 4:238
optical quality of, 4:239
- Diffuse optical tomography (DOT), 5:247
- Diffusible factors, in engineered tissue, 3:199
- Diffusing capacity, lung, 5:438–439
- Diffusion imaging, 4:290–291
- Diffusion impedance, at electrode-electrolyte interface, 1:125
- Diffusion tensor imaging (DTI), 4:291; 6:270
- Diffusion-weighted pulse sequences, 4:291
- Digital angiography, 2:421–426
future of, 2:426
history of, 2:421–422
- Digital archival media, 5:347
- Digital cameras, medical uses for, 5:296
- Digital colposcopy, 2:201–202
- Digital communication systems, in office automation systems, 5:155–156
- Digital ECG recorder, 1:13
- Digital examination, for anorectal manometry, 1:63
- Digital filtering, in EEG biofeedback instrumentation, 1:171
- Digital image archival, 5:346–347
- Digital image quality, 5:338–339
- Digital image receptors, quality control of, 6:578–579
- Digital imaging, in medical photography, 5:293
- Digital Imaging and Communications in Medicine (DICOM) standard, 5:119, 527, 332, 333, 335, 555
- Digitally reconstructed radiographs (DRRs), 5:456
- Digital mammography, full field, 4:303–304
- Digital polysomnography, 6:212–213, 219
- Digital radiograph, preview, 2:237
- Digital radiography (DR), 5:343–346
compared with computed radiography, 5:345–346
system characteristics in, 5:344
- Digital reconstructed radiographs (DRRs), 5:586–587
- Digital recorders, 6:49–51, 59–60
- Digital spot mammography, 4:303
- Digital subtraction angiography (DSA), 2:422–425
- Digital video, 5:144–146
- Digital video disk (DVD), 5:348
- Digitization, 5:112
- Digitized signal transmission, 1:420
- Digitizing, in exercise training, 1:394–395
- Dilators, vaginal, 6:153–154
- Dilution signal, decay of, 2:29
- Dimethyl siloxane, in silicone rubbers, 1:337
- Dipping, circadian blood pressure variability and, 1:15
- Direct-acting (label-free) IMFET, 4:100–102
workings of, 4:103–104
- Direct blood pressure measurements, 4:569
- Direct colorimetric measurement, 2:192
- Direct Conversion digital radiography, 5:343
- Direct coronary artery bypass, minimally invasive, 4:541–542
- Direct current (dc), in cardiopulmonary resuscitation, 2:36–37. *See also* DC offset voltage standards; dc power
- Direct current osteogenesis, 1:560–561
- Direct current SQUIDs, 1:231–232
- Direct Fourier reconstruction, 2:247
- Directives, Nuclear Regulatory Commission, 2:171
- Direct kinetic measurement, 2:194
- Direct Linear Transformation, 1:395
- Direct model reference adaptive controller (DMRAC), 2:505
- Direct ultrasound ablation, 6:365
- Disasters, as a hospital problem, 6:114
- Discrete variables, testing hypothesis on, 6:248–251
- Discrimination learning, 1:167
- Discrimination training, 1:177
- Discussion boards, in office automation systems, 5:155
- Disease
effects on gas transport, 6:106–107
implant failure and, 1:112
- Disease progression, drug effectiveness and, 5:273
- Disease states, respiratory mechanics in, 6:106
- Disease treatment, using systemic hyperthermia, 4:58–59
- Disinfection, 6:279t
use of UV in, 6:486–487
versus sterilization, 6:274
- Disk disease
degenerative, 6:232–233
implants for, 6:234–237
- Disk electrode field, in electrosurgery, 3:174
- Disk replacement, total, 6:237
- Disk sensor, in solid electrolyte cell oxygen analyzers, 5:205
- Dispersed phase, of resin-based composites, 6:95
- Displacement, in exercise, 1:392
- Displacement magnetometers, in neonatal respiratory monitoring, 5:16
- Display, ultrasound, 3:10–13
- Display systems, for lung sounds, 4:279–280
- Display workstation, teleradiology, 6:305
- Disposable drug infusion pump, 2:502
- Disposable electrodes, 1:148
modern, 1:139–141
- Disposal, of radioactive material, 2:170–171
- Dissipation constant (DC), in thermistors, 6:321
- Dissolution, of bioactive glasses, 1:286–287
- Distal catheters, for hydrocephalus, 4:12
- Distortion product otoacoustic emissions (DPOAEs), 1:101, 102
- Disuse syndrome, 1:389–390
- Divergence errors, in computed tomography, 2:257
- Diverticulosis, 3:391–392
- DNA. *See also* DNA sequencing
ancient, 5:384–385
contamination of, 5:382
molecule, 2:427
proteins, RNA and, 1:217
scanning tunneling microscopy and, 4:517–519
- DNA arrays. *See also* DNA microarrays
ex situ fabrication of, 4:367
in situ fabrication of, 4:366–367
- DNA binding agents, 1:574–575
- DNA-computing-based DNA sequencing, 2:435
- DNA detection, with nanoparticles, 2:434
- DNA measurements, automated cytology for, 2:403
- DNA microarrays, 1:223–224, 225f. *See also* DNA arrays
- DNA polymerase reaction, 5:380–381
- DNA sequence analysis software, 2:432t
- DNA sequencing, 2:427–437
with atomic force microscopy, 2:435
commercial state of the art in, 2:431–433
evaluating techniques of, 2:431
by fluorescence microscopy, 2:435
future of, 2:434–435
gel electrophoresis and, 2:427–428
by hybridization, 2:433
mass spectrometry based, 2:434
at the nanoscale, 2:434–435
principles of, 2:428–431
types of, 2:428
- DNA sequencing equipment, evaluating, 2:431t
- Documentation, in biomedical equipment maintenance, 3:227–228
- Dominant frequency/power, EGG, 3:90
- Donation after cardiac death (DACD), 4:270
- Donnan osmotic pressure, 2:64
- Doppler echocardiography, 3:4, 6–10
clinical uses of, 3:20–22
- Doppler effect, 3:1
- Doppler imaging, 6:462–465
color flow, 6:464–465
power, 6:465
pulse wave, 6:463–464
- Doppler measurement, of cutaneous blood flow, 2:378–384
- Doppler principle, 2:378
- Doppler ultrasound (US)
continuous wave, 5:241–242
in prosthetic heart valve testing, 3:432–434
transabdominal, 3:290–291
- Doppler ultrasound flow measurement, 3:330–332
- Doppler volume flow meters, 3:327
- Dose calculation
in high-dosage-rate brachytherapy, 1:597–598
radiation, 5:458

- in three-dimensional conformal radiotherapy, 6:34
- Dose calculation functions, relationships between, 2:130–131
- Dose conversion, between materials, 5:471
- Dose display, in radiation dose planning, 5:462–463
- Dose equivalent, 5:467
- Dose fractionation, effect of, 2:260
- Dose functions, cobalt-60 unit, 2:130–131
- Dose gradient, 5:582
- Dose homogeneity, 5:583
- Dose modifying devices, 5:594–599
- Dose optimization, in high-dosage-rate brachytherapy, 1:598
- Dose prescription, in three-dimensional conformal radiotherapy, 6:33
- Dose–volume histogram analysis, 5:547
- Dose–volume histograms (DVHs), 5:462–463, 580–581
- Dosimeter characteristics, required for quality assurance, 5:482
- Dosimeters
 - calibration of, 5:474–476
 - chemical, 5:474
 - Fricke, 5:473
 - personnel, 5:517t
 - reference or secondary, 5:471–474
 - thermoluminescent, 5:473–474
 - X-ray, 6:585
- Dosimetric formalism, 5:421, 423
- Dosimetry. *See also* Radiation dosimetry
 - BetaCath IVB system, 1:605
 - in boron neutron capture therapy, 1:572–573
 - Checkmate IVB system, 1:604
 - experimental, 1:614–615
 - gel, 5:484–488
 - in the Guidant Galileo IVB system, 1:608–609
 - of ¹³¹I-MIBG therapy, 5:569–570
 - internal, 5:493
 - in intraoperative radiotherapy, 6:23
 - in new drug development, 5:571–572
 - in prostate seed implants, 5:420–424
 - in radioimmunotherapy, 5:570–571
 - for radionuclide therapy, 5:567–571
 - radiopharmaceutical, 5:565–574
 - theoretical, 1:613–614
- Dosimetry calibration, in X-ray therapy equipment, 6:584–586
- Dosimetry systems, quality assurance procedures requiring, 5:481–482
- Double blind clinical trials, 6:262
- Double-layer capacitance, 1:123
- Drive mechanism, ventilator, 6:506
- Dropfoot, correction of, 3:126–127
- Drug carriers
 - ethosomal, 2:477
 - liposomal, 2:466–473
 - niosomal, 2:473–477
 - particle, 2:480–486
 - ultradeforvable vesicular, 2:479–480
 - vesicular, 2:466–480
- Drug–cyclodextrins complex, preparation of, 2:453–454
- Drug delivery
 - biomaterials in, 1:277–278
 - controlled, 2:437–440
 - intravenous, 2:448
 - oral, 2:446–447
 - polymers for, 5:390–391
 - smart polymers for, 1:340
 - stimuli-responsive hydrogels for, 5:391
 - supramolecular aggregates for, 2:460–464
 - transdermal, 2:442–446
 - use of nanoparticles in, 5:5
- Drug delivery devices
 - colloidal, 2:464–486
 - economics of, 2:440
 - implantable, 2:439–440, 440–442
 - nanoengineered, 2:504
- Drug delivery microchip, 1:424, 425f
- Drug delivery systems, 2:437–495. *See also* Drug delivery devices
 - biodegradable polymeric, 2:504
 - development of, 2:438
 - future perspectives on, 2:486
 - implantable microfabricated, 2:504
 - microelectro-mechanical, 2:440–452
 - microemulsions as, 2:461–462
 - microflow regulator for, 2:504
 - molecular, 2:452–460
 - principles of controlled drug delivery, 2:437–440
- Drug discovery, optical biosensors in, 5:173
- Drug effect control, 1:48–49
- Drug effectiveness, disease progression and, 5:273
- Drug infusion pump, disposable, 2:502
- Drug infusion systems, 2:495–508
 - advancements in controller design for, 2:504–507
 - common, 2:496–502
 - new developments in, 2:502–504
 - tiny, 2:502–503
- Drug/lead discovery, use of microarrays in, 4:370
- Drug-resistant hypertension, 1:15
- Drugs
 - exposure-effect link in, 5:272–273
 - implant failure and, 1:112
 - monitoring in anesthesia, 1:44, 48
- Drug screening, cell-based, 4:140–142
- Drug testing, microbioreactors for, 4:393
- Drug toxicity, as a complication of
 - parenteral nutrition, 5:128
- Dry electrodes, 1:134
- Dry heat sterilization, 6:275
- Drying, of foam scaffolds, 1:293
- DTA profile, 1:359
- d-T (deuteron-tritium) reaction, in neutron production, 5:60–61
- Dual-chamber arrhythmia detection, 1:77–79
- Dual-chamber pacemakers, 1:77
- Dual-energy scanning, 2:244
- Dual-energy X-ray absorptiometry, 1:552–553
- Dual-parameter histograms, 2:401
- Dual-photon absorptiometry, 1:551–552
- Dual Purkinje image measurement, 3:277–278
- Duchenne, Guillaume, 1:143, 429
- Duke activity scale index (DASI), 3:252t
- Dumbbell oxygen analyzers, 5:201
- Duplex Doppler ultrasonography, to assess genital engorgement, 6:153
- Duplex scanning, 3:1
- Duration of movement, in exercise, 1:392
- Dye-dilution method, 2:32
- Dyes
 - Alexa Fluor, 4:467–468
 - fluorescent, 4:466–467
 - with general chemistry analyzers, 1:20
- Dynamic acoustic immittance, 1:100–101
- Dynamic area telethermometry technique, 6:352
- Dynamic compression, pulmonary, 5:432
- Dynamic CT scan, 2:238
- Dynamic lung compliance, 6:517
- Dynamic programming approach, to sequence alignment, 1:218–219
- Dynamic range control, 5:341–342
- Dynamic response, checking, 4:570–572
- Dynamic SIMS, 1:350
- Dynamic thermatomes, 6:349
- Dynamic tracers, 6:430
- Dysarthria, 2:211
- Dyshemoglobins, 5:211–212
- Dyslexia, 2:211
- Dysrhythmias, cardiac arrest, 2:44–48
- Dyssynergia, 1:68
- EADL systems. *See also* Electronic aids to daily living (EADLs)
 - feature control in, 3:212–213
 - subsumed devices in, 3:213–214
- Ear. *See also* Hearing entries; Oto- entries
 - anatomy of, 1:94
 - disorders of, 1:94–95, 96
 - sensorineural mechanism of, 1:94–95
- Ear oximeter, 1:471
- Ear oximetry, 1:471
- Ear thermometers, infrared, 6:361
- e-beam irradiation, 6:278
- Eccentric contraction, 1:392
- ECG data analysis program, 3:46–47. *See also* Electrocardiograms (ECGs)
- ECG data flow/storage, 3:45–46
- ECG electrodes, standards for, 1:158–160
- ECG machine, early, 1:137f
- ECG paper recording, in exercise stress testing, 3:250
- ECG recording, in exercise stress testing, 3:248. *See also* Echocardiogram (ECG) recorder
- ECG recording instruments, in exercise stress testing, 3:249
- ECG sensors/amplifiers, 1:175
- ECG signals, instrumentation to record, 3:39–41
- ECG signal telemonitoring, 3:51
 - ECG systems, diagnostic computer-based, 3:41–50
- Echocardiogram (ECG) recorder
 - cassette-type, 1:13
 - portable, 1:13
- Echocardiographic examination, 3:14–18
- Echocardiographic instrumentation, 3:5–6

- Echocardiography, 2:38; 3:1–24, 481f
 clinical formats of, 3:2–3
 clinical uses of, 3:19–22
 Doppler, 3:4, 6–10
 in exercise stress testing, 3:255
 principles of, 3:3–5
 signal processing, display, and management in, 3:10–13
 specialized clinical data related to, 3:22–23
 two-dimensional, 3:20
- ECT device specifications, 3:58t. *See also* Electroconvulsive therapy (ECT)
- ECT stimulus, 3:58–59
- Edema, impedance plethysmography and, 4:128
- Edge effects, with metal electrodes, 1:146–148
- EDGE system electrodes, 1:147
- Edison, Thomas, 1:299
- Education
 biomedical engineering, 1:403–409
 medical device, 6:119
 nuclear medicine, 5:122
- Educational assessment, for augmentative and alternative communication systems, 2:208
- EEG amplifiers, 1:172. *See also* Electroencephalograms (EEGs); Electroencephalography (EEG)
- EEG analysis
 inter-user variability in, 3:68
 techniques of, 3:69–72
- EEG biofeedback (neurofeedback), biofeedback clinical outcome literature related to, 1:181–182
- EEG biofeedback instrumentation, 1:171–173
- EEG burst and analyzing methods, 3:72–80
- EEG differential amplifier, 1:172
- EEG displays, 1:172
- EEG electrodes, 1:171–172
- EEG index-based neurological monitors, 5:35
- EEG monitoring, scientific basis for, 3:67–68
- EEG monitors
 classification of, 5:32–33
 types of, 5:35–39
- EEG potentials, generation of, 3:65
- EEG rhythms, 4:557t; 5:33t
- EEG signals, 3:65–67
- Effective atomic number method, of formulating tissue substitutes, 5:256
- Effective blood flow (EBF), 2:16–17
- Effective dose, 5:505
- Effective energy, X-ray beam, 6:599
- Effective orifice area (EOA), of heart valves, 3:415–416
- Effective thermal conductivity (k_{eff}), 1:193
- Effective thermal diffusivity (α_{eff}), 1:193
- Effective thermal property measurements, 1:193–195
 error analysis of, 1:195–196
- Effector T cells
 activation and polarization of, 4:112–113
 redirecting to tumor, 4:116–117
 trafficking and proliferation after adoptive transfer, 4:113–114
- EGG data analysis, 3:88–91. *See also* Electrogastragram (EGG)
- EGG electrodes, 3:88
- EGG frequency range, 3:91–92
- EGG parameters, 3:89–90
 methods to obtain, 3:90–91
- EGG power distribution, percentage of, 3:90
- EGG recording equipment, 3:86–87
- Einthoven, Willem, 1:137
- Ektacytometer, 1:504
- ELA *Defender*, 1:77
- Elam–Safar studies, 2:36
- Elan Medipad technology, 2:503
- Elastic fibers, in ligaments and tendons, 4:242–243
- Elasticity, of bone, 1:529
- Elastic modulus
 of diseased arterial walls, 1:89–91
 of normal arterial walls, 1:88–89
- Elastic resistance strain gages, 6:284–285
- Elastin
 in arterial walls, 1:85
 in skin, 6:204
- Elastohydrodynamic lubrication, 6:417
- Elastomeric infusers, 2:500–501
- Elastomers, 1:331, 332–333
 polyurethane, 1:337
- Elbow radial head (RH) prostheses, 1:304
- Electrical activity, cardiac, 2:43–44
- Electrical anesthesia, 3:27–30
- Electrical effects, electrophoretic, 3:134
- Electrical events, in the cardiac cycle, 4:163
- Electrical field exposure limits, 2:183t
- Electrical impedance plethysmography, 1:121
- Electrical impedance tomography (EIT), 1:121, 134, 203; 4:121; 6:361
- Electrical inspection, in X-ray equipment, 6:563
- Electrically excitable tissues, stimulation during electrosurgery, 3:172–173
- Electrical nerve stimulation, transcutaneous, 3:26–27
- Electrical nerve stimulators
 implantable, 3:26
 transcutaneous, 3:25–26
- Electrical noise, at electrode interface, 1:130
- Electrical oxygen analyzers, 6:524
- Electrical power, as a hospital problem, 6:113
- Electrical resistance, in cellular parameter measurement, 2:396
- Electrical resistance thermometers, 6:356–357
- Electrical resistivity measurement, of SMA transition temperature, 1:7
- Electrical signals, in engineered tissue, 3:199–200
- Electrical stimulation. *See also* Functional electrical stimulation (FES)
 in cochlear prostheses, 2:138
 for scoliosis, 6:130
- Electrical strain gages
 capacitance and inductance, 6:283
 elastic resistance, 6:284–285
 resistance, 6:284
 semiconductor, 6:283–284
- Electric amplitude-frequency selection, in transcutaneous electrical nerve stimulation, 6:449–450
- Electric artificial heart, 3:457–458
- Electric field distributions, in electrosurgery, 3:173–175
- Electricity
 basic terms related to, 1:465–466
 bladder control using, 1:430
 as a hospital problem, 6:111
 muscular stimulation using, 1:429
- Electric motors, in ventilators, 6:507
- Electric ventilators, 6:506
- Electric ventricular assist devices, 3:452–456
- Electroanalgesia. *See also* Neurostimulatory techniques; Systemic electroanalgesia
 evolution of, 6:441–443
 obstetric, 3:31–32
- Electroanatomic mapping system, 6:372
- Electro-Cap™, 1:171
- Electrocardiogram arrhythmia monitoring, 4:568
- Electrocardiogram biopotential signal, 1:175
- Electrocardiogram monitoring, 4:567–568
- Electrocardiograms (ECGs), 1:168; 2:44. *See also* ECG entries; Fetal electrocardiogram (FECG)
 abdominal, 3:291–293
 acute coronary events and, 2:51, 56
 history of, 1:137
 recording at home, 3:527–528
- Electrocardiographic computer systems, 3:34–53
 diagnostic, 3:41–50
 monitoring, 3:50–51
- Electrocardiograph surface electrode testers, 1:160
- Electrocardiography (ECG). *See also* ECG entries
 basics of, 3:35–39
 exercise (stress), 3:47–48
 high resolution, 3:49–50
 12-lead clinical, 3:42
- Electrochemical detection, in microdialysis sampling, 4:410–411
- Electrochemical gas sensors, 4:324–325
- Electrochemical oxygen analyzers, 5:202–206
- Electroconvulsive therapy (ECT), 3:53–62. *See also* ECT entries
 complications of, 3:56
 conditions of increased risk with, 3:56
 consent for, 3:55
 history of, 3:54
 indications for, 3:55
 mechanism of action of, 3:59–60
 medications and, 3:56–57
 monitoring of, 3:57
 pre-ECT evaluation, 3:54–55
 seizure response and, 3:59

- Electrocutaneous electrical stimulation, 6:296
- Electrode arrays, high density, 2:138
- Electrode assembly, in visual prostheses, 6:539
- Electrode–cell interface model, 4:138–139
- Electrode contact impedance standards, 1:162
- Electrode-electrolyte impedance, 1:123–124
- Electrode-electrolyte interface, 1:122–131; 3:106
simple equivalent circuit model of, 1:124–125
- Electrode-electrolyte potential, 1:122–123
- Electrode gels, 1:141
effects of, 1:134–136
- Electrode impedance, in
electroneurography, 3:117–119
- Electrode metals, 1:129–131
- Electrode placement, in transcutaneous electrical nerve stimulation, 6:446–448
- Electrodermal activity (EDA), 1:173, 174
- Electrodermal biofeedback amplifier, 1:173
- Electrodermal biofeedback
instrumentation, 1:173–174
- Electrodermal electrodes, 1:173
- Electrodes. *See also* CO₂ electrodes;
Microelectrodes
basic types of, 6:446t
biomedical, 1:120–166
bipolar and tetrapolar, 1:198–199
blood gas, 1:465–467
circumferential cuff, 3:123–124
in conductance catheter systems, 1:206–207
conductive, 1:145–146, 146–148
current density distribution under, 1:144–146, 148
designing biomedical, 1:120–121, 137–158
dry, 1:134
EEG, 1:171–172
EGG, 3:88
in electrical anesthesia, 3:29
for electrical stimulation, 3:351–352, 354t
electrodermal, 1:173
EMG, 1:169; 3:102–105
external electrostimulation, 1:142–146
extraneural, 3:123–125
flat interface nerve, 3:124
flexible, 2:1–2
garment, 1:149
gel-less, 1:134
hook, 3:125
implanted, 1:120–121, 149–158; 3:353–357
for intraspinal stimulation, 3:357
intra-neural, 3:120–121
longitudinal intrafascicular, 3:111, 121–122
modern designs for, 1:146–149, 152–154
modern disposable, 1:139–141
in neurological monitors, 5:33–34
noble metal, 1:126, 129–131
placement for biofeedback, 1:177–178
regenerating, 3:119
reshaping cuff, 3:124–125
resistive, 1:148–149
sieve, 3:119–120
silicon-based, 3:122–123
skin temperature and, 1:135
slowly penetrating interfascicular, 3:124
solid conductive adhesive, 1:141
standards for, 1:158–162
surface, 3:352–353
testing, 1:133, 134
in transcutaneous electrical nerve stimulation, 6:443–446
wearable, 1:141–142
- Electrode–skin interface, electrical
properties of, 1:122–137
- Electrode testers, 1:160
- Electroencephalogram movement, sexual
arousal and, 6:161
- Electroencephalograms (EEGs),
1:238–239; 5:32. *See also* EEG entries
in anesthesia monitoring, 1:44
- Electroencephalography (EEG), 1:97;
3:62–83. *See also* EEG entries
biofeedback training and, 1:168
burst and analyzing methods for,
3:72–80
clinical, 3:67
early work in, 1:98
fetal, 3:299–300
inter-individual variability in, 3:68
logistical and technical considerations
related to, 3:68
origin of, 3:63–65
problems associated with, 3:68
in sleep laboratory, 6:209–210
volume conduction of, 3:65
- Electro-explosion, nanoparticle fabrication
via, 5:2–3
- Electrogastrogram (EGG), 3:83–98. *See
also* EGG entries
abnormal, 3:92–93
in adults, 3:91–92
clinical role of, 3:93
electrophysiology of the stomach,
3:84–85
future prospects for, 3:94–95
historic review of, 3:84
in infants and children, 3:93–94
measurement of, 3:85–88
- Electrohydraulic generator, 4:259–260
- Electrokinetic effect, 1:139
- Electrolyte. *See* Electrode-electrolyte
entries
- Electromagnet conversion, 4:431
- Electromagnetic brain stimulation, future
directions for, 3:60
- Electromagnetic devices, noninvasive,
1:564–568
- Electromagnetic fields
growth of, 3:543
volumetric heating by, 1:190
- Electromagnetic flowmeters, 3:322–342
volume flow measurement using,
3:324–329
- Electromagnetic flow probes, 3:324–325
- Electromagnetic generator, 4:260
- Electromagnetic heating, in interstitial
hyperthermia, 4:34–35
- Electromagnetic interference (EMI)
as a hospital problem, 6:114
in thermocouples, 6:344
- Electromagnetic interference
environment, with heart rate
variability instrumentation, 1:176
- Electromagnetic lenses, in electron
microscopy, 4:480
- Electromagnetic osteogenesis, 1:561–562
- Electromagnetic radiation
as a hospital problem, 6:111
in incubators, 4:156
versus ultrasound, 4:63
- Electromechanical uncoupling, gastric,
3:85
- Electrometers, for kilovoltage X-ray
therapy measurements, 6:585
- Electromyogram, 6:66
uterine, 3:295–296
- Electromyographic (EMG) activity, 1:168.
See also EMG entries; MEG–EMG
coherence
- Electromyography (EMG), 3:98–109.
See also EMG entries
in erectile dysfunction diagnosis, 6:157
historical perspective on, 3:99–100
in neonatal respiratory monitoring, 5:17
in sleep laboratory, 6:210
vaginal, 6:152
- Electron beam computed tomography,
2:232–233
- Electron-beam radiotherapy, 6:4–6
- Electron beams
beam quality specifiers for, 5:476
in electron microscopy, 4:480
magnetically confined, 6:11–12
- Electron beam–specimen interaction, in
the scanning electron microscope,
4:483
- Electron capture detector (ECD), 4:325
- Electron detection, using the scanning
electron microscope, 4:484
- Electroneurography (ENG), 3:109–132.
See also ENG recording
configurations; Neuro-electronic
interface; Neuroprosthetic
applications
long-term peripheral nerve interfaces in,
3:119–125
- Electroneurostimulation, 3:25
- Electron gun, 4:482
- Electronic aids to daily living (EADLs),
3:210–215. *See also* EADL systems
controlling, 3:214–215
future of, 3:215
- Electronic artificial larynx, 4:230–231
- Electronic attenuation coefficient, 6:593
- Electronic blood cell counters, 2:84–85
- Electronic gas flow sensors, in anesthesia
machines, 1:40–41
- Electronic interfaces, capacitive sensor, 2:7

- Electronic low vision aids, 1:444–445
- Electronic mail, in office automation systems, 5:155
- Electronic Medical Record (EMR), 5:549–550, 551, 552
- Electronic nose, 4:381
- Electronic patient records, 4:310–311
- Electronic portal imaging detectors (EPIDs), 4:89–90; 5:599
applications of, 4:92–96
- Electronic portal imaging device (EPID), 2:506; 5:478, 484
- Electronic portal imaging technology, physical aspects of, 4:90
- Electronic signal processing, of uterine contractions, 3:295
- Electronic stethoscope, 5:288–289
- Electronics thermistors, interfacing with personal computers, 6:332–333
- Electronic travel aids (ETAs)
conventional, 1:448–450
intelligent, 1:450–451
- Electronic vaporizers, in anesthesia machines, 1:41
- Electron lenses, 4:482–483
- Electron microscopy, 4:478–488. *See also* Scanning electron microscope (SEM); Transmission electron microscope (TEM)
diagnostic, 4:484
prospects for, 4:486–488
theory of, 4:479–481
- Electron optical column, in the scanning electron microscope, 4:482–483
- Electrooculogram (EOG), 3:150
clinical applications of, 3:153–154
- Electrooculography (EOG), 3:266–267; 5:140–142
in sleep laboratory, 6:210
- Electrooptic blood cell measurements, 2:83–84
- Electropharmaceutical anesthesia, in long duration microsurgery, 3:32
- Electrophoresis, 2:105–106; 3:132–141
enhancing resolution in, 3:134–138
equipment and procedures in, 3:138–139
evaluation of, 3:139–141
with separation/detection methods, 2:106–107
theory behind, 3:132–138
- Electrophoretic-based DNA sequencing methods, 2:431–433
- Electrophoretic deposition, for porous biomaterial fabrication, 5:401
- Electrophysiologic audiometry, 1:97–99
- Electrophysiology, 3:141–149
cardiac action potential and, 3:143–144
computational model impact in ventricular cells, 3:148
computational modeling of murine ventricular action potentials, 3:146–147
current-clamp technique in, 3:142–143
dissemination of computational models, 3:147–148
mathematical modeling of cardiac cells, 3:144–145
murine cardiac ventricular cell research, 3:145–146
quantifying ionic cell mechanisms, 3:142
resting membrane potential and, 3:141–142
stomach, 3:84–85
voltage-clamp technique in, 3:142
- Electroretinogram (ERG), 3:150–152. *See also* Multifocal ERG (mfERG); Pattern electroretinogram (PERG)
clinical applications of, 3:154–155
Electroretinography, 3:150–156. *See also* Electroretinogram (ERG)
clinical applications of, 3:153–155
electrooculogram and, 3:150
techniques in, 3:150
- Electrosensitive hydrogels, 5:391
- Electro-sensitivity, perceived, 5:69
- Electro-spinning, nanoparticle fabrication via, 5:3
- Electrospinning technique, for porous biomaterial fabrication, 5:400
- Electrostatic conversion, 4:432–433
- Electrosurgery
advanced principles of, 3:170–176
alternate site burns associated with, 3:171–173
dental, 3:160
engineering principles of, 3:161–170
explosion hazard during, 3:172
general, 3:159–160
gynecologic, 3:160–161
hazards and remedies associated with, 3:170–171
interference with instrumentation, 3:173
minor, 3:158–159
representative electric field distributions in, 3:173–175
representative surgical procedures using, 3:164–166
rf generators for, 3:161–164
stimulation of electrically excitable tissues during, 3:172–173
urologic, 3:160
- Electrosurgical cutting processes, 3:169–170
- Electrosurgical devices, standards for, 1:161–162
- Electrosurgical dispersive electrode, 1:147f
- Electrosurgical generators, early, 3:157
- Electrosurgical unit (ESU), 3:156–177. *See also* Electrosurgery
ablation, coagulation, and tissue fusion in, 3:166–169
clinical applications of, 3:157–161
historical background of, 3:157
safety appliances in, 3:163–164
- Electrotactile displays, 6:297
- Electrotactile stimulation, 6:295–297
- Elemental equivalence, in tissue substitute formulation, 5:255–256
- Elevated body temperature, physiological effects of, 4:46–48
- Elgiloy electrodes, 1:130
- Ellipsometry, for characterizing surfaces, 1:352
- Embryonic stem cells
in cartilage regeneration, 2:73
in tissue engineering, 6:383
- Emergency Cardiac Care (ECC) system, 2:48, 51
- Emergency care simulator (ECS), 1:46
- Emergency Medical System (EMS)
personnel, emergency cardiac care and, 2:48–49, 51, 52
- Emergency medical technicians (EMTs), 2:49
- Emergency service instruments, in high-dose-rate remote afterloaders, 1:596
- Emergency switches, in high-dose-rate remote afterloaders, 1:595
- EMG amplifiers, 1:169. *See also* Electromyography (EMG)
- EMG biofeedback
instrumentation for, 1:169–170
with tension and headache, 1:176
- EMG electrodes, 1:169; 3:102–105
choosing, 3:104
configuration of, 3:105–106
locating, 3:104–105
maintenance of, 3:104
- EMG signal, 3:100–102
analysis techniques for, 3:107
detection of, 3:105–107
- Emission computerized tomography (ECT), 5:108
- Emission spectra, measurements of, 4:498–500
- Emission wavelengths, 3:346
- EMIT enzyme immunoassay, 1:21
- Emphysema, pulmonary interstitial, 3:506–507
- Empirical Rule Effect Size (ERES), 5:451
- Enamel, tooth, 1:324
- Encapsulated hemoglobins, 1:520
- Endoanal cushion, 1:68
- Endocardial catheter electrodes, 1:151–152
- Endocrine function, hyperthermia and, 4:48
- Endocrine systems modeling, 5:319–320
- Endogenic ionic current, 1:199
- Endometrial ablation, 6:375
- Endoradiosonde, 1:417
- End-organ damage, hypertension with, 1:15
- Endoscopes, 3:177–189. *See also* “Optical fiber” devices
future directions for, 3:186–187
history of, 3:178–179
light delivery with optical fibers, 3:179
medical applications using, 3:180–186
types of, 4:538t
- Endoscopic procedures, 4:535–537
- Endoscopic wireless pill, 1:423–424
- Endoscopy
fiber optics in, 3:307
gastrointestinal, 3:183–184
- Endothelial incorporation, in vascular graft prostheses, 6:498–500
- Endothelium derived relaxant factor (EDRF), 1:517
- Endotracheal intubation, 1:29
- Endovascular ablation, 6:376–378
- Endovascular stent-grafts, 6:495–496

- End-tidal CO₂, 2:(ETCO₂), 20, 21
- Energetic electron beam, 6:3
- Energy, Anger camera, 1:58. *See also*
Surface energy
- Energy absorption coefficients, 6:598
- Energy-dispersive spectrometer (EDS),
1:356
- Energy-dispersive X-ray analysis (EDX),
4:435
- Energy fluence, photon, 6:591
- Energy parameters, X-ray beam, 6:599
- Energy resolution, Anger camera, 1:60
- Energy storage, in human body, 1:191
- Energy Storage and Return (ESAR)
prosthetic feet, 4:553
- Energy transfer coefficients, 6:598
- ENG recording configurations, 3:112–117.
See also Electroneurography (ENG)
- Engine, microheat, 4:433
- Engineered tissue, 3:189–210
biocompatibility of, 3:201
bioreactor technology and, 3:200
cryopreservation of, 3:201–202
examples of, 3:202–205
future prospects for, 3:205–206
growth of, 3:191
history of, 3:189–190
immune concerns related to, 3:192
properties of, 3:200–201
scaffolds in, 3:194–200
signals in, 3:198–200
spatial organization in, 3:199
theory behind, 3:190–202
- Engineering, biosurface, 1:409–417
- Engineering Accreditation Commission,
1:404
- Engineering evaluation form, 3:219t
- Engineering in Medicine and Biology
Society (EMBS), 4:316
- Enhanced resolution techniques, in
electrophoresis, 3:134–138
- Enterococci, 1:114
- Enterprise level PACS, 6:309
- Entropy-based neurological monitor, 5:39
- Environmental control, 3:210–215
feature control in EADL systems,
3:212–213
power switching and, 3:210–212
subsumed devices in EADL systems,
3:213–214
- Environmental noise, reducing, 1:234–235,
236f
- Environmental pathogens, classification
of, 1:113–114
- Environmental probes, fluorescent,
4:468–469
- Environmental Protection Agency (EPA),
2:155
- Environmental rounds, conducting, 3:228
- Environmental temperature, for infant
incubators, 4:148
- Environment-induced cracking (EIC),
1:311
- Enzymatic conversion, 2:193–194
- Enzyme activity, in white blood cells, 2:411
- Enzyme electrode glucose sensors
based on conductive polymers, 3:399–400
based on oxygen detection, 3:400–402
based on peroxide detection, 3:398–399
implantable, 3:398–402
- Enzyme immunoassays (EIAs), 1:21
- Enzyme-linked immunosorbent assays
(ELISA), 4:384
- Eosinophils, 2:81, 411
- Epicel, as a skin substitute, 6:177
- Epidemiology, of hydrocephalus, 4:1
- Epidermal growth factor (EGF), 1:376–377
- Epidermis, 1:131
electrical properties of, 1:132
morphology and function of, 6:187–188
properties of, 6:169
- EPID images, 4:93
- Epilepsy, magnetoencephalography and,
1:245
- Epimysial electrodes, 3:353–355
- Epinephrine, 1:150
- Epiretinal implants, 6:534–535
- Episodic hypertension, 1:15
- Epoxy resin-based tissue substitute
manufacture, 5:256–257
- Equal pressure point concept, 6:102–103
- Equations, shock assessment, 6:167
- Equilibration method, 2:110
- Equilibrium potential, 1:122–123
- Equipment. *See also* Apparatus;
Biomedical equipment maintenance;
Instrumentation
anorectal manometry, 1:62–63
atomic absorption spectrometry,
3:319–321
blood collection, 1:457–458
blood pressure monitoring, 4:569–570
DNA sequencing, 2:431t
EGG recording, 3:86–87
electrophoresis, 3:138–139
esophageal manometry, 3:230
exercise stress testing, 3:247–256
flame atomic emission spectrometry,
3:319
force spectroscopy, 4:510
gas system, 3:379–380
hyperbaric medicine, 4:27
intraaortic balloon pump, 4:166–167
microarray, 4:367–370
MRI, 5:78
near-field imaging, 4:436–439
neutron activation analysis, 5:43–49
piezoelectric sensor, 5:362–363
prosthetic heart valve testing, 3:430–431
spinal cord stimulation, 6:225–226
thermodilution cardiac output
measurement, 2:29–32
ultraviolet therapy, 6:484–485
X-ray, 6:550–560
- Equipment acquisition, 3:216–222
implementation process in, 3:222
justification process in, 3:216–217
process outline for, 3:216t
selection process in, 3:217–222
- Equipment evaluation scoring form,
3:222t
- Equipment user evaluation form, 3:220t
- Equivalent circuit analysis, 5:361–362
- Equivalent circuit model, of skin, 1:132
- Equivalent current dipole (ECD),
1:239–240
- Equivalent dose, 5:505
- Erectile dysfunction (ED)
instruments and measurement of,
6:155
treatment of, 6:158–159
- Erosion corrosion, 1:311
- Error analysis, of effective thermal
property measurements, 1:195–196
- Error related negativity (ERN), 3:237
- Erythema, 6:475
- Erythrocyte elongation, 1:504. *See also* Red
blood cells (RBCs)
- Erythrocyte filtration, 1:504
- Erythrocyte orientation, blood flow
conductivity based on, 1:208
- Escherichia coli*, 1:114
- Esophageal manometry, 3:229–233
anatomy and physiology related to,
3:229–230
conducting, 3:230–232
equipment for, 3:230
indications for, 3:230
interpreting results of, 3:232–233
- Esophageal pressure, 6:516
- Esophageal pressure measurements, 6:520
in sleep laboratory, 6:212
- Esophageal speech, 4:230
- Esophagitis, 3:390
- Esophagus temperature monitoring,
6:318
- Esterase enzymes, 2:411
- Ethosomal drug carriers, 2:477
- Ethosomes
formulative aspects of, 2:477–478
therapeutic potentialities of, 2:478–479
- Ethylene oxide sterilization, 6:276–277
- E-Trans technology, 2:503
- Euler angles, 4:210
- European Society of Hypertension, 1:489
- EVADIA Group artificial pancreas,
5:229–230
- Evaluation
of CAD schemes, 2:298–302
of electrophoresis, 3:139–141
of gastrointestinal hemorrhage, 3:385
of neutron activation analysis, 5:47–48
of visual prostheses, 6:545
- Evanescent-wave spectroscopy, 5:161–162
- Event-related oscillations, 3:239–241
- Event-related potentials (ERPs), 3:236
analysis of, 3:237–238
- Everything-On-Line (EOL), 5:350
- Evoked potentials, 3:233–246
auditory, 1:97–98, 98–99
cognition and, 3:236–237
ERP analysis and, 3:237–238
event-related oscillations and,
3:239–241
generation of, 3:235
omitted, 3:237
recording, 3:234–235
sensory, 3:235–236
single-trial analysis of, 3:241–242
source localization and, 3:238–239
wavelet transform and, 3:239–241

- Excitation
 in functional electrical stimulation, 3:348–350
 in linear variable differential transformers, 4:253
- Excitation frequency, in linear variable differential transformers, 4:255
- Excitation microscopy, multiphoton, 2:93–94
- Excitation wavelengths, 3:346
- Executive Information System (EIS), 5:131
- Exercise(s)
 feedback control of, 1:398–400
 principles for, 1:391–393
 for rehabilitation, 1:398
 for scoliosis, 6:130
 thermoregulation and, 1:191
- Exercise (stress) electrocardiography, 3:47–48
- Exercise equipment
 feedback control of, 1:398–400
 hydraulic, 1:397–398, 398–399
 weight-based, 1:397
- Exercise fitness biomechanics, 1:384–403
 exercise and training principles, 1:391–393
 feedback control of exercise, 1:398–400
 future developments in, 1:400–401
 high technology tools and, 1:393–398
 quantifying motion, 1:384–391
- Exercise physiology, basic principles of, 3:251t
- Exercise protocols, in exercise stress testing, 3:251
- Exercise stress testing, 3:246–263
 angiographic disease prediction and, 3:256–257
 biomarkers in, 3:255–256
 blood pressure measurement in, 3:247–248
 computerization in, 3:249
 ECG recording in, 3:248, 249, 250
 echocardiography in, 3:255
 equipment for, 3:247–256
 evaluation in, 3:256–262
 lead systems in, 3:248–249
 modalities in, 3:250–251
 noise in, 3:249–250
 nuclear techniques in, 3:254–255
 postexercise period in, 3:252–253
 prediction of restenosis with, 3:258–259
 protocols in, 3:251–252
- Exercise test, ACC/AHA guidelines for prognostic use of, 3:259–262
- Exercise test modalities, in exercise stress testing, 3:250–251
- Exopolymer production, in biofilms, 1:115
- Expected oxygen tension, 6:105
- Experimental dosimetry, 1:614–615
- Experimental neural prosthetic systems, 3:128–129
- Expert systems, machine-based, 2:317
- Expiration, forced, 5:432
- Expiratory reserve volume (ERV), 6:100
- Exposure limits, electrical and magnetic field, 2:183t
- External anal sphincter (EAS), 1:62, 65
- External beam radiation therapy (EBRT), 5:418
- External beam treatment delivery, 5:481–482
- External biosignal monitoring electrodes, 1:130, 137–139
- External defibrillators, 2:406–407
- External electrostimulation electrodes, 1:142–146
 history of, 1:142–144
- Externally applied hyperthermia, 4:68–73
- External measurement devices, for female sexual behavior assessment, 6:153
- External pacemakers, temporary, 5:218
- External pacing electrodes, 1:131
- External processor, in visual prostheses, 6:537
- External stimulation electrodes, 1:131
- Extracellular currents, 3:110–111
- Extracellular matrix (ECM), 1:366;
 3:198–199; 6:180, 383–384
 implant-induced alterations of, 1:110
- Extracellular matrix analogs
 characteristics of, 6:184–187
 chemical composition of, 6:186
- Extracellular matrix-based bioscaffolds,
 disinfection of, 6:279
- Extracellular matrix proteins, 5:402t
- Extracellular water (ECW) volumes, in dialysis patients, 1:212
- Extracoronary tooth restorations, 6:425
- Extra-corporal photochemotherapy (ECP), 6:486
- Extracorporeal hyperthermia devices, 4:73–76
- Extracorporeal measurement systems,
 optical sensors in, 5:170
- Extracranial stereotactic targeting, 5:577
- Extra low interstitial (ELI) grade, 1:105
- Extraneural electrodes, 3:123–125
- Extravascular transducers, 1:485
- Extremely low frequency (ELF) radiation,
 5:68–69
 use in medical diagnosis and therapy, 5:71
- Extrinsic probes, 4:497–498
- Extrinsic spatial resolution, Anger camera, 1:60
- Extrusion technique, vesicles by, 2:469
- Eye(s). *See also* Visual entries
 effect of UV exposure on, 6:475
 features of, 3:268–269
- Eye movement(s)
 calibrating and analyzing, 5:146
 recording, 5:137–149
 saccades, 5:137–139
 torsional, 3:275
 version and vergence in, 5:137
- Eye movement measurement techniques,
 3:263–286. *See also* Eye orientation
 comparison of, 3:283–284
 electro oculography, 3:266–267
 feature recognition, 3:269–270
 optical sensors, 3:267–268
 optical techniques, 3:267
 scleral search coil technique,
 3:265–266
- Eye-movement recording technologies,
 5:140–146
- Eye orientation
 as a function of feature shape, 3:273
 as a function of pupil or corneal reflection position, 3:271
 as a function of reflectivity pattern movement, 3:273
 as a function of relative first and fourth Purkinje image positions, 3:272–273
 as a function of relative pupil and corneal reflection positions, 3:271–272
 as a function of relative pupil or iris and facial landmark positions, 3:273
- Eye shields, 5:597–599
- Eye surgery, 4:530–532
- Eye tracking systems
 calibration of, 3:275–276
 compatibility with eyeglasses and contact lenses, 3:276
 dual Purkinje image measurement, 3:277–278
 head mounted, video based, 3:279–280
 illumination safety of, 3:276
 implementations of, 3:276–283
 photo electric, reflectivity pattern (limbus), 3:276–277
 remote, video-based, 3:280–283
 using two-dimensional video sensor arrays, 3:278–279
 video-based with fMRI, 3:283
- Fabrication methods/techniques
 polymeric-scaffold, 5:390
 for porous biomaterials, 5:400–401
- Fabrication technologies, for capacitive sensors, 2:2–5
- Fabric electrodes, 1:142
- Fabry–Perot interferometry, 6:359
- Face recognition, 3:270
- Factor analysis, in tracer kinetics, 6:435
- Fahrenheit temperature scale, 6:313
- Failsafe mechanism, anesthesia machine, 1:36–37
- Failure hazards, of medical devices, 3:538
- Failure
 of biomaterials, 1:278–281
 of hip prostheses, 1:314–315
- Failure mechanisms, of ligaments and tendons, 4:247–248
- Fall, Magnus, 1:430
- Faraday, Michael, 1:429
- Faraday's law, 1:309
- Faradic stimulation, 1:429
- Farado-puncture, 1:150
- Fast neutron therapy, 6:6–7
 facilities for, 5:61–62
 origins of, 5:50–51
 radiobiological rationale for, 5:51–53
- Fast neutron therapy beams, production of, 5:57–58
- Fast separations, in microdialysis sampling, 4:410
- Fast X-ray computed tomography (CT), in peripheral vascular noninvasive measurements, 5:246
- Fat pads, 4:199

- Fatigue, due to corrosion, 1:311
- Fatigue issues, public awareness of, 3:543
- Fatigue strength, of bone, 1:531
- Fatigue wear, 1:315
- Fauchard, Pierre, 1:326
- FDA Modernization Act (FDAMA), 4:315.
See also Food and Drug Administration (FDA)
- FDA quality system regulations, 2:150–152
- FDA-regulated entities, 2:142
- F (Fisher) distribution, 6:255t
- Feature-based algorithms, in arrhythmia analysis, 1:76–77
- Feature extraction
 alternatives to, 2:296
 in automated cytology, 2:403
 in computer-assisted detection/diagnosis, 2:295–296
 EEG, 3:74–75
- Feature extraction/classification, using IR imaging, 6:352
- Feature recognition systems, 3:269–270
- Fecal incontinence
 biofeedback clinical outcome literature related to, 1:180, 182
 manometric features of, 1:65
- Feedback control
 of exercise, 1:398–400
 in incubators, 4:153
 of movement, 1:385–386
 in thermoregulation, 1:191
- Feedback controlled drug delivery, 2:439
- Feedback operation, in the atomic force microscope, 4:506–508
- Feet, prosthetic, 4:552–553
- Fegler, G., 2:25
- Feldspar, in porcelain, 1:326
- Feldspathic porcelain, 1:326
- Female condoms, 2:339–340
- Female human sexual behavior
 instruments and measurement of, 6:149–150
 treatment of, 6:153–155
- Female sexual behavior assessment, 6:150–153
 external measurement devices for, 6:153
 internal devices for, 6:150–152
- Female sexual dysfunction, treatment articles on, 6:154–155
- Fenn, Wallace O., 5:430
- FES devices/systems, 3:350–358. *See also* Functional electrical stimulation (FES)
- Fetal activity monitoring, 3:299
- Fetal blood gas monitoring, 3:298–299
- Fetal cardiocography, 3:296–298
 clinical applications of, 3:297–298
- Fetal electrocardiogram (FECG), 3:288
- Fetal electroencephalography, 3:299–300
- Fetal heart, acoustic pickup of, 3:291
- Fetal heart monitoring, signal processing in, 3:288–290
- Fetal heart rate (FHR), 1:475
 direct determination of, 3:288
 indirect sensors of, 3:290–293
 monitoring of, 3:287–293
- Fetal hemoglobin, 2:40
- Fetal magnetocardiography (fMCG), 1:246
- Fetal magnetoencephalography (fMEG), 1:246–247
- Fetal microblood analysis, 3:298
- Fetal monitoring, 3:287–301
 alternative methods of, 3:298–300
 by a human observer, 3:287
 physiologic variables in, 3:287
 uterine contractions in, 3:293–296
- Fetal pulse oximetry, intrapartum, 1:475–476
- Fetal studies, 1:246–247
- Fetal surgery, 4:533
- Fever-range whole-body hyperthermia, 4:58
- FFR_{myo}-FFR_{myo-g} correlations, 2:357–359
- Fiber-based confocal microscopy, 3:309–310
- Fiber-matrix interactions, in ligaments and tendons, 4:243
- Fiber-optic catheters, mixed-venous, 5:165–166
- Fiber-optic fluoroimmunoassay systems, 4:378
- Fiber-optic probes, temperature measurement with, 6:358–359
- Fiber optics, 3:301–315. *See also* Optical fiber entries
 diagnostic applications of, 3:306–313
 general principles of, 3:302–304
 illumination applications for, 3:304–306
 in spectroscopy, 3:311
 physics of, 3:302
 surgical applications of, 3:313–314
 therapeutic applications of, 3:313–314
- Fiber-optic sensors, 5:206
- Fibers, types of, 3:303–304
- Fibrillation, ventricular, 1:79
- Fibrillation detection interval (FDI), 1:72
- Fibrin, in tissue engineering, 1:370; 6:384
- Fibrinous inflammation, implant-related, 1:116
- Fibroblasts, tissue regeneration and, 1:109
- Fibrocartilage, 2:63
 composition and structure of, 2:65–66
- Fibrochondrocytes, 2:63
- Fibromyalgia, biofeedback clinical outcome literature related to, 1:179–180
- “Fibrous encapsulation,” 1:109
- Fibrous joints, 4:199, 203f
- Fick, Adolf, 2:12–13
- Fick cardiac output technique, 2:12–21
 assumptions when using, 2:14–16
 flow-dependent parameters for, 2:18–19
 history of, 2:12–13
 intracardiac shunts and, 2:16–18
 physiology of, 2:13
 practical considerations for using, 2:14
 variations of, 2:19–21
- Fick principle, 2:21, 24
- Fidia Advanced Biopolymer, 1:340
- FID signal, 5:75, 77
- Field effect transistors (FETs), reference, 4:194–195. *See also* Immunologically sensitive field-effect transistors; Ion-sensitive field-effect transistors (ISFETs)
- Field-flow fractionation (FFF), 2:107–108
- Field geometries, electrosurgical, 3:173–174
- Field-shaping devices, 5:594–599
- Filament, in X-ray tubes, 6:600–601
- Filament evaporation, in X-ray tubes, 6:608
- File Transfer Protocol (FTP), in office automation systems, 5:156
- Filler-matrix interface, in resin-based composites, 6:95
- Film(s)
 CT, 2:239
 physical characteristics of, 6:140–141
 radiographic, 5:483
 in screen-film systems, 6:140–145
- Film contrast/latitude, in screen-film systems, 6:143–144
- Film digitizers, 5:336–337
- Film dosimetry, 5:477
- Film processing, for screen-film systems, 6:145–147
- Film processors, quality control of, 6:578
- Filtered backprojection (FBP)
 reconstruction technique, 2:249–250; 5:115
- Filtering, in neonatal respiration monitoring, 5:21
- Filtering techniques, for signal noise, 1:202
- Filters
 acoustooptic tunable, 4:457–459
 in CT reconstruction methods, 2:247–248
- Filtration, in microbio reactors, 4:391
- Finetech-Brindley (VOCARE) bladder system, 1:432, 438–441
- Finger blood pressure monitoring, 1:489
- Fingertip electro tactile displays, 6:296, 297
- Finite element analysis (FEA), of human joints, 4:218–219
- Finite element models
 patient-specific and task-dependent morphological, 4:219–220
 of spine stabilization procedures, 3:585
- First aid procedures, in blood collection/processing, 1:456
- First-order gradiometers, 1:233, 236
- Fistulas, upper airway, 3:507
- Fitness, need for, 1:389. *See also* Exercise entries
- Fixed-fixed computed tomography, 2:232–233
- Fixed-rotate computed tomography, 2:232
- Fixed threshold detection, in neonatal respiration monitoring, 5:20–21
- Flail chest, 2:38
- Flame atomic emission spectrometry, 3:315–317
 equipment for, 3:319
 medical applications of, 3:321
 theoretical basis for, 3:317–318
- Flame ionization detector (FID), 4:325
- Flame photometric detector (FPD), 4:325
- Flaps, hyperbaric medicine and, 4:24–25
- Flat interface nerve electrode (FINE), 3:124
- Flat panel technology, 4:91–92
- Fleish pneumotachograph, 6:521

- Fleish pneumotachometers, 5:369
- Flexible electrodes, 2:1–2
- Flexible endoscopes, 4:536
- Flexible imaging bundles, fabrication of, 3:307–308
- Flexion–extension
of the middle and lower cervical spine, 3:555–556
of the occipital–atlantoaxial complex, 3:554
- Floating electrodes, 1:139
- Flood field image, Anger camera, 1:58
- Flow controllers, anesthesia machine, 1:35
- Flow cytometry, 2:397
- Flow-dependent parameters, for Fick cardiac output technique, 2:18–19
- Flow dynamics
of heart valve prostheses, 3:426–437
past mechanical heart valves, 3:417–421
- Flow mapping, magnetic resonance, 3:330–332
- Flow measurement
indirect techniques for, 5:378
invasive (inline volume), 3:327–329
velocity, 3:329–340
- Flowmeters
anesthesia machine, 1:35
Doppler volume, 3:327
electromagnetic, 3:322–342
flutter, 5:371
transit time volume, 3:325–327
vortex shedding, 5:370–371
- Flowmetry, laser Doppler, 2:378
- Flow monitoring device, 2:504
- Flow probes, electromagnetic, 3:324–325
- Flow rate, constant, 2:29
- Flow ratio, pulmonary–systemic, 2:17–18
- Flow sensors, lift force gas, 5:371
- Flow transducers, 5:435–436
- Flow visualization, in heart valve prostheses, 3:434
- Flow–volume curves, 5:437
- Fluence, particle, 5:465
- Fluence rate, photon, 6:591
- Fluid challenge, in shock assessment, 6:166–167
- Fluid control components, in microbioreactors, 4:389–390
- Fluid-filled catheter transducer systems, 4:578–579
- Fluidic oscillator, 5:371
- Fluidics stations, in array platforms, 4:369
- Fluid mechanics, 4:420–422
- Fluid-resistance pneumotachometers, 5:369–370
- Fluid restriction, total parenteral nutrition regimen for, 5:132
- Fluids
measuring rheological properties of, 1:504–506
Newtonian, 1:500–501, 504–505
non-Newtonian, 1:500–501
pumping, 4:423–425
- Fluid shifts, impedance plethysmography and, 4:129
- Fluid status, intrathoracic, 1:205–206
- Fluorescence, 2:91; 3:346
for dialysates, 4:411
environmental effect on, 4:490
imaging via, 3:346
molecular biology applications of, 3:346
signal detection in, 2:414
- Fluorescence correlation spectroscopy (FCS), 2:97–98
- Fluorescence detection, for blood cell counting, 2:85–86
- Fluorescence detectors, 4:492–493
- Fluorescence excitation light sources, 4:491
- Fluorescence *in situ* hybridization (FISH), 4:445, 446
- Fluorescence lifetime imaging microscopy, 2:95–97
- Fluorescence measurements, 3:342–347
practical applications of, 3:345–346
theory and instrumentation, 3:343–345
- Fluorescence microscopes
designs of, 4:490–494
optical components of, 4:491–492
- Fluorescence microscopy, 2:91–98; 3:344–345; 4:488–503. *See also* Fluorescence microscopes; Fluorescent probes
advanced functional imaging modalities based on, 4:498–502
confocal, 4:493–494
DNA sequencing by, 2:435
spectroscopic principles of, 4:489–490
two-photon, 4:494
wide-field, 2:92
- Fluorescence microscopy configurations, advanced, 4:493–494
- Fluorescence polarization, 3:346
- Fluorescence quenching oxygen analyzers, 5:206–207
- Fluorescence resonance energy transfer (FRET), 4:502
- Fluorescence spectroscopy, 4:489–490
fiber optics in, 3:311
- Fluorescence techniques, in diagnosis, 6:480–481
- Fluorescence WBC counting technique, 2:412
- Fluorescent cell probes, 2:392t, 393t
- Fluorescent compounds, detection of, 3:345
- Fluorescent dyes, for confocal microscopy, 4:466–467
- Fluorescent emission, 3:346
- Fluorescent environmental probes, in confocal microscopy, 4:468–469
- Fluorescent excitation, 3:346
- Fluorescent lamps, ultraviolet radiation from, 6:477
- Fluorescent lifetime, 3:346
- Fluorescent probes, 4:494–498
classification of, 4:497–498
optical factors in selecting, 4:495–497
- Fluorescent proteins, confocal microscopy and, 4:471–472
- Fluoro CT, 2:238
- Fluoroimmunoassay systems, fiber-optic, 4:378
- Fluorophore intensity measurements, 4:498
- Fluorophores, 2:91–98
basic characteristics of, 4:465–466
for confocal microscopy, 4:464–465
- Fluoroscopic units, requirements for, 2:177t
- Fluoroscopic X-ray equipment, quality control of, 6:570
- Fluoroscopic X-ray output, 6:571
- Fluoroscopy, 4:87–89; 6:559
phantom materials in, 5:266
- Fluosol DA, 1:514, 515
- Flutter flowmeter, 5:371
- Flux analysis, 1:225–226
- Flux transformers, 1:232–234
- Foams, sol–cel-derived bioactive glass, 1:292–293
- Focal spot loading, in X-ray tubes, 6:606–607
- Focal spot sizes, in X-ray equipment, 6:564
- FOcal Undetermined System Solution (FOCUSS) algorithm, 1:240
- Focus, in X-ray tubes, 6:603–605
- Focused radiation fields, 4:65
- Focusing cup, in X-ray tubes, 6:601
- Foil electrodes, 1:144
- Fold recognition, 1:221
- Follicles, 1:133
- Food and Drug Administration (FDA), 2:155–156. *See also* FDA entries
on biomaterials, 1:268–270
blood glucose monitors approved by, 1:16
human factors engineering perspective of, 3:537–538
- Food industry, monoclonal antibodies in, 4:607
- Foot structure, stability of, 4:221–222
- Force, in exercise, 1:392
- Forced expiration, 5:432
- Force spectroscopy, 4:505–506, 510–513
applications of, 4:510–513
evaluation of, 4:514–515
- Foreign body response, biomaterial failure related to, 1:280
- Forster resonance energy transfer (FRET), 3:346
- Forward simulation, 4:217–218
- Forward solution, 1:239
- Fourier transform infrared spectroscopy (FTIR), 1:356, 357, 358, 359f, 362, 363f
- Fourth Purkinje image recognition, 3:270
- Fractionation, field-flow, 2:107–108
- Fracture fixation, 1:257
- Fracture orthoses, 6:90f
- Frameless stereotactic technique, 6:269–270
- Free hemoglobin (FHb), 1:516, 517
- Free induction decay (FID), 5:74. *See also* FID signal
- Free-radical polymerization, 1:331
- Freeze and thawed multilamellar vesicles (FAT-MLVs), 2:469
- Freeze–thaw cycles, in cryosurgery, 2:367–368
- Freezing. *See also* Cryosurgery
body/tissue damage from, 1:192
effect on tissue, 2:362–363
tissue injury from, 6:369

- Frequency dependence, of impedance, 3:117–119
- Frequency domain analysis, 3:108
- Fresh gas decoupling, in anesthesia machines, 1:40
- Fretting-corrosion, 1:311
- Freund phalloplethysmograph, 6:154f
- Fricke dosimeter, 5:473
- Fricke gels, 5:484–485
- Friction, 1:314
- Friction coefficients, of diamond-like carbon coatings, 1:318
- Friedman test, for matched samples, 6:259
- Fringe effects, with metal electrodes, 1:146–148
- Frontal electrode placement, for biofeedback, 1:177
- Frontalis electromyogram (FEMG), in anesthesia monitoring, 1:44
- Frontal plane, scoliosis and, 6:122
- Fuel cell, in solid electrolyte cell oxygen analyzers, 5:204–205
- Full field digital mammography, 4:303–304
- Full field digital mammography (FFDM) systems, 4:302, 304–305
imaging characteristics of, 4:305t
- Full-body CT screening, 2:264
- Fuller, Buckminster, 1:298
- Fullerenes, 1:297–298, 305–306
- Fullerite, 1:298
- Full-field optokinetic response, 5:140
- Full film lubrication, 1:315
- Full-thickness injuries, 2:72
- Functional angiological thermatomes, 6:349
- Functional ankylosis, 1:328
- Functional deficit scale, injury-related, 6:179–180
- Functional electrical stimulation (FES), 3:347–366. *See also* FES devices/systems
controllers and control strategies in, 3:357–358
electrode designs for, 3:351–357
theory and application of, 3:348–350
therapeutic effects of, 3:358–359
- Functional magnetic resonance imaging (fMRI), 4:289–290
- Functional mapping, presurgical, 1:244–245
- Functional residual capacity (FRC), 5:437
by plethysmography, 5:438
- Functional stereotaxis, 6:268
- Fundus camera, 5:291
- Fundus photos
clinical evaluation of age-related macular degeneration in, 5:293–294
stereo, 5:293
- Fundus reflectometry, 5:135–136
clinical applications of, 5:136
- Fungi, electron microscopic diagnostic criteria for, 4:485
- Fusion bonding, 2:3, 4f
- Fusion cage stabilization, interbody, 3:578–579
- Fuzzy logic systems, 2:317–319
- Gadolinium, tracer kinetics of, 6:432–433
- Galen, 2:35
- Gallium arsenide sensors, 6:359
- Galvani, Luigi, 1:197, 429
- Galvanic corrosion, biomaterial failure from, 1:278
- Galvanic fuel cell, 6:524
- Galvanic oxygen analyzers, 5:202–203
- Galvanic skin response, sexual arousal and, 6:161
- Galvanism, 1:143
- Galvanometer, 1:137
- Galvanometric recorders, 6:55–58
- Galvano-puncture, 1:150
- Gamma cameras, 1:51; 5:96–100. *See also* Anger camera
acquisitions possible with, 5:113
animal, 5:104–105
types of, 5:98–100
types of acquisition from, 5:98
- Gamma emission, radioactive decay and, 5:560
- Gamma frequency band, 1:244
- Gamma Knife, 3:367–377
clinical use of, 3:373–375
early history of, 3:368–372
evaluation of, 3:376
quality control/quality assurance for, 3:375
risk analysis for, 3:376
theory behind, 3:372–373
- GammaKnife unit, 5:578
- Gamma-ray emitting radionuclides, Anger camera and, 1:51
- Gamma rays, 5:303
Anger camera and, 1:52–53, 55–56
- Gamma spectrum, neutron activation analysis of, 5:46–47
- Gamma sterilization, 6:277–278
- Gamma Unit Model B, 3:370–371
- Ganglion cell dysfunction, 3:155
- Gantry, in CT scanners, 2:234
- Garment electrodes, 1:149
- Gas. *See also* Gases
measurement of carbon dioxide in, 6:526
measurement of oxygen in, 6:524–525
- Gas adsorption chromatography, 2:104
- Gas analysis, 5:436
- Gas analyzers, nondispersible infrared, 5:436. *See also* Medical gas analyzers
- Gas bubbles, mathematical and physical modeling of, 6:467–468
- Gas chromatography (GC), 2:103–104; 4:325–327; 6:525
- Gas chromatography–mass spectrometer (GCMS), 5:207–208
- Gas compression, in anesthesia machines, 1:40
- Gaseous oxygen sensors, 5:207–208
- Gases, 1:465. *See also* Gas
ionization by radioisotope, 6:521–522
medical, 3:379
partial pressures of, 6:104
- Gas exchange assessment, parameters used in, 6:518–520
- Gas-filled ionization detectors, 2:236
- Gas-foaming method, scaffold fabrication via, 1:378
- Gas laws, 5:433
- Gas-liquid chromatography, 2:104
- Gas monitor methods, 4:323
- Gas partial pressure, 1:465
- Gas proportional counters, 5:510–511
- Gas proportioning system, anesthesia machine, 1:37
- Gas sensors, electrochemical, 4:324–325
- Gas systems, 3:377–381
components of, 3:378–379
installation of, 3:380
maintenance of, 3:383–384
performance criteria and standards for, 3:380–381
pressurized, 3:377–381
source equipment for, 3:379–380
- Gas transport, effects of disease on, 6:106–107
- Gas transport/exchange, in the respiratory system, 6:103–106
- Gastric dysrhythmias, 3:85
percentage of, 3:90
- Gastric myoelectrical activity, normal and abnormal, 3:84–85
- Gastric slow waves, percentage of, 3:90
- Gastritis, 3:390
- Gastroesophageal varices, 3:385–388
- Gastrointestinal disorders, biofeedback
clinical outcome literature related to, 1:180
- Gastrointestinal endoscopy, 3:183–184
- Gastrointestinal hemorrhage, 3:384–393
from aortoenteric fistula, 3:390
evaluation and resuscitation for, 3:385
from hemobilia, 3:391
localization of, 3:386t
upper, 3:385–388
- Gastrointestinal system, biomagnetic measurements and, 1:248
- Gas volume measuring devices, 5:434–435
- Gate control theory of pain, 6:440
- Gauges, anesthesia machine, 1:35
- Geiger-Müller counters, 5:511
- Gel-casting, of hydroxyapatite, 1:291
- Gel dosimeters, characteristics of, 5:489
- Gel dosimetry, 5:478–479, 484–488
applications of, 5:489–494
complications associated with, 5:494–495
from imaging procedures, 5:489
- Gel electrophoresis, DNA sequencing and, 2:427–428
- Gel-less electrodes, 1:134
- Gel phantom, 5:264
- Gels. *See also* Electrode gels
equivalence and energy dependence of, 5:494
Fricke, 5:484–485
in stereotactic radiosurgery, 5:490–491
polymer, 5:485–486
- GEMISCH multi-user database
programming language, 1:44
- Gene clustering, 1:224
- Gene delivery, in tissue engineering, 6:389

- Gene expression patterns, *1:226*
- General anesthesia, typical process of delivering, *1:29*
- General electrosurgery, *3:159–160*
- Generalization
of biofeedback response, *1:178*
operant conditioning and, *1:167*
- Generators
electrohydraulic, *4:259–260*
electromagnetic, *4:260*
micropower, *4:430–433*
piezoelectric, *4:260*
thermoelectric, *4:430*
- Genes
classification of, *1:224*
expression of, *4:361*
genetic code and, *1:217*
structure of, *4:361–362*
- Genetic code, *1:217t*
- Genetic control, by bioactive materials, *1:290*
- Genetic engineering, recombinant, *4:600–601*
- Genetic expressible probes, *4:498*
- Genetic network, *1:225*
- Genital engorgement, *6:153*
- Genome analysis, *1:221–223*
- Genome annotation, *1:222*
- Genome assembly, *1:221–222*
- Genome sequencing, *2:433–434*
- Genomic instability, ionizing radiation and, *4:184*
- Genomics, comparative, *1:222–223*
- Genosensors, piezoelectric, *5:365–366*
- Geometric blurring, in computed tomography, *2:251*
- Geometric penumbra, *2:131*
- Geometric phantom, *5:264–265*
- Geometry specification, in Monte Carlo simulation, *5:534–535*
- Geriatric setting, use of thermistors in, *6:339*
- Geriatric systems modeling, *5:323*
- Germanium (Ge) detectors, *5:517*
- Germany, infrared imaging in, *6:353*
- Gill–Thomas–Cosman (GTC) relocatable head ring, *5:594*
- Glass ceramics, *1:287*
as biomaterials, *1:273*
in tissue engineering, *1:374*
- Glasses. *See also* Bioglass
bioactive, *1:284, 285*
as biomaterials, *1:273*
dissolution and bioactivity of, *1:286–287*
- Glass frits, *1:357, 358*
- Glass ionomer cements (GICs), *1:324*
- Glass ionomers, as dental fillings, *1:324*
- Glass-transition temperature, *1:331*
- Glaucomatous field defects, *6:530*
- Global Positioning System (GPS), *1:451–453*
- Glow discharge plasma modification, *1:345*
- Glucose intolerance, total parenteral nutrition regimen for, *5:132*
- Glycolic–lactic acid copolymers, as biomaterials, *1:107*
- Glucose monitoring
ambulatory, *1:16–17*
home, *3:395–396*
- Glucose-sensitive microtransponder, *1:424, 425f*
- Glucose sensors, *1:16–17; 3:393–406*
blood glucose prediction and, *3:402*
challenges for development of, *3:402*
ideal, *3:395*
new, *3:393–395*
noninvasive, *1:17*
optical, *5:164–165*
sensing methodologies and, *3:395–402*
- Glucose system, *5:126–127*
- Glycosaminoglycan (GAG) side chains, *2:64*
- Gold, in dentistry, *1:322*
- Goldmann perimetry, *6:529*
- Gold thin-film electrodes, *1:156–157*
- Goodness of fit test, *6:248–249*
- “Good Samaritan” legislation, *2:49*
- Gorlin formula, *2:19*
- Gott, Vincent, *1:302*
- Government regulation. *See* Regulation
- GPS navigational aids, for the visually impaired, *1:451–453*
- Gradient recalled echo (GRE) imaging, *4:288–289*
- Gradient separation, of peripheral blood mononuclear cells, *1:461–464*
- Gradiometers, *1:233–234, 235–236*
- Graft dysfunction, after liver transplantation, *4:271*
- Grafted polymer layers, *1:347*
- Grafts
hyperbaric medicine and, *4:25*
vein, *6:494*
- Gram-negative bacteria, *1:319–320, 371*
- Gram-positive bacteria, *1:320*
- “Granuloma pouch,” *1:111*
- Graphene structure, *1:297*
- Graphic audiograms, *1:95–96*
- Graphic recorders, *6:48–62*
analogue, *6:55–59*
analogue and digital, *6:49–51*
chart abscissa generation in, *6:51*
digital, *6:59–60*
evaluation of, *6:60–62*
galvanometric, *6:55–58*
graphic quality of, *6:53–55*
manufacturers of, *6:61t*
recording accuracy in, *6:51–53*
translational servorecorders, *6:58–59*
- Graphic recording, fundamental aspects of, *6:49–55*
- Graphics display, in neurological monitors, *5:34*
- Graphite(s), *1:296, 297*
synthetic, *1:299*
- Graphite-loaded polyesters, in electrodes, *1:131*
- Gravity drip drug infusion systems, *2:497–498*
- Green Cross Corporation, *1:514*
- Green strain tensor, *1:88*
- Grid computing, *6:309*
- Gross tumor volume (GTV), *5:545, 546; 6:32*
- Grounding, in electroencephalography, *3:116–117*
- Ground substance, in skin, *6:204*
- Group change, versus individual change, *5:451*
- Groupware systems, in office automation systems, *5:156–158*
- Growth factors
for modulating cell behavior, *2:74*
in tissue engineering, *1:366*
- Guardian real time system (Guardian® RT), *1:16*
- Guidances, Nuclear Regulatory Commission, *2:171*
- Guidant Galileo delivery device, *1:607–608*
- Guidant Galileo IVB system, *1:605–609*
- Guidant ICD algorithm, *1:75*
- Guidant Ventak AV III DR algorithm, *1:78*
- GuideCane, *1:450–451*
- Guidewire diagnostics
in coronary angioplasty, *2:349–360*
issues in, *2:353–354*
- Guide wire perturbation, *1:609–610*
- Guidewires
in coronary angioplasty, *2:349–350*
increased pressure drop and reduced hyperemic flow due to, *2:355–357*
in microsurgery, *4:532*
nickel–titanium shape memory alloy, *1:8*
- Guiding catheter, in coronary angioplasty, *2:349*
- Guluronate junction zone, *1:370f*
- Guyton model, *5:302–303*
- Gynecological diseases, cryosurgical treatment of, *2:374*
- Gynecologic electrosurgery, *3:160–161*
- Gynecology, high intensity focus ultrasound for, *4:78*
- Haar (square) wavelet, *1:75*
- Hagen, Gotthilf Heinrich Ludwig, *1:504*
- Hagen-Poiseuille law, *1:504*
- Hair follicles, *1:133*
- Haldane effect, *1:469*
- Half-cell potential, *1:122–123*
- Half-value layer (HVL), *6:592, 599*
determination in X-ray equipment, *6:566*
- Hamilton, W. F., *2:25*
- Hand-cycling, *4:549*
- Hand grasp, restoration of, *3:127*
- Handheld electrodes, *1:143*
- Hand surface temperature biofeedback, with tension and headache, *1:176*
- HAPEX composite, *1:289*
- Haptic feedback, in microsurgery, *4:528–529*
- Haptic icons, *6:299*
- Hard copy output device, in neurological monitors, *5:34*
- Hard wall contact, *1:351*
- Hardware
computer networking, *5:350–351*
medical image display, *5:353–355*
phonocardiography, *5:289*
visual prosthesis, *6:536–539*
- Harmonic excitation, coded, *6:469–471*
- Harmonic imaging, *3:4–5; 6:468–469*

- Haversian bone, 1:528
remodeling of, 1:534–535
- Haversian canals, 1:524
- Hazard assessment, ultraviolet radiation, 6:487–488
- HBOC-201 (Hemopure), clinical trial of, 1:518–519
- Headache. *See also* Migraine headache
biofeedback clinical outcome literature related to, 1:178–179
biofeedback training and, 1:167–168
- Head and neck cancer, treatment planning for, 5:538
- Headgear, in continuous positive airway pressure, 2:335
- Head holder, 5:588–589
- Head Injury Management Guidelines, 4:578
- Head motion, measuring point-of-gaze in the presence of, 3:273–275
- Head-mounted video-based eye tracking systems, 3:279–280
- Head stereotactic localizer, 5:594
- Healing
regeneration versus repair in, 6:180
response to biomaterials, 1:108–109
- Health, in older people, 1:390
- Healthcare, problems with, 6:109–114
- Healthcare information systems, 4:310
- Healthcare process, increased patient participation in, 3:542–543
- Health Insurance Portability and Accountability Act (HIPAA), 1:456
- Health issues, nanoparticle-related, 5:7–10
- Health On the Net Code of Conduct (HONcode), 4:309
- Health Physics Society, 4:320
- Health status measures, clinical significance of, 5:444–445
- Healthy organs
composed of separately critical voxels, 6:46
integral response for, 6:45–46
- Healthy subjects, EGG in, 3:91–92
- Healthy tissue integral dose, 6:43
- Hearing. *See also* Audiometry; Ear entries
assessment of, 2:215
assistive devices related to, 2:222
- Hearing aids, 2:224
- Hearing analysis software, 2:214
- Hearing disorders, new directions in, 2:224
- Hearing impairment, categories of, 2:211
- Hearing intervention, 2:218–219
- Heart. *See also* Artificial heart; Cardiac entries; Cardio- entries; Myocardial entries; Pacemaker entries
anatomy and function of, 3:450
conducting system of, 2:43
electropuncture of, 1:150
hemodynamic monitoring of, 4:566–567
pressures and flows in, 3:450t
as a pump, 3:477–483
sounds and murmurs from, 5:279–282
- Heart-arterial coupling, 3:492–494
- Heart beat detection, computer-based, 3:50–51
- Heart disease(s), 1:388. *See also* Electrocardiography
diagnosis of, 3:41
- Heart–lung machines, 3:459–462
CPB circuit and, 3:461–462
future directions for, 3:462
- HeartMate XVE, 3:453
- Heart performance, control of, 4:566
- Heart rate (HR), 2:12
control of, 4:566–567
- Heart rate home health care devices, 3:528–529
- Heart rate instrumentation, 1:174
- Heart rate monitoring, fetal, 3:287–293
- Heart rate sensors/amplifiers, 1:175
- Heart rate studies, 3:257t
- Heart rate variability (HRV) biofeedback, 1:182
- Heart rate variability instruments, 1:175–176
- Heart rate variability training, 1:174, 175
- Heart rhythm disorders, cryosurgical treatment of, January 18, 2006: 375–376
- Heart sounds, 5:279–282
processing of, 5:287–288
recording of, 5:283–287
- Heart sound transducers, 5:283–286
- Heart valve function, dynamics of, 3:417
- Heart valve prostheses, 3:407–426. *See also* Prosthetic heart valves
functional characteristics of, 3:415–423
future directions in, 3:435–436
ideal design of, 3:410–415
in vitro flow dynamics of, 3:426–437
in vitro testing of, 3:430–435
technology related to, 3:428–430
- Heart valves. *See also* Heart valve prostheses
anatomy of, 3:407–410
bioprosthetic, 3:413–415
computational simulation of function of, 3:422–423
effective orifice area of, 3:415–416
hemodynamics of, 3:427–428
mechanical, 3:411–413
native structure of, 3:427–428
tissue engineered, 6:392
- Heart vibrations, fundamental aspects of, 5:282–283
- Heat. *See also* Temperature entries; Therm- entries
defined, 6:312
as a thermal dilution indicator, 2:26–27
- Heat and cold therapy, 3:462–477
history of, 3:463–464
physiological effects of, 3:464–466
temperature-change devices and, 3:466–475
- Heat balance/production/loss, in infant incubators, 4:146–148
- Heat dissipation, in X-ray tubes, 6:606–608
- Heat flowmeter, in cryosurgery, 2:372
- Heat generation, metabolic, 1:189, 191
- Heating
as a hospital problem, 6:112
superficial, 3:467
tissue injury from, 6:368–369
- Heat loss
by human body, 1:191
predictable, 2:28
- Heat-related hyperemic response, 2:379
- Heat sterilization, 6:275–276
dry, 6:275
moist, 6:275–276
- Heat transfer
effects of blood perfusion on, 1:189–190
in human body, 1:191
- Heat transfer components, in microbio reactors, 4:389
- Heavy ion radiotherapy, 6:1–13
in cancer therapy, 6:10–11
electron-beam, 6:4–6
fast and slow neutron radiotherapy, 6:6–8
future developments in, 6:11–13
- Heimlich maneuver, 2:55–56
- Helical computed tomography, 2:233
- Helical CT scan, 2:238
- Helical tomotherapy, 6:398–399
- Helicobacter pylori* urease analyzer (HPUA), 4:105
- Helium dilution, 5:438
- Hematocrit (HCT), 2:87
effect on blood viscosity, 1:501f
impedance plethysmography and, 4:130
- Hematocrit measurement, 1:208
optical sensors in, 5:170–171
- Hematogenous infection, 1:117–118
- Hematological malignancies, treatment of, 4:606
- Hematology, automated cytology in, 2:404
- Hematology analyzers, 2:89t
automated, 2:410
- Hematology systems, WBC differential analysis on, 2:414–419
- Hematopoietic stem cells (HSCs), 6:382
- Hemiarthroplasty, 3:514
- Hemiarthroplasty hip replacements, 3:522
- Hemisurface hip replacements, 3:522
- Hemobilia, 3:391
- Hemocytometer, 2:82–83
- Hemodynamic evaluation, of prosthetic heart valves, 3:439–441
- Hemodynamic monitoring, 4:565–576. *See also* Blood pressure monitoring
bedside, 4:567–568
cardiac output determination in, 4:573–574
checking dynamic response in, 4:570–572
computerized decision support in, 4:575
heart, 4:566–567
measurements in, 4:568–569
neonatal, 4:593–594
signal amplification, processing, and display in, 4:572–573
theory behind, 4:566
- Hemodynamic monitoring system, technical management of, 4:593–594
- Hemodynamic parameters, alarming based on, 4:574
- Hemodynamics, 3:477–497. *See also* Blood entries; Heart
arterial system and, 3:483–492
heart–arterial coupling and, 3:492–494

- Hemodynamics (*Continued*)
heart valve, 3:427–428
peripheral, 4:127–128
- Hemoglobin (Hb), 1:500; 2:14
in blood oximetry, 1:470–471
in blood oxygen transport, 1:467–468, 469
cardiopulmonary resuscitation and, 2:40–41
encapsulated, 1:520
oxygen saturation of, 6:523
PFCs and, 1:514
- Hemoglobin concentration (HGB), 2:87
- Hemoglobin solutions, 1:516–520
in clinical trials, 1:517t, 518–520
duration of action of, 1:518
formulation of, 1:516–517
hematopoietic effect of, 1:518
hemodynamic effects of, 1:517–518
- Hemolink, clinical trial of, 1:519–520
- Hemolysis, 1:461
- Hemopure, clinical trial of, 1:518–519
- Hemorrhage, gastrointestinal, 3:384–393
- Homeostasis, biomaterial failure and, 1:280
- Henderson–Hasselbalch method, 2:110
- Henry's law, 1:465; 4:19
- Hepatic artery thrombosis, 4:271. *See also* Liver entries
- Hepatic failure, diseases associated with, 4:268t
- Hepatic vein thrombosis, after liver transplantation, 4:271–272
- Hepatitis, from banked blood, 1:512–513
- Heterografts, 1:283
- Heuristic sequence alignment methods, 1:219
- Hewlett-Packard ear oximeter, 1:471
- Hi-ART II CT imaging system, 6:399–401
- Hi-ART II tomotherapy unit, 6:398–399
- Hidden Markov Model (HMM), 1:222
- Hierarchical storage management/compression, 5:349
- HIFU apparatus, 4:80f. *See also* High intensity focus ultrasound (HIFU)
- HIFU devices, 4:75–76
- High angle annular dark field (HAADF) mode, 4:478
- High blood pressure. *See* Hypertension entries
- High contrast spatial resolution, in CT scanners, 6:576–577
- High Density Avalanche Chamber (HIDAC) PET system, 5:411
- High density electrode arrays, 2:138
- High density polyethylene (HDPE), 1:333
- High dose rate (HDR) brachytherapy, 1:590–601; 6:26–27. *See also* High-dose-rate remote afterloaders
advantages and disadvantages of, 1:600
costs associated with, 1:599–600
quality assurance in, 1:599
shielding in, 1:598–599
treatment planning system for, 1:597–598
- High-dose-rate remote afterloaders
components of, 1:590–594
features of, 1:592t
safety features of, 1:594–596
- Higher brain function, MEG studies of, 1:244
- Higher order gradients, noise reduction using, 1:235–236
- High frequency current, early experiments with, 3:157
- High frequency flow interruption (HFFI), 3:500, 503
- High frequency jet ventilation (HFJV), 3:500
- High frequency jet ventilators (HFJVs), 3:501–502
during and after cardiac surgery, 3:507
- High frequency oscillatory ventilators (HFOVs), 3:500, 502–503
- High frequency ventilation (HFV), 3:497–514
airway pressure monitoring during, 3:505
equipment for, 3:500–505
future outlook for, 3:512
in children and adults, 3:508–509
increasing, 3:499–500
risks associated with, 3:509–511
theoretical basis for, 3:498–500
in neonates, 3:505–508
- High frequency ventilators
design classifications for, 3:500–502
design philosophy for clinical applications of, 3:503–504
limitations of, 3:504–505, 511
safety and effectiveness of, 3:509
working of, 3:498–500
- High intensity focus ultrasound (HIFU), 4:62, 75–76; 6:365. *See also* HIFU entries
additional applications of, 4:81
future perspectives in use of, 4:81–83
history of, 4:75
medical applications of, 4:76–83
principles of, 4:75
- Highly oriented pyrolytic graphite (HOPG), 1:300
- High molecular weight boron delivery agents, 1:576
- High-performance liquid chromatography (HPLC), 2:103, 104–105
- High-performance silicon micropump, 2:504
- High resolution cytometry, in cellular parameter measurement, 2:399–400
- High resolution electrocardiography, 3:49–50
- High resolution temperature measurements, using lock-in amplifiers, 6:330
- High resolution transmission electron microscopy (HRTEM), 4:478
- High technology tools, for exercise fitness, 1:393–398
- Hip implant bearings
friction factors and lubrication regimes in, 3:521t
types of, 3:521–522
- Hip implants, 1:271f
- Hip joint replacements, biomaterials for, 3:515, 516
- Hip joints. *See also* Artificial hip joints
anatomy and environment of, 3:515
ceramic-on-ceramic, 3:521
cross-linked polyethylene, 3:521
metal-on-metal, 3:520–521
stability of, 4:220–221
UHMWPE-on-metal/ceramic, 3:518–520
- Hip joint wear debris, biological response of, 3:518
- Hip prostheses
failure of, 1:314–315
fixation of, 5:195–196
- Histogram analysis, 2:401–402
- Histograms, 2:401
- Hitachi CLAS, 1:24
- HIV infection, of whole blood, 1:460. *See also* Human immunodeficiency virus (HIV) patients
- HIV transmission, from banked blood, 1:512
- HL-7 system-to-system interface, 1:24
- HMMER alignment tool, 1:222
- Hold-down force, in arterial tonometry, 6:404
- Holger–Nielson method, 2:36
- Hologic/Lorad full field digital mammography system, 4:305
- Holter analyzer (scanner), 1:13
- Holter monitor
ambulatory monitoring with, 1:12–13
clinical application of, 1:13t
- Home glucose monitoring, 3:395–396
- Home health care devices, 3:525–536
blood components, 3:531–532
blood pressure, 3:526–527
body fat, 3:530–531
body temperature, 3:529–530
body weight, 3:532–533
daily activity and sleep, 3:534
electrocardiogram, 3:527–528
heart and pulse rate, 3:528–529
nutrition, 3:533
respiration therapy and oxygen therapy, 3:534–535
urine components, 3:532
- Homeostasis, 1:167
- Home parenteral nutrition, 5:133
- Homogeneous atelectatic lung disease, 3:505
- Homogeneous obstructive lung disease, 3:507
- Homogeneous restrictive lung disease, 3:506
- Homograft, 1:283
- Homology, 1:221
- Homopolymers, 5:388
- Hook electrode, 3:125
- Hooke's law, 2:69
- Hoppe–Seyler, Felix, 1:469
- Horsley–Clark stereotactic device, 6:265f
- Hospital environment, problems with, 6:113–114
- Hospital facilities, problems with, 6:112–113

- Hospital information system (HIS), interfacing with radiology information system, 5:335
- Hospital safety program, 6:109–122. *See also* Total hospital safety program medical devices and instrumentation related to, 6:117–119 patient safety in, 6:119–120 tools for, 6:119
- Host-related risk factors, for implant-related infections, 1:114–115
- Hot packs, 1:190
- Hotspots, current density, 1:121
- Human allograft (cadaver skin), as a skin substitute, 6:173–174
- Human amnion, as a skin substitute, 6:174
- Human anatomy, 1:384f
- Human body, oxygen transport in, 5:209–210
- Human body temperature profile, simulated, 6:348f
- Human clinical models, of spine stabilization procedures, 3:582–585
- Human factors
alarms and, 3:540–541
automation and, 3:541
case studies of, 3:538–539
cognitive task analysis and, 3:540
labeling and, 3:541
legal influences that affect, 3:539
in medical devices, 3:536–547
methods to improve design of, 3:539–540
new issues involving, 3:542–544
reporting and, 3:541–542
user testing and, 3:540
work domain analysis and, 3:540
- Human Factors and Ergonomics Society, 4:320
- Human factors engineering perspective, 3:537–538
- Human immunodeficiency virus (HIV) patients, medical procedures for, 3:543–544. *See also* HIV entries
- Humanitarian use devices, 2:146
- Human joints. *See also* Joints characteristics of, 4:206t finite element analysis of, 4:218–219
- Humanoid phantom, 5:265–266
- Human organ weights, 5:572t
- Human patient simulators (HPS), 1:45–46
- Human performance, analyzing, 1:394
- Human spine, 6:230
- Human spine biomechanics. *See* Spine biomechanics
- Human thermal models, 6:347–348
- Humidity, in continuous positive airway pressure, 2:335–336
- Humoral immune response, 1:113
- Hunter–Roth intramyocardial electrode, 1:151
- Hurter and Driffeld curve (H&D curve), 6:142–143
- Hutchinson, 5:John, 430
- Hyaff, 1:340
- Hyaline cartilage, composition and structure of, 2:63–65, 66
- Hyaluronan (HA), 1:340
in tissue engineering, 1:370
- Hyaluronic acid, in tissue engineering, 6:384
- Hybrid animal imaging devices, 5:106
- Hybrid approach, to developing visual prostheses, 6:535
- Hybrid imaging instruments, 5:103–104
- Hybridization, DNA sequencing by, 2:433
- Hybridization stations, in array platforms, 4:369
- Hybrid lenses, in intraocular lenses, 4:239
- Hybrid SPECT/CT systems, computers in, 5:119–121
- Hydraulic exercise equipment, 1:397–398, 398–399
- Hydrocephalus, 4:1–18
devices for treating, 4:8–14
diagnosis with imaging, 4:5–8
diagnostic methods related to, 4:3–5
epidemiology of, 4:1
nonobstructive, 4:9
pathophysiology of, 4:2–3
physiology of, 4:1–2
symptoms of, 4:2–3
third ventriculostomy in, 4:14–15
- Hydrocolloids, in hydrogels, 1:135
- Hydrodynamic lubrication, 6:417
- Hydrofluoric acid (HF), 1:292; 2:4
in hydrophobic bonding, 2:3
- Hydrogel electrodes, EMG, 1:169
- Hydrogel layers, 1:348
- Hydrogels, 1:134, 135–136, 141, 276, 277, 338–339
electrosensitive, 5:391
in engineered tissue, 3:196
pH-sensitive, 5:391
stimuli-responsive, 5:391
temperature-sensitive, 5:391
- Hydrogenated diamond-like carbon (HDLC), 1:313–314
- Hydrogen electrode, standard, 1:122
- Hydrogen fluoride (HF), 2:4
in hydrophobic bonding, 2:3
- Hydrogen-induced cracking (HIC), 1:311
- Hydron, 1:339
- Hydrophilic polymers, 1:348
- Hydrophilic surfaces, 1:343
- Hydrophobic bonding, 2:3
- Hydrophobic polymers, 1:314, 348
- Hydrophobic protein sites, 1:344
- Hydrophobic surfaces, 1:343, 344
- Hydrotherapy
contrast bath, 3:464
convection, 3:468
- Hydroxides, alkaline, 1:32, 33
- Hydroxyapatite (HA), 1:260, 261, 314
as biomaterial, 1:108, 110
gel-casting of, 1:291
in vitro biochemistry behavior of, 1:364–365
sintered, 1:287
synthetic, 1:284, 285
- Hydroxyapatite coatings, 1:355, 356, 357–365
- Hygiene, use of UV in, 6:486–487. *See also* Health
- Hylan, 1:340
- Hyperalgesia, 6:439–440
- Hyperbaric chamber, 4:21
- Hyperbaric chamber facility design, 4:27
- Hyperbaric medicine, 4:18–29
acute exceptional blood loss anemia and, 4:26
acute thermal burns and, 4:26
acute traumatic ischemias and, 4:23–26
air embolism in, 4:22
approved indications for, 4:20–21
carbon monoxide poisoning and, 4:26
contraindications for, 4:21–22
cyanide poisoning and, 4:27
decompression illness and, 4:23
flaps and, 4:24–25
frontiers and investigational uses of, 4:27–28
historical background of, 4:18
physics of, 4:18–19
physiology of, 4:19–20
radiation tissue damage and osteoradionecrosis in, 4:26
refractory osteomyelitis and, 4:26
skin grafts and, 4:25
wounds and, 4:23–24
- Hyperbaric oxygenation (HBO), 4:29–33
- Hyperbaric oxygen therapy, approved uses for, 4:18t
- Hyperemic flow, during guidewire diagnostics, 2:355–357
- Hyperemic response
heat-related, 2:379
nerve-axon-related, 2:379–380
- Hyperglycemia, 1:17
- Hyperlipidemia, parenteral nutrition and, 5:132–133
- Hyperpolarized contrast agents, in magnetic resonance imaging, 4:294
- Hypertension. *See also* Ambulatory blood pressure monitoring arterial mechanics in, 1:89–90 biofeedback clinical outcome literature related to, 1:180 borderline, 1:15 defined, 1:491 drug-resistant, 1:15 with end-organ damage, 1:15 episodic, 1:15 labile, 1:15 in pregnancy, 1:16 white-coat, 1:15
- Hypertension monitoring, transthoracic bioimpedance and, 1:202
- Hyperthermia, 1:188–189. *See also* Systemic hyperthermia; Ultrasonic hyperthermia application modes for, 4:68–75 chemotherapy in conjunction with, 4:68 devices, 4:75–76 effect on tumors, 4:49 endocrine function and, 4:48 externally applied, 4:68–73 future perspectives in use of, 4:81–83 immune system and, 4:50–51 interstitial, 4:34–42

- Hyperthermia (*Continued*)
 medical applications of, 4:67
 nanoparticle therapy and, 4:49t
 radiation coupled with, 4:67–68
 step-down sensitization in, 4:49
- Hyperthermia and chemotherapy clinical trials, 4:70t
- Hyperthermia and radiation clinical trials, 4:69t
- Hyperthermia devices, commercially available, 4:47t
- Hyperthermia systems, commercially available, 4:45–46
- Hypertrophy, developing, 1:391
- Hypoglycemia, 1:17
- Hypoperfusion, 6:164
- Hypopnea, 6:220. *See also* Obstructive sleep apnea/hypopnea syndrome (OSAHS)
- Hypotension, orthostatic, 1:16
- Hypothalamus, thermoregulation and, 1:190, 191
- Hypothermia, 1:188–189, 191
- Hypothesis testing, parametrical, 6:251
- Hypothetical distribution
 comparing sample distribution to, 6:257–258
 one-sample *t*-test to compare one data sample to, 6:251–253
- Hypovolemia, 2:48
- Hypovolemic shock, 6:164
 management of, 6:168
- Hysteresis, in surface analysis, 1:348–349
- ¹²⁵I (iodine-125), physical characteristics of, 5:419–420. *See also* ¹³¹I-MIBG therapy
- Ice bath, for thermodilution cardiac output measurement, 2:30–31
- Icterus, 1:461
- Ideal scaffold, 1:290–291
 sol-gel-derived bioactive glass foam as, 1:292–293
- Idiopathic scoliosis etiology, theories of, 6:128–129
- IL-2 receptor blockers, 4:274–275
- Illnesses, preventable, 1:390. *See also* Disease entries
- Illuminating system, in the transmission electron microscope, 4:481
- Illumination fibers, requirements for, 3:306
- Image acquisition
 future trends in, 5:346
 for picture archiving and communication systems, 5:335–336
- Image analysis
 in automated cytology, 2:402–403
 computers in, 5:117–118
- Image artifacts, 5:342–343
- Image capture, in teleradiology, 6:305
- Image compression, in teleradiology, 6:306–307
- Image contrast, in magnetic resonance imaging, 4:286–288
- Image data privacy, in teleradiology, 6:307–308
- Image data security, in teleradiology, 6:309–310
- Image data sets, 5:346
- Image display, computers in, 5:117–118
- Image enhancement and restoration algorithms, smart, 6:351
- Image fusion, PET-CT, 2:244
- Image generation, Anger camera, 1:53f, 55–56, 57f
- Image Grabber, 6:536–537
- Image guidance, in tomotherapy, 6:401
- Image guided surgery (IGS), 4:538–539
- Image-in-flow WBC counting technique, 2:412
- Image-intensifier assembly (IIA), in fluoroscopic X-ray systems, 6:570
- Image intensifiers, 4:87–89
- Image intensifier systems, 6:557
- Image magnification, in computed tomography, 2:251–252
- Image noise, in computed tomography, 2:252
- Image processing, 5:346
 algorithms for, 5:340
 computers in, 5:117–118
 in visual prostheses, 6:541–542
- Image quality
 CT, 2:233–250
 in digital radiography, 5:344–345
- Image receptors, 6:556–557
- Image recognition, fourth Purkinje, 3:270
- Image recording system
 in the scanning electron microscope, 4:484
 in the transmission electron microscope, 4:482
- Image registration, 5:457–458
- Images, transmission through optical fibers, 3:307
- Image segmentation, 5:340
- Imaging. *See also* Magnetic resonance imaging; Mammography
 during ablative treatment, 6:374–375
 to assess genital engorgement, 6:153
 cardiac, 6:371
 cellular, 2:90–101
 contrast, 6:465–471
 diagnosis of hydrocephalus with, 4:5–8
 diffusion, 4:290–291
 diffusion tensor, 4:291
 Doppler, 6:462–465
 in erectile assessment, 6:157–158
 fiber optics in, 3:307
 fluorescence, 3:346
 functional magnetic resonance, 4:289–290
 gradient recalled echo, 4:288–289
 harmonic, 6:468–469
 nuclear medicine, 5:108–114
 nuclear perfusion, 3:254–255
 peripheral vascular noninvasive measurements, 5:245–249
 for radiation oncology, 6:30–32
 SPECT, 5:100–101
 stereotactic, 5:577–578
 temporal-spectral, 5:247–249
 3D, 6:465
- ultrasonic, 6:453–473
 vascular, 2:426; 4:291–292
- Imaging artifacts, 5:494
- Imaging atomic force microscope (AFM), 4:504–505, 506–510
- Imaging atomic force microscopy, evaluation of, 4:513–514
- Imaging books/reports, 4:343–344
- Imaging devices/equipment/instruments, 4:89–97
 animal, 5:104–106
 electronic portal imaging detectors, 4:89–90; 5:484
 hybrid, 5:103–104
 liquid ionization chambers, 4:90–92
 near-field, 4:436–439
 physical aspects of, 4:90
- Imaging fiber bundles
 flexible, 3:307–308
 rigid, 3:308–309
- Imaging modalities, based on fluorescent microscopy, 4:498–502
- Imaging modes, 6:468–471
- Imaging sequences, in magnetic resonance imaging, 4:288
- Imaging system, in the transmission electron microscope, 4:481
- Imaging techniques, for exercise stress testing, 3:253
- Imaging units, regulations related to, 2:174
- ¹³¹I-MIBG therapy, dosimetry of, 5:569–570. *See also* ¹²⁵I (iodine-125)
- Immature cells, in tissue engineering, 6:382
- Imittance measurement, acoustic, 1:99–101
- Immune adjuvant, use with T cell administration, 4:113
- Immune reactions. *See also* Immune response
 from banked blood, 1:513
 to implants, 1:112
- Immune response, 1:113. *See also* Immune reactions
 biomaterial failure related to, 1:280–281
 cells participating in, 4:597
- Immune system
 antibodies and, 4:597–599
 effect of UV exposure on, 6:476
 whole-body hyperthermia and, 4:50–51
- Immunization, 4:597
- Immunoassay analyzers, 1:21, 22t
- Immunoassays
 development of, 4:375
 in microbial detection, 4:376
- Immunoblotting, 4:603
- Immunocytochemistry, 2:411
- Immunolectron microscopy, 4:602
- Immuno-electrophoresis, 4:603
- Immunofluorescence, 3:345–346; 4:601–602
- Immunohistochemistry, 4:602
- Immunologically sensitive field-effect transistors (IMFETs), 4:98–110
 direct-acting (label-free), 4:100–102, 103–104
 future directions for, 4:108–109

- indirect-sensing, 4:102–103, 104–107
 practical limitations of, 4:107–108
 theory behind, 4:99–103
- Immunology, automated cytology in, 2:405
- Immunomodulation, transfusion-related, 1:513
- Immunosensors, 5:172–173
 piezoelectric sensors as, 5:363–365
- Immunosuppressive agents, 4:274t
 alternative, 4:275
- Immunosuppressive medications, 4:273
- Immunotherapy, 4:111–120
 activation and polarization of effector T cells, 4:112–113
 adoptive T cell immunotherapy of cancer in lymphopenic host, 4:115–116
 antitumor reactivity of T cell subsets, 4:114–115
 cancer, 4:605
 combined, 4:117–118
 effector T cell trafficking and proliferation, 4:113–114
 induction of tumor-reactive pre-effector T cells, 4:111–112
 redirecting effector T cells to tumor, 4:116–117
 use of immune adjuvant with T cell administration, 4:113
- Impedance, 1:99. *See also* Tissue impedance
 acoustic, 4:66t; 6:454–455
 electrode-electrolyte, 1:123–124
 at electrode-skin interface, 1:121
 frequency dependence of, 3:117–119
 interface, 1:124
 limit current of linearity and, 1:128–129
 skin, 1:131–137
- Impedance analysis, 3:483–485
 of tissue and suspended cells, 4:135–137
- Impedance cardiography waveforms, 1:201f
- Impedance mapping, 6:372
- Impedance plethysmography, 4:120–132; 5:238–239
 bioimpedance measurement fundamentals, 4:122–124
 characteristics of bioimpedance, 4:124–126
 instrumentation and applications for, 4:127–131
 laboratory applications of, 4:130
 methodology of, 4:120–122
 model-based relations for volume determination, 4:126–127
- Impedance plot, complex, 1:124
- Impedance–resistance measurements, in cryosurgery, 2:370–372
- Impedance spectroscopy, 4:132–144
 in adherent cell monitoring, 4:137–138
 data presentation and analysis related to, 4:134–135
 electrode–cell interface model and, 4:138–139
 instrumentation for, 4:133–134, 138
 for monitoring cell attachment/spreading, 4:139–140
 theory behind, 4:133
- as a transducer in cell-based drug screening, 4:140–142
- Implantable applications, capacitive electronic interfaces for, 2:7–10
- Implantable atrial defibrillators, 1:80
- Implantable biotelemetry systems, 1:421
- Implantable cardioverter defibrillators (ICDs), 1:69–70, 71–79
- Implantable controlled drug delivery devices, 2:439–440
- Implantable defibrillators, 2:407–409
- Implantable devices
 biocompatibility and, 1:104
 minute ventilation in, 1:204–205
- Implantable drug delivery devices
 history of, 2:440–442
 microelectro-mechanical systems for, 2:449–452
- Implantable electrical nerve stimulators, 3:26
- Implantable enzyme electrode, glucose sensors, 3:398–402
- Implantable lead system, 1:434–435
- Implantable microcapsules, 4:394
- Implantable microfabricated drug delivery system, 2:504
- Implantable neurostimulator, 1:434
- Implantable optical glucose sensors, 3:397
- Implantable pulse generator (IPG), 1:432; 6:225–226
- Implantable pump drug infusion systems, 2:500
- Implantation, spinal cord stimulator, 6:227–228
- Implantation-related risk factors, for implant-related infections, 1:114
- Implant–bone interface, in spine stabilization procedures, 3:569–572
- “Implant bursitis,” 1:109
- Implant-contiguous wound sepsis, prevention of, 1:117–118
- Implant-derived particles, effects of, 1:111
- Implanted biosignal monitoring electrodes, 1:130
- Implanted electric bone treatments, 1:562
- Implanted electrodes, 1:120–121. *See also* Implant electrodes
 in functional electrical stimulation, 3:353–357
- Implanted nerve electrodes, 3:355–357
- Implanted stimulation electrodes, 1:121
- Implant electrodes, 1:149–158. *See also* Implant electrodes
 historical background of, 1:150–152
- Implant failure, correlating with disease states and drugs, 1:112
- Implant-grade steel, 1:311–312
- Implant-induced alterations, of the mechanical environment, 1:110
- Implant movement, effects of, 1:111
- Implant-related infections, 1:113–118
 biofilm and, 1:115
 common bacteria that cause, 1:114
 diagnosis of, 1:116
 fate of material during, 1:116
 latent, 1:115
 outcomes of, 1:115–116
- preventing, 1:117–118
 risk factors for, 1:114–115
 treating, 1:116–117
- Implants. *See also* Cochlear prostheses
 bioactive glasses in, 1:286
 biocompatibility of, 1:256
 bone cement, 1:540
 carcinogenicity and, 1:112–113
 ceramic, 1:260–262
 cochlear, 2:133, 139
 composite, 1:262–264
 contraceptive, 2:344–345
 dental, 1:327–329; 6:426–427
 for degenerative bone disease, 6:234
 for degenerative disk disease, 6:234–237
 for developmental spine deformities, 6:234
 epiretinal, 6:534–535
 history of, 1:255
 immune reactions to, 1:112
 interfacial stability of, 1:284
 long-term problems with, 1:255
 materials for, 1:255
 prostate seed, 5:418–429
 resorbable, 1:255–256
 spinal, 6:229–240
 subretinal, 6:534
 tissue response to, 1:109–110
 transplants versus, 1:283–284
- Implant sites, characteristics of, 1:108
- Implant surfaces, 1:130
- Impressed currents, 1:238
- Impulse response, in compartment modeling, 6:433
- Incandescent ultraviolet radiation, 6:476
- Incident-accident investigation/reporting, 6:118–119
 methodology for, 6:119
- Inclusion complexes, formation of, 2:453
- Inconsistent data, in computed tomography, 2:255
- Incontinence, biofeedback clinical outcome literature related to, 1:182
- Incubator dynamics, 4:153–155. *See also* Infant incubators
- Incubator studies, 4:150–152
- Independent Component Analysis (ICA), 1:241–242
- Independent living aids, for the sight impaired, 1:447–448
- Indexers, in remote afterloaders, 1:591–592
- “Indication for use” statement, 2:146
- Indicator dilution, principle of, 2:26
- Indicator dilution approach, in peripheral vascular noninvasive measurements, 5:246
- Indicator dilution cardiac output measurement, fundamental equations for, 2:22
- Indicator dilution curve, fundamental equations for, 2:22–23
- Indicator dilution equations, application of, 2:23
- Indicator-mediated transducers, in optical sensors, 5:163

- Indirect blood pressure measurement, 1:485, 486–489; 4:568–569
- Indirect colorimetric measurement, 2:192–194
- Indirect Conversion digital radiography, 5:343
- Indirect enzyme activity measurements, 2:195–196
- Indirect-sensing IMFET, 4:102–103. *See also* Immunologically sensitive field-effect transistors (IMFETs) workings of, 4:104–107
- Individual change, versus group change, 5:451
- Indoor navigational aids, 1:453
- Induced fluorescence, 6:359
- Induced organ synthesis, for organ function loss, 6:183
- Inductance pneumography, 6:521
- Inductance respirometry, in neonatal respiratory monitoring, 5:16
- Inductance strain gages, 6:283
- Inductively coupled plasma (ICP), 1:356
- Inductive plethysmography, 4:131; 5:378
- Inductive signaling, in tissue engineering, 6:387–388
- Industrial-Medical-Scientific (ISM) frequencies, 1:190
- Inert bioceramics, 1:284
- Inert biomaterials, 1:104
- Infant cardiopulmonary resuscitation (CPR), 2:57–58
- Infant heat transfer, infant incubators and, 4:148
- Infant incubators, 4:144–157
design actors related to, 4:146–150
dynamics of, 4:153–155
electromagnetic radiation in, 4:156
heat balance/production/loss in, 4:146–148
history of, 4:144–146
nonthermal environment of, 4:155–156
as sensory microenvironments, 4:155–156
for specialized purposes, 4:156
studies related to, 4:150–152
- Infants
EGG in, 3:93–94
indications for liver transplantation in, 4:269t
- Infant skin servo-controlled (ISC) incubator, 4:153
- Infections
from banked blood, 1:512
hematogenous, 1:117–118
implant-related, 1:113–118
- Infectious diseases, electron microscopic diagnosis of, 4:484–485
- Inferior lead ST-segment depression, in exercise stress testing, 3:248
- Inflammation
biomaterial failure related to, 1:280
in implant-related infections, 1:116
- Inflammatory bowel disease, 3:392
- Inflammatory-reparative response, tissue regeneration and, 1:109
- Inflation/deflation control, in blood pressure measurement, 1:487–488
- Informatics, medical, 4:309
- Information Notices, Nuclear Regulatory Commission, 2:171
- Information, radiation protection, 5:501
- Information systems
in healthcare, 4:310
radiology, 5:549–559
- Information technology, computers in, 5:118–119. *See* Computer information technology
- Informed consent, for contraceptive devices, 2:338
- Infrared imaging, 6:346–347
pathophysiological-based understanding of, 6:347–349
- Infrared light absorption sensors, 2:20
- Infrared light absorption technique, 1:478–479
- Infrared/optical spectroscopy, 4:327–329
- Infrared (IR) radiation, 5:66–67. *See also* IR entries
in biotelemetry systems, 1:421–422
use in medical diagnosis and therapy, 5:70
- Infrared reflectance, 5:142–144
- Infrared technologies, new generation, 6:349–353
- Infrared thermography, 6:360
- Infrared thermometers, 6:316
- Infrared tympanic thermometers (ITTs), 6:361
- Infusers, elastomeric, 2:500–501
- Infusion pump drug infusion systems, 2:498–502
- Inhalational anesthetics, 1:28
history of, 3:377–378
- Initial values, law of, 1:167
- Injection, needleless, 2:503
- Injurious forces/mechanisms, as a hospital problem, 6:109–110
- Injury
to cartilage, 2:72
mammalian response to, 6:179–184
from nonionizing radiation, 5:69
raised intracranial pressure related to, 4:580–582
scale of functional deficit related to, 6:179–180
- Injury currents, 1:211
- Injury risk, during exercise, 1:397
- Inkjet technology, for cell patterning, 1:412
- Inline volume flow measurement, 3:327–329
- Inner Helmholtz plane (IHP), 1:123
- Innervation, of engineered tissue, 3:192
- Innovative medical devices, codes and regulations related to, 2:152
- Input power, ventilator, 6:506
- Input radiation levels, 6:572
- In situ* bone regeneration, 1:291
- Instability coefficients, EGG, 3:90
- Instantaneous blood pressure measurement, 5:235
- Instantaneous center of rotation, 4:207
- Institute for Biological Engineering (IBE), 4:316
- Institute for Diabetes Technology artificial pancreas, 5:229
- Institute for Medical Technology Innovation, 4:320
- Institute of Electrical and Electronics Engineers, 4:320
- Institute of Environmental Sciences and Technology, 4:320
- Instrumentation. *See also* Apparatus; Equipment; Instruments. *See also* Sexual instrumentation
acoustic immittance measurement, 1:99
anterior and posterior spinal, 3:579–580
auditory evoked potential, 1:98
biofeedback, 1:168–169
biomagnetic, 1:231–242
for cell defining, enumerating, and isolating, 4:603–604
colorimetry, 2:189–190
ECG, 3:39–41
echocardiographic, 3:5–6
EEG biofeedback, 1:171–173
electrodermal biofeedback, 1:173–174
electrosurgery interference with, 3:173
for endoscopic procedures, 4:535–537
fluorescence-measurement, 3:343–345
in a hospital safety program, 6:117–119
impedance plethysmography, 4:127–131
impedance spectroscopy, 4:133–134, 138
laser Doppler flowmetry, 2:379–381
linear variable differential transformer, 4:253–254
microdialysis sampling, 4:402
for minimally invasive surgery, 4:535–539
for non-endoscopic procedures, 4:537–538
nuclear medicine, 5:90–106
optical sensor, 5:163–164
otoacoustic emission, 1:102
positron emission tomography, 5:415–417
pulmonary physiology, 5:434–435
radiation protection, 5:500–520
sexual, 6:149–163
single photon emission computed tomography, 2:280–281
speech audiometry, 1:96
spinal, 6:132–133
temperature biofeedback, 1:170–171
- Instruments. *See also* Instrumentation
colposcopy, 2:200–202
critical care, 1:22t
heart rate variability, 1:175–176
respiratory sinus arrhythmia, 1:174
- Instrument Society of America, 4:320
- Insufficient data quantity, in computed tomography, 2:253–255
- Insulin-like growth factor II (IGF-II), 1:290
- Insulin pump, with ambulatory blood glucose monitor, 1:17
- Integra, as a skin substitute, 6:177–178
- Integral dose, healthy tissue, 6:43
- Integral response, for a healthy organ, 6:45–46

- Integrated circuits, in temperature measurement electronics, 6:331–332
- Integrated-circuit temperature sensor, 4:157–162
 applications for, 4:160–161
 circuits and devices in, 4:159–160
 future of, 4:161–162
 theory behind, 4:158–159
- Integrated gas analyzers, anesthesia machine, 1:40
- Integrated human–machine–environment systems, in anesthesia monitoring, 1:44
- Integrated monitors, anesthesia machine, 1:39–40
- Integrated optic oxygen sensor chip, 5:206
- Integrating the Healthcare Enterprise (IHE), 5:334
- Integration
 in electromyography, 3:107
 in microbioreactors, 4:392
- Integrator–inverter circuit, 1:194
- Integrity, in teleradiology, 6:307–308
- Intelligent electronic travel aids, 1:450–451
- Intelligent human–machine interface, in anesthesia monitoring, 1:44
- Intelligent navigational aids, for the visually impaired, 1:453
- Intensifying screen
 absorption efficiency and conversion efficiency of, 6:139–140
 physical properties of, 6:139
 in screen-film systems, 6:138–140
- Intensity-modulated arc therapy (IMAT), 5:492, 523
- Intensity-modulated radiation therapy (IMRT), 5:520–525, 460–461
 computed tomography simulation for, 2:273–274
 gel dosimeters in, 5:491–492
- Interaction coefficients, in radiation measurement, 5:465–466
- Interaction types, in radiation therapy dose calculation, 5:535
- Interactive localization device (ILD), 6:270
- Interbody cages, 6:235
 lumbar, 3:580–581
- Interbody fusion cage stabilization, 3:578–579
- Intercalated disks, 2:43
- Intercapillary space, 2:199–200
- Intercompartmental fluid shifts, impedance plethysmography and, 4:129
- Intercostal EMG measurements, 6:521
- Interface electronics, in biotelemetry systems, 1:418
- Interface impedance, 1:124
- Interfaces, in continuous positive airway pressure, 2:335
- Interfacial bonding, of bioactive glasses, 1:286
- Interference, in biomedical electrodes, 1:121
- Intergranular corrosion, 1:311
- Interictal spikes, 1:245
- Interlaminar hooks, 3:569–570
- Intermediate zones, of articular cartilage, 2:64, 65
- Intermittent mandatory ventilation (IMV), with anesthesia machines, 1:38
- Intermolecular forces, surface protein adsorption and, 1:344–345
- Internal anal sphincter (IAS), 1:62, 64
- Internal measurement devices, for female sexual behavior assessment, 6:150–152
- Internal noise standards, 1:158–159, 160
- Internal rate of return, for Mount Sinai total laboratory automation, 1:27
- Internal sensors, in neonatal blood gas measurement, 5:23–24
- International Atomic Energy Agency (IAEA), 2:154
- International Biometric Society ENAR, 4:320
- International College of Surgeons, 4:320
- International Commission on Radiation Protection (ICRP), 2:153
- International Commission on Radiation Units and Measurements (ICRU), 2:153–154; 5:542
- International Council for Science (ICSU), 4:316
- International Electrotechnical Commission (IEC)/American National Standards Institute (ANSI), 2:154
 audiometry standards by, 1:92
- International Federation for Medical and Biological Engineering (IFMBE), 4:316
- International IR imaging activities, 6:353–354
- International Organization for Medical Physics (IOMP), 4:316
- International pacemaker codes, 5:218t
- International radiation concepts/units, 5:504t
- International Society for Magnetic Resonance in Medicine, 4:320
- International Society for Neuronal Regulation (ISNR), 1:172
- International Society for Optical Engineering (SPIE), 4:321
- International Standards Organization (ISO), audiometry standards by, 1:92
- International Union for Physical and Engineering Sciences in Medicine (IUPESM), 4:316
- Internet, medical information on, 4:309–310
- Internet access
 for the sight impaired, 1:447
 using augmentative and alternative communication systems, 2:207
- Interoperative electron-beam radiation therapy (IOERT), 6:5
- Interstitial hyperthermia, 4:34–42
 clinical studies with microwave and RF devices, 4:37
 electromagnetic heating and, 4:34–35
 laser devices and, 4:37–39
 microwave devices and, 4:36–37
 radio frequency devices and, 4:35–36
 thermal dose and heat transfer in, 4:34
- Interstitial hyperthermia application, 4:74
- Interstitial ultrasound, 4:39–40
- Interstitial ultrasound thermal therapy, 6:365
- Interterritorial matrix (ITM), 2:65
- Interval distribution, in EEG analysis, 3:69
- Interval-amplitude analysis, EEG, 3:69
- Interval modulator, 5:221
- Interventional cardiology, 1:601
- Interventional MRI, 6:270–271
- Intervertebral disk ablation, 6:378
- Intervertebral disk (IVD) prostheses, fixation of, 5:197
- Intervertebral disks, 2:63; 3:551–552; 6:230
 artificial, 3:587–588
 lumbar spine anatomy of, 3:551f
- Intervertebral stiffness matrix, 6:126
- Intestinal tract
 effects of parenteral nutrition on, 5:130–131
 nanoparticles in, 5:9
- Intima layer, in arterial walls, 1:85
- Intraabdominal pressure, 1:64
- Intraaortic balloon pump, 4:162–171.
See also Cardiac cycle
 automatic control of, 4:169–170
 clinical applications of, 4:165–166
 complications associated with, 4:166
 control of, 4:167
 counterpulsation principle and, 4:164
 equipment for clinical application of, 4:166–167
 future developments related to, 4:170
 historical perspective on, 4:164–165
 indications and contraindications for, 4:165–166
 left ventricular pump failure and, 4:164
 myocardial oxygen balance and, 4:163–164
 optimization of, 4:168–170
 timing associated with, 4:167–168
 triggering in, 4:167
- Intraarterial administration, of boron containing agents, 1:576–577
- Intraarterial catheter, in neonatal blood gas measurement, 5:23
- Intraarterial pO_2 measurement, continuous, 5:212–213
- INTRABEAM System, for intracranial lesions, 6:24–26
- Intracardiac echocardiographic examination, 3:17
- Intracardiac impedance, 1:207
- Intracardiac shunts
 detection and assessment of, 2:16–18
 Fick cardiac output technique and, 2:14–15
- Intracardiac transducers, 3:2
- Intracavitary hyperthermia application, 4:73–74
- Intracavitary hyperthermia devices, 4:75–76
- Intracellular water (ICW) volumes, in dialysis patients, 1:212

- Intracerebral delivery, of boron containing agents, 1:577
- Intracoronary tooth restorations, 6:424–425
- Intracranial compliance and pressure (ICP), 4:2
MRI-based measurement of, 4:6–8
- Intracranial hemorrhage monitoring, neonatal, 5:28
- Intracranial lesions, INTRABEAM System for, 6:24–26
- Intracranial pressure
analysis methods for, 4:582–584
in hydrocephalus, 4:2
monitoring devices associated with, 4:578–580
monitoring physiology of, 4:576–578
raised, 4:580–582
- Intracranial pressure monitoring, 4:576–588
literature related to, 4:580–582
model-based analysis methods for, 4:584–585
neonatal, 5:27–28
- Intracranial stereotactic procedures, 5:576–577
- Intracranial targeting, radiosurgery for, 5:575
- Intraesophageal pressure, in neonatal respiratory monitoring, 5:17–18
- Intrafascicular electrodes, 3:111
- Intramuscular electrodes, 3:353–355
- Intraneural electrodes, 3:120–121
- Intraocular lenses (IOLs), 4:234–241
accommodation and, 4:239
definition and function of, 4:235
historical overview of, 4:235–236
materials used for, 4:236
multifocal, 4:238–239
optical quality of, 4:237–238
parameters for, 4:236–237
- Intraoperative pressure (IOP), 1:423
- Intraoperative planning, for prostate seed implants, 5:419
- Intraoperative radiotherapy (IORT), 6:13–29. *See also* IORT entries
beam alignment devices in, 6:22–23
clinical rationale for, 6:15–16
clinical results of, 6:27–28
defined, 6:13
future directions for, 6:28
historical review of, 6:15–16
radiation safety issues in, 6:23–24
treatment apparatus for, 6:13–15
treatment applicators for, 6:21–22
treatment logistics in, 6:23
user surveys of, 6:16–17
- Intrapartum fetal pulse oximetry, 1:475–476
- Intrapleural space, 6:101–102
- Intraspinal stimulation, electrodes for, 3:357
- Intrathoracic bioimpedance, 1:204–208
- Intrathoracic fluid status, 1:205–206
- Intrathoracic impedance, 1:205
- Intrauterine devices, 2:345–346.
See also IUD
- Intrauterine surgical techniques, 4:171–181
for amnioexchange, 4:176–178
for amniopatch, 4:178
for bipolar umbilical cord occlusion, 4:179
for congenital cystic adenomatoid malformation and sacrococcygeal teratoma, 4:172–173
for lower urinary tract obstruction, 4:173–175
for myelomeningocele, 4:175–176
for twin-to-twin transfusion syndrome, 4:178–179
for valvuloplasty, 4:176
- Intravascular brachytherapy (IVB), 1:601–618
advanced topics in, 1:613–616
in the drug-eluting stent era, 1:615–616
terms related to, 1:616
- Intravascular brachytherapy devices
design considerations for, 1:609–613
theory and description of, 1:602–609
- Intravascular neonatal blood gas/biochemical sensors, 4:591–592
- Intravascular optical blood gas catheters, 5:168–169
- Intravascular temperature monitoring, 6:318
- Intravascular transducers, 1:485; 3:2
- Intravenous artificial pancreas, 5:226
- Intravenous drug delivery, microparticles for, 2:448
- Intravenous dyes, pulse oximetry and, 5:212
- Intrinsic heart rate, in rate-based arrhythmia detection, 1:71
- Intrinsic plexus, 1:68
- Intrinsic probes, 4:497
- Intrinsic spatial resolution, Anger camera, 1:60
- Invasive blood gas measurements, 1:465
- Invasive electric bone treatments, 1:562–563
- Invasive neonatal monitoring, future trends in, 4:595
- Invasive techniques, in neonatal blood gas measurement, 5:22–24
- Inventory, in biomedical equipment maintenance, 3:224
- Inverse AR filtering, 3:72
- “Inversion Method,” 2:35
- In vitro* arterial elasticity measurement, 1:85–86
- In vitro* effector T cell activation/polarization, 4:112–113
- In vitro* flow dynamics, of heart valve prostheses, 3:426–437
- In vitro* studies, of skin mechanical properties, 6:205–206
- In vitro* synthesis, for organ function loss, 6:183
- In vitro* testing, of prosthetic heart valves, 3:430–435
- In vivo* arterial elasticity measurement, 1:86–87
- In vivo* dosimetry, EPID system, 4:96
- In vivo* studies, of skin mechanical properties, 6:206
- In vivo* tissue compatibility assessment, 1:110
- In vivo* tumor-reactive pre-effector T cell induction, 4:111–112
- Iodine. *See* ¹²⁵I (iodine-125); ¹³¹I-MIBG therapy
- Ion-beam assisted deposition (IBAD), 1:346–347
- Ion channels, contributions to cardiac action potential, 3:143–144
- Ion cyclotron resonance (ICR), 1:560, 565–568
- Ion diffusion, at electrode-electrolyte interface, 1:125
- Ion-exchange chromatography, 2:104
- Ionic cell mechanisms, techniques to quantify, 3:142
- Ionic current
at electrode-electrolyte interface, 1:127
endogenic, 1:199
- Ion implantation, 1:346–347
- Ionization chambers, 4:90–92; 5:508–510
in radiotherapy dosimetry, 5:472–473
- Ionized oxygen electrode, 6:524–525
- Ionizing radiation. *See also* Linear no-threshold model of radiation effects
biological effects of, 4:181–185
biology/bystander effects of, 4:182
detection methods for, 5:93
detector materials used to measure, 5:93t
genomic instability and, 4:184
low dose exposures to, 4:183–184
time, dose, and fractionation associated with, 4:182–183
U.S. government divisions regulating, 2:154–158
- Ionizing radiation protection
recommendations/ regulations, organizations involved in, 2:153–158
- Ionizing radiation sterilization, 6:277–278
- Ionometry, in determining absorbed dose and air kerma, 5:468–469
- Ions, mobilities of, 3:132–134
- Ion selective electrode (ISE), 6:527
- Ion-sensitive field-effect transistors (ISFETs), 4:98, 185–198. *See also* ISFET entries
applications for, 4:195–196
development of, 4:189–190
miniature reference electrodes in, 4:193–194
packaging of, 4:191
site-binding model and, 4:186–189
- IORT dosimetry, 6:23. *See also* Intraoperative radiotherapy (IORT)
- IORT modalities, institutions practicing, 6:16t
- IORT techniques, 6:24–27
- IORT technology, 6:17–24
early, 6:17
recent, 6:17

- IORT units
 conventional versus mobile, 6:21
 dedicated, 6:17–18
 mobile, 6:18–20
 IP telephony, in office automation systems, 5:159
 IR absorption technique, 6:526. *See also* Infrared entries
 IR-based breast screening, concept
 validation in, 6:352–353
 IR cameras, 6:361
 IR fibers, 3:304
 IR image processing, smart, 6:351
 IR imaging
 in early breast cancer detection, 6:349
 for feature extraction/classification, 6:352
 IR imaging activities, international, 6:353–354
 Irradiance, weighted, 6:478–479
 Irradiation
 e-beam, 6:278
 in neutron activation analysis, 5:44
 Irritable bowel syndrome, biofeedback
 clinical outcome literature related to, 1:180
 Irritative zone, 1:245
 Isaacs, James, 2:37
 Ischemia detection, 1:210–211
 Ischemias, hyperbaric medicine and, 4:25–26
 Ischemic atherosclerotic disease, 2:50–51
 Ischemic colitis, 3:392
 Ischemic injury, after liver transplantation, 4:272
 Ischemic stroke, 2:52–53
 ISFET circuits, 4:191–193. *See also* Ion-sensitive field-effect transistors (ISFETs)
 ISFET field-effect transistor-based chemical sensor, 6:527
 Isocyanates, for polyurethanes, 1:337
 Isodose charts, 2:131–132
 Isodose curves, 6:589
 Isoelectric focusing (IEF), 2:106
 Isokinetic training, 1:392, 393
 Isometric exercises, 1:391–392
 Isometric muscle tests, 6:63
 Isosbestic point, 1:470
 Isotactic polypropylene, 1:335
 Isotonic resistive training, 1:392, 393
 Isotope selection, for prostate seed implants, 5:419
 Isotropic fluidized-bed pyrolytic carbons, 1:300, 302–303
 Italy, infrared imaging in, 6:354
 Iterative reconstruction techniques, 2:246–247
 Itrel I neurostimulator, 1:432
 IUD, with danazol, 6:154. *See also* Intrauterine devices

 Japan, infrared imaging in, 6:353
 Japanese clinical trials, of BNCT for brain tumors, 1:579

 Jaw. *See* Mastication; Masticatory system; Tooth/jaw biomechanics
 Joint biomechanics, 4:199–228
 Joint Commission on Accreditation of Healthcare Organizations (JCAHO), 2:154; 6:560
 recommendations, 6:115
 Joint Committee on Engineering in Medicine and Biology (JCEMB), 4:312
 Joint degeneration, mechanical factors associated with, 4:212–213
 Joint distribution problem, 4:215–220
 Joint lubrication regimes, 6:417
 Joint mechanics, theoretical analysis in, 4:213–214
 Joint motion, 4:204, 209–210. *See also* Joint movement; Kinematics
 equations of, 4:211–212
 planar, 4:207
 Joint movement, control of, 3:128
 Joint prostheses, wear of, 1:314–315
 Joint replacement, 2:73
 total, 4:540–541
 Joints, 2:63
 anatomical models of, 4:216
 cartilaginous, 4:202f
 characteristics of, 4:206t
 combined rotations of, 4:210
 diarthroidal, 4:212
 effects of motion and external loading on, 4:204
 fibrous, 4:203f
 intraarticular structures of, 4:199–201
 kinematics of, 4:204–211
 kinetics of, 4:211–212
 mathematical and mechanical models of, 4:212–220
 phenomenological models of, 4:216
 rotation in 3D space, 4:209–210
 stability of, 4:220–224
 surface modeling of, 4:214–215
 synovial, 4:200–201f
 types of, 4:199–201
 Joint system, terminology and definitions related to, 4:205
 Josephson, Brian, 1:231
 Journal bearing wear geometry, 1:316
 Journals
 biomaterials-related, 1:269t
 medical physics, 4:335–337

 Kalman filtering, 3:72
 Keidel vacuum tube, 1:455
 Kelvin temperature scale, 6:312
 Keratoplasty, 6:378
 Kety compartment model, 6:431
 Kilovoltage calibration, in X-ray equipment, 6:568–569
 Kinematic analysis, 1:394
 data collection in, 4:207–209
 methods of, 4:207–209
 Kinematics, joint, 4:204–211
 Kinesins, 5:183–184
 Kinetic energy released per unit mass (kerma), 5:466
 Kinetic imaging analyses, 5:414
 Kinetic measurements, 2:194–196

 Kinetics, joint, 4:211–212
 Kinetic tracer curves, analysis of, 6:435
 Kitchen environments, virtual reality, 6:76
 KLAS information systems reports, 5:558t
 Knee joint, stability of, 4:221
 Knee prostheses, fixation of, 5:196
 Knee studies, 4:213
 Knowledge Work System (KWS), 5:151
 Korea, infrared imaging in, 6:353
 Korotkoff sounds, 1:14, 486
 Kramer, Kurt, 1:469
 Krimer's electropuncture of the heart, 1:150
 Krogh, Marie, 5:430
 Kruskal–Wallis test, for unmatched samples, 6:259
 Kurzweil, Ray, 1:453–454
 Kyphoplasty, 6:234
 Kyphosis, 6:232

 Label-free IMFET, 4:100–102
 Labeling
 codes and regulations for, 2:147, 148–149
 human factors and, 3:541
 Labile cells, tissue regeneration and, 1:109
 Labile hypertension, 1:15
 Lab-on-a-chip (LOC) systems, 4:420
 Laboratory applications, of impedance plethysmography, 4:130
 Laboratory automation, financial perspective on, 1:24
 Laboratory automation system (LAS), 1:23–24
 Laboratory grade flowmeter, calibration using, 5:374
 Laboratory Information System (LIS), 1:23–24
 Labra, 4:199
 Lactic–glycolic acid copolymers, as biomaterials, 1:107
 Lambert–Beer law, 1:470
 Lamellae, 1:524
 Laminar flow, 1:504–505
 in microbioreactors, 4:390
 Language
 assessment of, 2:215–216
 assistive devices related to, 2:221–222
 language analysis software, 2:214
 language assessment, for augmentative and alternative communication systems, 2:208
 language disorders, 2:211
 new directions in, 2:224–225
 language impairment, specific, 2:211
 language intervention, 2:219–220
 language learning devices, 2:221
 language use devices, 2:221
 laparoscopic cholecystectomy, 4:525
 laparoscopy, 3:185; 4:539
 Laplace transform, 1:194
 large bore CT, 5:529
 large diameter blood vessels, in coronary artery replacement, 6:394
 large-scale finite element (LSFE) models, 4:220
 laryngeal pathology, 2:211

- Laryngeal prosthetic devices, 4:229–234
 electronic artificial larynx, 4:230–231
 voice prostheses, 4:231–234
- Laryngectomy, voice after, 4:230
- Larynx, electronic artificial, 4:230–231
- Laser ablation, 4:528; 6:365–366, 378
 nanoparticle fabrication via, 5:3
- Laser cane, 1:449
- Laser classification schemes, 2:182t
- Laser devices, interstitial hyperthermia
 and, 4:37–39
- Laser Doppler flowmetry (LDF), 2:378
 clinical applications of, 2:382–383
 instrumentation for, 2:379–381
 in peripheral vascular noninvasive
 measurements, 5:244–245
- Laser Doppler velocimetry, 3:332–335
- Laser emission, lines of, 2:398t
- Laser eye correction, uses of UV in, 6:486
- Laser heating devices, 3:470
- Laser light scattering, for blood cell
 counting, 2:85–86
- Laser probe, single-point, 2:379, 381
- Laser refractive surgery, 6:378
- Laser regulations, state, 2:158t
- Lasers
 books/reports on, 4:346
 monochromatic radiation from, 6:478
 use in medical diagnosis and therapy,
 5:70
- Laser scanning confocal microscopy
 (LSCM), 2:92–93; 4:452–453
- Laser scanning method
 of cutaneous blood flow measurement,
 2:380–381
 validation of, 2:381
- Latency, in eye movements, 5:138
- Latent image formation, in screen-film
 systems, 6:141–142
- Latent implant-related infection, 1:115
- Lateral bending
 of the middle and lower cervical spine,
 3:556
 of the occipital-atlantoaxial complex, 3:554
- Lateral force microscopy (LFM), 4:508
- Lateral resolution, in confocal microscopy,
 4:463–464
- Law of initial values, 1:167
- Law of large numbers, 5:534
- Lead connector, 5:222
- Lead discovery, use of microarrays in,
 4:370
- Lead field theory, 1:203–204
- Leads, in exercise stress testing,
 3:248–249
- Lead systems, in exercise stress testing,
 3:248
- Leakage radiation, 6:564–565
- Leak circuit modification, in continuous
 positive airway pressure, 2:329–330
- Learning disabilities, 2:211
 cognitive training for students with, 6:72
- Lea's shield, 2:341
- Lecithin organogel, 2:463–464
- Left-to-right intracardiac shunt, 2:16
- Left ventricle (LV) systolic function, ways
 of characterizing, 3:480
- Left ventricular pump failure,
 pathophysiology of, 4:164
- Legal influences, on human factors, 3:539
- Legal issues
 related to computer-based patient
 records, 4:357–358
 in teleradiology, 6:310
- Legislation, sunbed-related, 6:485–486
- Leksell Gamma Knife, 3:367–377. *See also*
 GammaKnife unit
- Lens aberrations, in electron microscopy,
 4:480–481
- Lenses
 intraocular, 4:234–241
 in the transmission electron microscope,
 4:481
- Lesion identification, in computer-assisted
 detection/diagnosis, 2:293–294
- Lesions
 intracranial, 6:24–26
 missed, 2:300
- Lesion segmentation, in computer-assisted
 detection/diagnosis, 2:294–295
- Letter-to-sound text conversion, 1:447
- Leucosep tubes, separation of peripheral
 blood mononuclear cells using,
 1:463–464
- Leukocytes, 1:503. *See also* White blood
 cells (WBCs)
- Liac mobile intraoperative radiotherapy
 unit, 6:20
- Lie detectors, 1:133
- Life processes, molecular, cellular, and
 systems levels of, 1:188
- Life Shirt, 1:142
- Life support systems, monitoring neonatal,
 5:29–30
- Lifetime resolved microscopy, 4:500–501
- Lifetime variables, in clinical studies,
 6:262
- Lift force gas flow sensor, 5:371
- Ligaments, 4:201
 collagen in, 4:241–242
 components of, 4:241–243
 elastic fibers in, 4:242–243
 failure mechanisms of, 4:247–248
 fiber-matrix interactions in, 4:243
 masticatory system, 6:417, 421–422
 measuring the properties of, 4:246–247
 properties of, 4:241–252
 proteoglycans in, 4:242
 repair of, 4:248
 skeletal, 4:243–245
- Ligand binding, 2:404t
- Ligation chain reaction (LCR), 5:385–386
- Light absorption, signal detection in,
 2:413–414
- Light delivery, with optical fibers, 3:179.
See also Fiber optics
- Light detectors, confocal microscope,
 4:455–457
- Light-emitting diodes (LEDs)
 in anesthesia machines, 1:41
 in biotelemetry systems, 1:422
 in pulse oximetry, 1:472
- Light field alignment, in X-ray equipment,
 6:565
- Light localizer illumination, in X-ray
 equipment, 6:566
- Light pipe, with Anger camera, 1:54, 55f
- Light scattering
 signal detection in, 2:413
 single-cell, 2:393
- Light scattering spectroscopy, 3:311–312
- Light scattering WBC counting technique,
 2:411–412
- Light sources
 fluorescence excitation, 4:491
 in optical sensors, 5:163–164
- Lillehei, C. Walton, 1: 432
- Limb plate electrode, 1:138
- Limbus trackers, 3:276–277
- Limit current of linearity, 1:128–129
- Limit voltage of linearity, 1:129
- Limit voltages, 1:128–129
- Limoge currents, 3:28–29
 transcutaneous cranial electrical
 stimulators using, 3:29–30
- Linear accelerators, 5:578, 579
 quality assurance for, 5:547
- Linear attenuation coefficient, 6:591–592
- Linear-energy transfer (LET), 6:2–3
- Linearity, currents of, 1:128–129
- Linearization, in ocular motor recording,
 5:146–147
- Linear low density polyethylene (LLDPE),
 1:333
- Linearly Constrained Minimum Variance
 (LCMV) beamformer, 1:241
- Linear no-threshold model of radiation
 effects, 4:183
- Linear polyethylene, as biomaterial,
 1:106–107
- Linear polymers, as biomaterials, 1:274
- Linear variable differential transformers
 (LVDTs), 4:252–257
 fabrication of, 4:252–253
 instrumentation for, 4:253–254
 medical applications of, 4:255–257
- Line immunoprobe assay (LIPA), 4:376
- Liners, dental, 1:324–325
- Lipemia, 1:461
- Lipid-based nanoparticles, 2:484–486
- Lipid system, 5:127
- Liposomal drug carriers, 2:466–473
- Liposomal formulations, lipid component
 used in, 2:467–468
- Liposomes
 main therapeutic applications of,
 2:470–473
 preparing, 2:468–469
 stability of, 2:469–470
 ultradeformable, 2:479–480
- Liposome system, characteristics of, 2:467t
- Liquid adsorption chromatography, 2:104
- Liquid chromatography (LC), 2:103, 104–
 105
 in microdialysis sample analysis, 4:409
- Liquid crystals (LCs), 6:315
- Liquid expansion thermometers, 6:313
- Liquid-in-glass thermometers, 6:356
- Liquid ionization chambers (LIC), 4:90–92
- Liquid-liquid chromatography, 2:104
- Liquids, surface energy of, 1:343

- Liquid-solid chromatography, 2:104
- Liquid ventilation, with
perfluorochemicals, 1:516
- Lister, Joseph, 1:267
- Literature, intracranial pressure
monitoring, 4:580–582
- Lithographie, Galvanik, Abformung
(LIGA), 4:527
- Lithotripsy, 4:258–266. *See also*
Lithotripters
acute and chronic injury with, 4:263t
advances in, 4:263–264
clinical results of, 4:261
history and evolution of, 4:258
principles of, 4:258–260
safety and efficacy advances in, 4:264
stone fragmentation in, 4:260–261
tissue injury with, 4:261–263
treatment strategy modifications in,
4:263–264
- Lithotripters
advances in, 4:263
literature comparison of, 4:262t
- Liver(s). *See also* Hepat- entries
bioartificial, 4:393
effects of parenteral nutrition on, 5:131
- Liver cancer, high intensity focus
ultrasound for, 4:76–77
- Liver susceptometer, 1:238f
- Liver tissue, engineered, 3:202–203
- Liver transplantation, 4:266–277
acute cellular rejection after, 4:272
alternative immunosuppressive agents
and, 4:275
antiproliferative agents and, 4:275
biliary complications following, 4:272
calcineurin inhibitors and, 4:275
chronic rejection following, 4:272–275
contraindications to, 4:269
corticosteroids and, 4:273–274
early complications of, 4:271
etiology of diseases requiring, 4:268–269
history of, 4:266–267
IL-2 receptor blockers and, 4:274–275
immunosuppressive medications and,
4:273
indications for, 4:267–268
initial results of, 4:267
ischemic and preservation injury
following, 4:272
portal and hepatic vein thrombosis after,
4:271–272
posttransplantation management
associated with, 4:271
recipient characteristics and
prioritization for, 4:269–270
recurrence of primary disease following,
4:273
source of organs for, 4:270–271
T-cell depleting agents and, 4:274
- Living dermal replacement (LDR),
6:191
- Living skin equivalent (LSE), 6:191–192,
381
- Loading, effect on joints, 4:204
- Local heating, of organs, 1:190
- Localized drug delivery, 2:438
- Localized infusion, using microdialysis
sampling, 4:404–405
- Locally weighted polynomial regression
(LOWESS) fit, 1:223
- Lock-in amplifiers, in temperature
measurement electronics, 6:330
- Long duration microsurgery,
electropharmaceutical anesthesia in,
3:32
- Longitudinal intrafascicular electrodes
(LIFE)s, 3:111, 121–122
- Long-term enzyme electrode glucose
sensors, based on oxygen detection,
3:400–402
- Long-term peripheral nerve interfaces,
3:119–125
- LORETA (LOW Resolution
Electromagnetic Tomography)
algorithm, 1:240
- Lost wax casting, 1:325
- Low contrast detectability, in computed
tomography, 2:253
- Low contrast resolution, in CT scanners,
6:577
- Low density polyethylene (LDPE), 1:333
- Low-dose X rays, cancer risks associated
with, 2:259–260
- Low energy X-rays, 6:580–581
- Lower back pain, biofeedback clinical
outcome literature related to, 1:179
- Lower cervical spine
anatomy of, 3:550–551
biomechanics of, 3:555–557
stabilization of, 3:575–578
- Lower cervical spine instability, role of
environmental factors in, 3:561
- Lower extremity prosthetics, 4:552–554
- Lower GI bleeding, 3:391–392
- Lower GI hemorrhage, diagnosis of, 3:387t
- Lower limb orthotic devices, 6:88t
- Lower limb technical analysis form, 6:90f,
91f
- Lower urinary tract obstruction (LUTO),
4:173–175
- Low friction arthroplasty, 3:515
- Low-molecular weight boron delivery
agents, 1:573–575
- Low pressure chemical vapor deposition
(LPCVD), 2:4; 4:190
- Low resolution flow systems, in cellular
parameter measurement, 2:399
- Low temperature isotropic (LTI) carbon, as
a biomaterial, 1:273
- Low vision
defined, 1:443–444
prevalence of, 1:444t
reading aids for, 1:444–445
- Lubricants, 1:315
- Lubrication, 1:314, 315
- Lubrication regimes, 6:417
- Lucy 3D precision phantom, 5:265
- Lumbar interbody cages, 3:580–581
- Lumbar spine
anatomy of, 3:552
biomechanics of, 3:557–558
stabilization procedures for,
3:579–581
- Lumbar spine instability, role of
environmental factors in, 3:561–562
- Lumbar spine region
degeneration–trauma in, 3:565–566
surgical procedures and, 3:568
- Luminescence/fluorescence, gas sensors,
4:329
- Luminous emitters, 3:468–469
- Lung(s). *See also* Pneumo- entries;
Pulmonary entries; Respiratory
system
dead space in, 3:498
diffusing capacity of, 5:438–439
nanoparticles in, 5:8–9
natural frequency of, 3:498–499
zones of, 2:41–42
- Lung cancer
low dose CT screening for, 2:263–264
treatment planning for, 5:537–538
- Lung capacities, 2:39
- Lung compliance (elasticity), 6:517
- Lung disease
homogeneous atelectatic, 3:505
homogeneous obstructive, 3:507
homogeneous restrictive, 3:506
nonhomogeneous atelectatic and
restrictive, 3:506–507
nonhomogeneous obstructive, 3:507
- Lung injury, prevention of, 3:505
- Lung mechanics, parameters used in
monitoring, 6:515–518
- Lung performance, impedance
plethysmography and, 4:128–129
- Lung sounds, 4:277–282
analysis of, 4:280–281
recording and display systems for,
4:279–280
results and clinical applications of,
4:281–282
sound transducers and, 4:278–279
stethoscope and, 4:278
- Lung volume, 5:437; 6:515–516
measurement of, 6:520–522
- Lymphocyte activation, biomaterial failure
related to, 1:280–281
- Lymphocytes, 1:507, 508f; 2:82
- Lymphopenic host, adoptive T cell
immunotherapy of cancer in,
4:115–116
- Lysing, in microbio reactors, 4:390
- Lysozyme, 1:340
- M20 source, 1:242
- M100 response, 1:244
- Machine-based expert systems, 2:317
- Machine-produced radiation, codes and
regulations for, 2:171–177
- Machines, FDA regulation of, 2:155–156
- Macroduct system, 2:386–387
- Macroflux technology, 2:503
- Macrophages, 1:113; 2:82
biomaterial failure and, 1:280
tissue regeneration and, 1:109
- Macroretentive features, in dental
implants, 1:328
- Macular degeneration, 1:445
age-related, 5:293

- Macular function, primary evaluation of, 3:155
- “Mad Cow” Disease, 1:513, 517
- Magnetically confined electron beams, 6:11–12
- Magnetic disk (MD), 5:347
- Magnetic field exposure limits, 2:183t
- Magnetic Field Tomography (MFT), 1:240
- Magnetic marker monitoring (MMM), 1:248
- Magnetic resonance (MR), in cryosurgery, 2:372. *See also* MR entries
- Magnetic resonance angiography (MRA), 4:291–292; 5:245–246
- Magnetic resonance flow mapping, 3:330–332
- Magnetic resonance imaging (MRI), 4:283–298. *See also* MR entries; MRI entries
- applications for, 4:289–297
- to assess genital engorgement, 6:153
- breast, 4:294
- cardiac, 4:292–294
- in diagnosing implant-related infection, 1:116
- hyperpolarized contrast agents in, 4:294
- image contrast in, 4:286–288
- imaging sequences in, 4:288
- interventional, 6:270–271
- of musculoskeletal disease, 4:296–297
- in peripheral vascular noninvasive measurements, 5:245–246
- in radiosurgery, 5:578
- rapid, 4:289
- signal generation in, 4:283–285
- signal-to-noise ratios in, 4:288
- spatial encoding in, 4:285–286
- and spectroscopy books/reports, 4:345
- stereotaxis based on, 6:267–268
- Magnetic resonance spectroscopy (MRS), 1:246. *See also* Nuclear magnetic resonance (NMR) spectroscopy; MRS entries
- Magnetic shielding, 1:234–235
- Magnetic source imaging, 1:242
- Magnetic susceptibility plethysmography, 4:131
- Magnetic wind, oxygen analyzers, 5:200–201
- Magneto-acoustic ball microrheometer, 1:506f
- Magnetocardiography (MCG), 1:231, 246
- fetal, 1:246
- Magnetodynamic (dumbbell or autobalance) oxygen analyzers, 5:201
- Magnetoencephalography (MEG), 1:231, 232, 233, 234, 236, 237–238, 242. *See also* MEG–EMG coherence
- clinical applications of, 1:244–246
- fetal, 1:246–247
- Magnetogastrography (MGG), 1:248
- Magnetometers, 1:234–235; 6:521
- displacement, 5:16
- Magnetopneumatic (differential pressure) oxygen analyzers, 5:201–202
- Magnetopneumography, biomagnetic measurements and, 1:248
- Magneto-position transducer, 1:175
- Magnification, in electron microscopy, 4:480
- Magnitude-squared coherence function, in separating ventricular fibrillation from tachycardia, 1:79
- Mailing lists, in office automation systems, 5:156
- Mainstream sampling techniques, 1:479–480
- Male condoms, 2:338–339
- Male erectile dysfunction (ED)
- circumferential versus volumetric assessment of, 6:155–156
- treatment of, 6:158–159
- Male human sexual behavior, instruments and measurement of, 6:155–162
- Malignancies, hematological, 4:606
- Malignant tumors, external beam radiotherapy options for, 6:5t
- Malleable metal foil electrodes, 1:144
- Mallory–Weiss tear, 3:390
- Malysed, John, 5:429–430
- Mammalian healing process, regeneration versus repair in, 6:180
- Mammals, thermoregulation in, 1:190–192
- Mammographic X-ray equipment, quality control of, 6:573
- Mammography, 4:298–307
- antiscatter grid in, 4:301–302
- computed radiography for, 4:302
- digital spot, 4:303
- full field digital, 4:303–304
- phantom materials in, 5:266
- physics of, 4:299–300
- screen-film, 4:300–301
- stereotactic breast biopsy, 4:302–303
- Mammography images, reading, 4:305
- Mammography Quality Standards Act of 1992 (MQSA), 4:298; 6:560
- Mammography system, regulations related to, 2:178t, 179t
- Management Information System (MIS), 5:151
- Mandated web accessibility, for the sight impaired, 1:447
- Mandible, deformations of, 6:424
- Mann–Whiney *U* test, for unpaired samples, 6:258–259
- Manometric data, during anorectal manometry, 1:67t
- Manometry
- anorectal, 1:62–69
- esophageal, 3:229–233
- Manual muscle testing, 6:64–65
- apparatus for, 6:65
- Manual wheelchairs, 4:546
- Mapping
- cardiac, 6:371
- electroanatomic, 6:372
- impedance, 6:372
- multiple electrode, 6:371
- Marey, Etienne Jules, 1:137
- Market, regulatory pathways to, 2:144–145
- Marketing approval, for medical devices, 2:146
- Martensite phase, 1:1, 2
- Martensite structure, for Ni–Ti shape memory alloy, 1:3–4
- Masimo SET monitors
- in pulse oximetry, 1:472, 473f
- Masks, in continuous positive airway pressure, 2:335
- Mason–Likar electrode placement, in exercise stress testing, 3:248
- Massachusetts General Hospital Utility Multi- Programming System (MUMPS), 1:44
- Mass and heat balance, in microbioreactors, 4:386–387
- Mass attenuation coefficient (μ/ρ), 6:593
- Massive parallel signature sequencing (MPSS), 2:434
- Massive skin loss, current treatment of, 6:189–190
- Mass recovery, in microdialysis sampling, 4:412
- Mass response, in piezoelectric sensors, 5:360
- Mass spectrography, 1:479
- Mass spectrometer, 5:436
- Mass spectrometry (MS), 6:525, 526. *See also* Mass spectroscopy
- for dialysates, 4:411
- Mass spectrometry-based DNA sequencing, 2:434
- Mass spectrometry, 4:329–330
- Mass transport
- in tissue engineering, 6:387
- via blood, 1:188
- Mastication
- biomechanical models of, 6:411, 417–424
- deformation and strain in, 6:419
- dynamic models of, 6:423–424
- external forces in, 6:419
- internal forces/stresses in, 6:419
- static models of, 6:422–423
- Masticatory system. *See also* Mastication
- dentition and supportive structures in, 6:411–414
- ligaments in, 6:417
- material properties of, 6:419–420
- mechanical properties of bone in, 6:420
- mechanical properties of cartilaginous tissue in, 6:420–421
- mechanical properties of ligaments in, 6:421–422
- musculature of, 6:415
- skelatal components of, 6:414
- temporomandibular joint in, 6:415–417
- Matched samples
- Friedman test for, 6:259
- one-way ANOVA for, 6:254–256
- MATCH-HHH-ED mnemonic, 2:48
- Material-related risk factors, for implant-related infections, 1:114
- Materials. *See also* Porous materials
- biocompatibility of, 1:104–120
- classes of, 1:104–108
- composite, 1:108
- dose conversion between, 5:471
- metallic, 1:104–105
- in orthopedic devices, 5:188–190

- Materials technology, for prosthetic heart valves, 3:441–442
- Mathematical cardiac torso (MCAT), 5:121
- Mathematical models
of cardiac cells, 3:144–145
of joints, 4:212–220
- Mathematics books/reports, 4:349
- MATLAB, 6:263
- Matrix-assisted laser desorption/ionization (MALDI), 2:106–107
- Matrix phase, of resin-based composites, 6:94–95
- Matter, X-ray interaction with, 6:590–599
- Matthes, Karl, 1:469, 471
- Mature cells, in tissue engineering, 6:382
- Maxillofacial resin materials, 1:327
- Maximal Expiratory Flow–Volume (MEFV), 6:102
- Maximal heart rate studies, 3:257t
- Maximal voluntary ventilation, 5:439
- Maximizing performance, 1:387
- Maximum dose to peripheral dose ratio (MDPD), 5:583
- Maximum intensity projection, in computed tomography, 2:239
- Maximum safe temperature rise standards, for electrodes, 1:161–162
- Maxwell, James C., 1:197
- Mean arterial pressure (MAP), 1:486, 487, 491–499
- Mean corpuscular hemoglobin (MCH), 2:87
- Mean corpuscular hemoglobin concentration (MCHC), 2:87
- Mean free path (MFP), 6:593
- Mean platelet volume (MPV), 2:87
- Measurements, in electrophoresis, 3:138–139
- Mechanical environment, implant-induced alterations of, 1:110
- Mechanical events, in the cardiac cycle, 4:163
- Mechanical heart valves, 3:411–413, 445–446
flow dynamics past, 3:417–421
versus bioprosthetic heart valves, 3:446–447
- Mechanical models, of joints, 4:212–220
- Mechanical modulation, in engineered tissue, 3:199
- Mechanical properties
of acrylic bone cement, 1:546–547
of surfaces, 1:343–344
- Mechanical signaling, in tissue engineering, 6:388
- Mechanical stimulation components, in microbioreactors, 4:390
- Mechanical strain gages, 6:283
- Mechanical tactile stimulators, 6:293–295
- Mechanical vaporizers, in anesthesia machines, 1:41
- Mechanical ventilation, 6:107
- Mechanical ventilatory support
adverse reactions to, 6:512–513
goals of, 6:509–510
indications for, 6:509
- Mechanical work, thermoregulation and, 1:191
- Mechanics, as a hospital problem, 6:112
- Mechanoreceptors, 6:291–293
- Meckel's diverticulum, 3:392
- Meconium aspiration syndrome (MAS), 3:507
- Media layer, in arterial walls, 1:85
- Media processing, in computer-based patient records systems, 4:356–357
- Medical applications
of capacitive microsensors, 2:2
of carbon biomaterials, 1:301–306
of linear variable differential transformers, 4:255–257
micropower for, 4:428–434
using endoscopy, 3:180–186
- Medical books/reports, 4:337–357. *See also*
Medical physics literature
computerized tomography, 4:344
computers, 4:349
diagnostic radiological physics, 4:341–343
imaging, 4:343–344
light and lasers, 4:346
magnetic resonance imaging and spectroscopy, 4:345
mathematics and statistics, 4:349
nuclear medicine, 4:344–345
public education, 4:350
radiation biology, 4:348–349
radiation measurements, 4:348
radiation oncology physics, 4:338–341
radiation physics, 4:350–351
radiation protection, 4:346–348
radiological physics, 4:349
ultrasound physics, 4:345–346
- Medical Device Amendments of 1976, 1:268–269, 270, 277
- Medical device codes/regulations, 2:141–153
challenges related to, 2:152
for clinical studies, 2:147
for clinical trials, 2:143–144
for device description, 2:146
enforcement and penalties related to, 2:142–143
general requirements of, 2:143
labeling-related, 2:147, 148–149
market-related, 2:144–145
for nonclinical laboratory studies, 2:143
for nonclinical studies, 2:147
premarket notification exemptions from, 2:145
- Medical device management systems, computerized, 6:118
- Medical Device Manufacturers Association, 4:320
- Medical device modifications, codes and regulations for, 2:150, 151t
- Medical device reporting/corrections/removals, codes and regulations for, 2:149–150
- Medical devices
bioceramics as, 1:284
classification of, 2:141–142
cognitive task analysis and, 3:540
conformance with standards, 2:142
defined, 2:141; 6:110–111
description of, 2:146–147
education and training related to, 6:119
effect of electromagnetic fields on, 3:543
failure hazards of, 3:538
human factors in, 3:536–547
humanitarian use, 2:146
as a hospital problem, 6:110
in a hospital safety program, 6:117–119
“indication for use” statement for, 2:146
marketing approval for, 2:146
methods to improve, 3:539–540
microarrays as, 4:370–371
new issues involving, 3:542–544
nickel–titanium shape memory alloy, 1:8–10
postmarket rules for, 2:148
postmarket surveillance of, 6:120
potential interaction of, 3:543
premarket approval of, 2:146
premarket “510(k)” notification for, 2:145–146
product development protocol for, 2:146
recall system for, 6:118
replacement planning for, 6:118
review standard for, 2:147–148
servicing, 6:118
temperatures and sterilization time for, 6:275t
update report requirements for, 2:148
use-related hazards of, 3:538
user testing and, 3:540
utilization and management of, 6:118
work domain analysis and, 3:540
- Medical device safety officer (MDSO), 6:120
- Medical device technologies, acquisition of, 6:117–118
- Medical device technology management, 6:117
- Medical diagnosis, use of nonionizing radiation in, 5:70–71
- Medical education, computers in, 4:307–311
- Medical engineering, historical developments in, 4:312–314
- Medical engineering societies/organizations, 4:311–321
- Medical gas analyzers, 4:322–335
colorimetry, 4:323–324
displays, alarms, calibration, and controls in, 4:334
electrochemical gas sensors, 4:324–325
emerging technologies in, 4:333
gas chromatography, 4:325–327
gas monitor methods, 4:323
infrared/optical spectroscopy, 4:327–329
luminescence/fluorescence, 4:329
mass spectroscopy, 4:329–330
nuclear magnetic resonance, 4:330–331
paramagnetic, 4:331
patient safety and, 4:334
radioactive ionization, 4:331–332
Raman laser spectroscopy, 4:332
solid-state, 4:332–333
- Medical gases, 3:379
as a hospital problem, 6:112–113
- Medical image archival, 5:346–347
- Medical image data files, 5:347

- Medical image display, 5:353–356
- Medical images, data size of, 6:304t
- Medical imaging, 5:406–407
- Medical imaging equipment, patient support structures in, 6:558–559
- Medical informatics, 4:309
- Medical information, inappropriate use of, 6:119
- Medical internal radiation dose (MIRD) committee, 5:565. *See also* MIRD schema
- Medical issues, in teleradiology, 6:310
- Medical microbiology, 4:372–375
- Medical monitoring, of electroconvulsive therapy, 3:57
- Medical photography, 5:291–299. *See also* Photomicrography
- ophthalmic, 5:291–294
- scanning laser ophthalmoscope and, 5:294
- telemedicine and, 5:295–296
- Medical physicists, training requirements for, 2:166t
- Medical physics, applications of Monte Carlo method in, 5:534
- Medical physics books/reports, 4:350
- Medical physics literature, 4:335–351. *See also* Medical books/reports
- medical and radiological physics books/reports, 4:337–338
- primary journals, 4:335–336
- secondary journals, 4:336–337
- Medical professionals, fatigue issues of, 3:543
- Medical radiation, 5:253
- Medical radiation protection products, 5:501t
- Medical records, biometric systems in management of, 4:358–359
- Medical record systems
- computer-based, 4:351–361
- paper-based, 4:351–352
- Medical Subject Headings (MeSH) vocabulary, 4:310
- Medical treatment options, scale of functional deficit and, 6:179–180
- Medications
- electroconvulsive therapy and, 3:56–57
- immunosuppressive, 4:273
- Medicine
- fiber optics in, 3:301–315
- infrared imaging in, 6:346–347
- international IR imaging activities in, 6:353–354
- scanning tunneling microscopy in, 4:517–519
- thermocouples in, 6:345–346
- ultraviolet radiation in, 6:473–490
- Medipad technology, 2:503
- MedSim-Eagle Patient Simulator, 1:46
- Medtronic EDGE system electrodes, 1:147
- Medtronic InterStim neurostimulator, 1:432–436
- Medtronic MiniMed artificial pancreas, 5:228
- Medtronic tined lead percutaneous implant, 1:436–438
- MEG–EMG coherence, 1:243. *See also* Magnetoencephalography (MEG)
- Melanoma, studies of BNCT for, 1:581–582
- Melt-derived bioactive glasses, 1:287
- porous, 1:291–292
- Melting temperature, of polymers, 1:331
- Membranes, nanoporous, 2:450–451
- Memory assessment/training, virtual reality for, 6:75–76
- Memory chips, to record glucose data, 1:17
- MEMS sensors, 4:532. *See also* Microelectromechanical systems (MEMS)
- Meniscus
- biomaterial scaffolds necessary for engineering, 2:74
- cells capable of generating, 2:73
- composition and structure of, 2:65–66
- functional tissue engineering of, 2:74–75
- mechanical properties of, 2:66–71
- properties of, 2:63–80
- repair strategies of, 2:73
- tissue engineering of, 2:73–75
- Mental retardation, 2:211
- M-Entropy module, 4:560–561
- Mercury, in dental amalgams, 1:322
- Mercury lamps, ultraviolet radiation from, 6:476–477
- MESAM 4 ambulatory cardiorespiratory monitor, 6:223
- Mesenchymal stem cells (MSCs), in cartilage regeneration, 2:73
- Messenger RNA (mRNA), structure of, 4:362
- Metabolic complications, of parenteral nutrition, 5:129–130
- Metabolic heat generation, 1:189, 191
- Metabolic network, 1:225
- Metabolic therapy, systemic hyperthermia and, 4:58
- Metabolic thermatomes, 6:349
- Metabolic tissue, engineered, 3:202–203
- Metacarpophalangeal (MCP) finger joint replacements, 1:303–304
- “Metal allergy,” 1:112
- Metal disk electrodes, 1:139
- Metal electrodes, reducing edge effects in, 1:146–148
- Metal foil electrodes, 1:144
- Metal halide arc lamps, ultraviolet radiation from, 6:476–477
- Metallic biomaterials, passivity of, 1:309
- Metallic ions, in implant subjects, 1:111–112
- Metal-on-metal hip joints, 3:520–521
- Metal-on-metal prostheses, 1:317–318
- Metal oxide semiconductor field effect transistor (MOSFET), 4:99, 185–186, 190. *See also* MOSFET dosimeters
- Metal plate electrodes, 1:144
- Metal probes, for external electrostimulation, 1:143
- Metals, 1:104–105
- alloyed with diamond-like coatings, 1:319–320
- anesthesia machine, 1:35
- as biomaterials, 1:270–272
- electrode, 1:129–131
- as prosthetic restorative materials, 1:325–326
- Methyl methacrylate (MMA), 1:541–542
- Michigan Probe, 1:155, 156
- Microarray analysis, 1:223–224, 225f
- fluorophores used for, 4:369t
- Microarray experiment, washing step of, 4:364
- Microarray images, quantification of, 4:365
- Microarrays, 4:361–371
- basic principles of, 4:361–366
- data analysis for, 4:366
- DNA, 2:433
- equipment related to, 4:367–370
- fabrication of, 4:366–367
- as medical devices 370, 4:371
- scanning, 4:365
- Microarray scanners, 4:369–370
- Microbatteries, 4:429–430
- Microbial detection, piezoelectric sensors for, 5:364
- Microbial detection systems, 4:372–383
- development of, 4:375–376
- electronic nose, 4:381
- fiber-optic fluoroimmunoassay systems, 4:378
- future trends in, 4:381–382
- microchip technology, 4:380–381
- nanoparticle-based bio-barcode technology, 4:379–380
- nucleic acid-based optical technologies, 4:376–378
- Microbioreactors, 4:383–400
- components of, 4:389–392
- design principles of, 4:385–389
- examples of, 4:393–395
- mechanical and material considerations for, 4:387–388
- operation principles of, 4:386–389
- for optimizing production conditions, 4:393
- for therapeutical applications, 4:394
- for toxicological and drug testing, 4:393
- for understanding biological responses, 4:394–395
- Microbolometer technology, 6:350
- Micro capillary systems, 2:434
- Microcapsules, implantable, 4:394
- Microcard project, 1:156
- Micro Cell Culture Analogs (CCAs), 4:393
- MicroCHIP drug delivery chip, 1:425f
- Microchips, drug delivery, 1:424, 425f
- Microchip technology, 4:380–381
- Microcirculation, 1:515–516
- Microcirculation systems modeling, 5:313
- Microcolony formation, in biofilms, 1:115
- Microcomputer, in neurological monitors, 5:34
- Microcontact printing (stamping), 1:409, 411, 412f
- Microcontrollers, in temperature measurement electronics, 6:330–331
- Microdialysis, 3:397
- Microdialysis membrane dimensions, 4:403t

- Microdialysis probes, sensor attachment to, 4:411
- Microdialysis sample quantitation, separations- based methods for, 4:409–410
- Microdialysis samples
analysis of, 4:409–412
peptide and protein analysis of, 4:411–412
- Microdialysis sampling, 4:400–420. *See also* Microdialysis samples
applications of, 4:412–413
clinical applications of, 4:413
detection types related to, 4:410–411
device calibration in, 4:405–409
evaluation and future use of, 4:413–414
instrumentation components in, 4:402
principles of operation in, 4:402–405
probe insertion trauma in, 4:408–409
recovery, delivery, and localized infusion in, 4:404–405
sample volume limitations in, 4:409
simplified view of, 4:400–401
uses for, 4:401t
- Microdosimetric measurements, in neutron beam therapy, 5:55
- Microdosimetry, in the MIRD schema, 5:567
- Microdrug delivery system, water-powered, 2:504
- Microelectrode arrays, 1:156
- Microelectrodes, 1:154–158
- Microelectro discharge machining (micro-EDM), 4:527–528
- Microelectromechanical (MEMS) based technology, 1:417–418, 4:22, 4:23, 4:24, 4:27
- Microelectro-mechanical drug delivery systems, 2:440–452
- Microelectromechanical systems (MEMS), 1:509–510; 2:1; 4:333, 527. *See also* MEMS sensors
fabrication of, 4:527–528
- Microemulsions, 2:460–461
as drug delivery systems, 2:461–462
ophthalmic application of, 2:463
transdermal application of, 2:462–463
- Microfabricated drug delivery system, implantable, 2:504
- Microfabrication, 4:392
of electrodes, 1:155, 157
photolithography in, 4:425
- Microflow regulator, for drug delivery systems, 2:504
- Microfluidic channels
fabricating, 4:425
studies of vascular diseases in, 4:395
- Microfluidic modeling, 4:422–423
- Microfluidic networks (μ FNs), patterning via, 1:412–413
- Microfluidics, 4:333, 4:20–427
biomedical applications of, 4:426–427
fabrication in, 4:425–426
pumping fluids, 4:423–425
theory of, 4:420–423
- Microfluidic systems, as assisted reproductive technologies, 4:394
- Microfuel cell, 4:429–430
- Microheat engine, 4:433
- Micromachined battery, 4:430
- Micromachining, surface, 2:3–5
- Micromachining technologies, 2:3
- Microneedles, for transdermal drug delivery, 2:442–446
- Microneurography, 3:120–121
- Microparticles
for intravenous drug delivery, 2:44
nanoporous, 2:448–449
nonporous, 2:448
for oral drug delivery, 2:446–447
- Micropipette technique, 1:506–507
- Microporous implant surfaces, 1:130
- Microporous platinumized platinum electrode, 1:154
- Micropower, for medical applications, 4:428–434
- Micropower generator, 4:430–433
- Micropumps, drug delivery, 2:442
- Microreservoirs, as drug delivery systems, 2:449–450
- Microrheometers, 1:505–506
- Microscope, surgical, 4:523–526
- Microscopy. *See also* Confocal microscopy; Electron microscopy; Fluorescence microscopy; Scanning force microscopy; Scanning tunneling microscopy
confocal fluorescence, 4:493–494
fluorescence, 2:91–98; 3:344–345
fluorescence lifetime imaging, 2:95–97
immunoelectron, 4:602
laser scanning confocal, 2:92–93
lateral force, 4:508
lifetime resolved, 4:500–501
multiphoton excitation, 2:93–94
polarization, 4:501–502
spectral imaging, 2:94–95
total internal reflection, 4:494
two-photon fluorescence, 4:494
wide-field deconvolution, 4:493
- Microsensors, capacitive, 2:1–12
- Microsphere burnt out method, for porous biomaterial fabrication, 5:401
- MicroSQUID systems, 1:239
- Micro stereo lithography processes, 4:528
- Microstimulator, single-channel, 1:424–427
- Microstimulator chips, 1:426–427
- Microstream CO₂ technology, 1:473
- Micro-stream technology, 1:480
- Microsurgery, 4:523–534. *See also* Eye surgery; Fetal surgery
applications of, 4:528
catheters/guidewires/stents in, 4:532–533
electropharmaceutical anesthesia in, 3:32
haptic feedback in, 4:528–529
surgical microscope and, 4:523–526
tissue sensing in, 4:529–530
tracking systems in, 4:530
- Micro-syringe, coronary, 2:503
- Micro-total-analysis systems (μ TAS), 4:420
- Microwave ablation, 6:364–365, 371, 374, 375–376
- Microwave devices
clinical studies with, 4:37
interstitial hyperthermia and, 4:36–37
- Microwave diathermy, 3:472
- Microwave exposure limits, 2:183t
- Microwave radiometry, 6:361
- Microwaves, therapeutic applications of, 1:190
- Micro wires, for intraspinal stimulation, 3:357
- Middle cervical spine, biomechanics of, 3:555–557
- Middle cervical spine instability, role of environmental factors in, 3:561
- Mid-wave IR (MWIR) region, 6:347
- Migraine headache
biofeedback clinical outcome literature related to, 1:178–179
biofeedback procedures for, 1:176–178
- Migration, of engineered tissue, 3:191
- Miller, Neal, 1:166
- Milliamperage accuracy, in X-ray equipment, 6:567–568
- Milliampere linearity, in X-ray equipment, 6:568
- Millikan, Glen Allan, 1:469
- MIMiC binary leaf collimator, 6:397
- MIMiC system, 5:596
- Miniature reference electrodes, in ion-sensitive field-effect transistors, 4:193–194
- Minimal Erythematous Dose (MED), 6:475
determination of, 6:481
- Minimally invasive direct coronary artery bypass, 4:541–542
- Minimally invasive surgery, 4:539–542
limitations of, 4:543
outcomes of, 4:543
- Minimally invasive surgical technology, 4:535–544. *See also* Minimally invasive surgery
instrumentation for, 4:535–539
new developments in, 4:542–543
- Minimum norm method, 1:240
- MINITAB package, 6:263
- Minor electrosurgery, 3:158–159
- Minute respiratory volume, lung, 2:39
- Minute ventilation (MV), in implantable devices, 1:204–205
- MIRDOSE software, 5:565, 566
- MIRD schema, for radiopharmaceutical dosimetry, 5:566–567. *See also* Medical internal radiation dose (MIRD) committee
- Mismatch negativity (MMN), 3:237
- MIT studies, of BNCT tolerance, 1:580–581
- Mixed lubrication, 1:315
- Mixed-venous fiber optic catheters, 5:165–166
- Mixed venous oxygen content (MVO₂), 2:16
- Mixers, 4:390
- Mixing, in microbioreactors, 4:390
- M-mode, 3:1
- M-mode echocardiography, clinical uses of, 3:19–20

- Mobetron mobile intraoperative radiotherapy unit, 6:19
- Mobile gamma cameras, 5:98–99
- Mobile intraoperative radiotherapy units, 6:18–20
- Mobility, clinical assessment of, 4:544–546
- Mobility aids, 1:448, 449t; 4:544–555
 - lower extremity prosthetics, 4:552–554
 - manual wheelchairs, 4:546
 - powered assist wheelchairs, 4:546–547
 - powered wheelchairs, 4:547
 - sports and recreation devices, 4:548–550
 - vehicle control systems, 4:550–551
 - walkers and rollators, 4:547–548
- Mobility-based enhanced resolution techniques, in electrophoresis, 3:135
- Mobility environments, virtual reality, 6:76
- Modality worklist, 5:335–336
- Model for End Stage Liver Disease (MELD) score, 4:269–270
- Modeling. *See also* Mathematical models of joints, 4:212–220
 - microfluidic, 4:422–423
 - role in pharmacokinetics and pharmacodynamics, 5:275–276
- Model predictive controller (MPC), 1:495–496
- Model reference adaptive control (MRAC), 1:494–495
- Modulated drug delivery, 2:439
- Modulation, in biotelemetry systems, 1:420, 421
- Moiré fringe photography, 6:127
- Moist heat sterilization (autoclave), 6:275–276
- Molecular biology applications, of fluorescence, 3:346
- Molecular drug delivery systems, 2:452–460
- Molecular filtration, sandwich design for, 2:451–452
- Molecular imaging, 5:406–407
- Molecular probes, monoclonal antibody molecules as, 4:604
- Molecular radiation biology/bystander effects, 4:182
- Molecular weight (MW)
 - of polymers, 1:331
 - of thermoplastics, 1:332
- Molecules, in polymers, 1:331
- Molybdenum-99/technetium-99m radionuclide generator, 1:52
- Monitoring. *See also* Monitors; Neonatal monitoring
 - ambulatory, 1:12–18
 - in anesthesia, 4:555–565
 - of electroconvulsive therapy, 3:57
 - of gas systems, 3:380
 - hemodynamic, 4:565–576
 - intracranial pressure, 4:576–588
 - umbilical artery and vein, 4:588–597
 - ventilatory, 6:514–528
- Monitoring/control components, in microbioreactors, 4:391–392
- Monitors
 - anesthesia machine, 1:37, 39–40
 - EEG, 5:32–33
 - neurological, 5:32–41
- Monochromatic radiation, 6:478
- Monochromator, 3:346
- Monoclonal antibodies, 4:597–608
 - additional uses for, 4:607
 - bioterrorism and, 4:607
 - in cancer detection, follow-up, and treatment, 4:605
 - in the food industry, 4:607
 - imaging tumors with, 4:605–606
 - immune system and, 4:597–599
 - produced in plants, 4:601
 - production of, 4:599–600
 - as serological and diagnostic probes, 4:601–604
 - as therapeutic agents, 4:604–605
- Monoclonal antibody molecules (Mabs), 4:601
 - as molecular probes, 4:604
 - in nuclear medicine, 4:605–606
- Monocular calibration, 5:146
- Monocytes, 2:82
- Monomer diffusion, in steep dose gradient, 5:495
- Monophasic anodic pulses, 1:128
- Monopolar recording configuration, 3:114
- Monopolar recordings, with EEG
 - biofeedback instrumentation, 1:171
- Monovision, contact lens designs for, 2:326–327
- Monte Carlo calculations
 - in medical physics, 5:534
 - in radiation therapy treatment planning, 5:534–542
- Monte Carlo radiation dose calculation, 5:459–460
- Monte Carlo simulations, 5:534–535
 - geometry specification in, 5:534–535
 - in theoretical dosimetry, 1:613–614
- Monte Carlo techniques, for radiation dosimetry, 5:471
- Morphine, 2:51
- Morphine analgesia, for chronic pain, 3:32
- Morphological analysis
 - dual-chamber with ventricular, 1:78
 - two-channel, 1:78–79
- Morphological finite element models, patient-specific and task-dependent, 4:219–220
- Morphological pattern recognition, in arrhythmia detection, 1:73–74
- Morphologic approaches, to arrhythmia analysis, 1:76–77
- Morphology discrimination (MD), 1:76
- Morphonemic text-to-speech conversion, 1:447
- Morton, W. T. G., 1:28
- Morton Inhaler, 1:28f
- MOSFET dosimeters, 5:477, 600–601. *See also* Metal oxide semiconductor field effect transistor (MOSFET)
- Motion. *See also* Movement
 - effect on joints, 4:204
 - quantifying, 1:384–391
- Motion artifact
 - pulse oximetry and, 5:212
 - of skin potential, 1:133–134
- Motion-mode imaging, 6:461–462
- Motion sensing pad, in neonatal respiratory monitoring, 5:16–17
- Motor proteins, 5:183–185
- Motor unit action potential, 3:100–101
- Motor unit action potential train, 3:101
- Mount Sinai Chemistry Automation Project
 - floor plan for, 1:25f
 - net present value of, 1:24–27
- Movement, age-dependent, 1:396. *See also* Motion entries
- Movement-evoked fields (MEFs), biomagnetic measurements and, 1:243
- Mowat Sensor, 1:450
- MP35N cobalt alloy, 1:271
- MRI-based measurement, of intracranial compliance and pressure, 4:6–8
- MRI contrast enhancement, nanoparticles in, 5:5. *See also* Magnetic resonance imaging (MRI)
- MRI equipment, 5:78
- MR imaging, of polymer gels, 5:486–487. *See also* Magnetic resonance imaging (MRI)
- MRI thermometry, 6:359–360
- MRS brain spectra, metabolites detected in, 5:85–87t, 89t. *See also* Magnetic resonance spectroscopy (MRS); Multi voxel MRS
- MR simulator, 5:530–531
- MR spatial encoding, 4:285
 - in the Fourier domain, 4:285–286
- MRS spectra, examples of, 5:87, 88f
- Müller, Hermann, 1:197
- Multichannel magnetocardiogram (MCG) systems, 1:237
- Multichannel systems, 1:420
 - with EEG biofeedback instrumentation, 1:171
- Multicolor optical coding, nanoparticles in, 5:6
- Multidetector computed tomography, 2:233
- Multidetector translate–rotate computed tomography, 2:231
- Multielectrode silicon probe, 1:155f
- Multielement transducers, 4:65
- Multifocal ERG (mfERG), 3:152–153, 155–156
- Multifocal intraocular lenses, 4:238–239
- Multifrequency method, of whole-body bioelectric impedance measurement, 1:212
- Multifrequency tympanometry, 1:101
- Multiimage CAD, 2:302
- Multiinput and multioutput (MIMO) models, 1:495
- Multilamellar vesicles, freeze and thawed, 2:469
- Multileaf collimators (MLCs), 5:521–523, 576, 579
- Multileaf Collimator (MLC) system, 5:596
- Multimodality CAD, 2:302–303
- Multiphoton excitation microscopy (MEM), 2:93–94
- Multiple electrode mapping, 6:371
- Multiple-element sensors, in arterial tonometry, 6:403–404, 405

- Multiple EM for Motif Elicitation (MEME), 1:222
- Multiple-head gamma cameras, 5:99–100
- Multiple laminar streams
in microbio reactors, 4:390
to study subcellular biology and chemotaxis, 4:394–395
- Multiple medical devices, potential interaction of, 3:543
- Multiple model adaptive control (MMAC), 1:495–496
- Multiple row and area detectors, 2:236–237
- Multiple sequence alignments, 1:219–220
- Multiple Signal Classification (MUSIC), 1:241
- Multiplexing, in biotelemetry systems, 1:420
- Multiplier transformation, 1:395
- Multipoint stainless-plate electrode, 1:138
- Multipoint video conferencing, 5:159
- Multipotent adult progenitor cells (MAPCs), 6:382–383
- Multiscale image contrast amplification (MUSICA), 5:342
- Multislice CT, 5:529
- Multivariate methods, 6:261–262
- Multiview radiography, for scoliosis, 6:126–127
- Multi voxel MRS, 5:79
- Multiwindow spatial registration, Anger camera, 1:60
- Mu rhythm, 1:243
- Murine cardiac ventricular cells, research in, 3:145–146
- Murine ventricular action potentials, computational modeling of, 3:146–147
- Muscle(s). *See also* Myo- entries
cardiac, 2:43
contraction of, 1:385, 391–392
electrodes in or on, 3:353–355
- Muscle assessment methods, advanced, 6:65–66
- Muscle cells, in arterial walls, 1:85
- Muscle dynamics, measuring, 6:65–66
- Muscle force assessment, stimulated, 6:66–70
- Muscle force assessment system, 6:68f
- Muscle overload, 1:392, 393
- Muscle performance, components of, 6:63
- Muscle tension, biofeedback training and, 1:167–168
- Muscle testing, 6:62–70
isometric, 6:63
manual, 6:64–65
- Muscle tissue, aging and, 1:389
- Muscle tone
cooling and, 3:466
therapeutic heat and, 3:465
- Muscular electrodes, 1:149
- Muscular endurance, 1:391
- Muscular strength, 1:391
- Muscular system, integration of, 1:391
- Musculoskeletal disease, magnetic resonance imaging of, 4:296–297
- Musculoskeletal function, evaluation of, 6:85–92
- Musculoskeletal system, 1:385
- Mutations, polymerase chain reaction and, 5:382
- Myelomeningocele, 4:175–176
- Myenteric plexus, 1:68
- Myocardial electrical impedance (MEI), 1:210–211
- Myocardial infarction, 2:50–51
stroke and, 2:52
- Myocardial ischemia, total parenteral nutrition regimen for, 5:132
- Myocardial oxygen balance, 4:163–164
- Myocardial oxygen supply/demand, 4:163–164
- Myocardial perfusion studies, 1:53f
- Myoelectrical activity, gastric, 3:84–85
- Myofascial pain dysfunction (MPD) syndrome, biofeedback clinical outcome literature related to, 1:179
- Myosins, 5:183–184
- Myotomal thermatomes, 6:349
- N20m source, 1:242–243
- Nakajima, Susumu, 1:471
- NanoChip[®] system, 4:380–381
- Nanoengineered drug delivery device, 2:504
- Nanomanipulation, alternatives to optical tweezers in, 5:182
- Nanomaterials, commercial exploration of, 5:7, 8t
- Nanoparticle-based bio-barcode technology, 4:379–380
- Nanoparticles, 5:1–10
applications of, 5:4–7
as biological tags, 5:5
in biomolecule/cell separation and purification, 5:5
in cancer therapy, 5:6
DNA detection with, 2:434
in drug delivery, 5:5
fabrication of, 5:2–4
future directions for, 5:7
health issues related to, 5:7–10
in the intestinal tract, 5:9
lipid-based, 2:484–486
in the lung, 5:8–9
in manipulation of cells and biomolecules, 5:6–7
in MRI contrast enhancement, 5:5
in multicolor optical coding for biological assays, 5:6
ports of bodily entry of, 5:7–10
in protein detection, 5:7
recent developments related to, 5:6–7
self-assembled, 5:3–4
skin penetration by, 5:9–10
therapeutic applications of, 2:481–484
in tissue engineering, 5:5, 6
in tumor destruction, 5:5
- Nanoparticle surface treatment, 5:4
- Nanoparticle therapy, whole-body hyperthermia and, 4:49t
- Nanopore sequencing, 2:435
- Nanoporous membranes
biocompatibility of, 2:451
zero-order kinetics through, 2:450–451
- Nanoporous microparticles, 2:448–449
- Nanoporous silicon membranes, as drug delivery systems, 2:450
- Nanoscale, DNA sequencing at, 2:434–435
- Nanosecond, 3:346
- Nanosensors, 1:427
- Nanostructured lipid carrier (NLC), 2:485–486
- Nanotechnology
for biotelemetry, 1:426–427
in gas analysis, 4:333
for microelectrodes, 1:157–158
- Nanotubes, 1:298, 299f, 305
- Narcotrend anesthesia monitoring system, 4:560
- Nasal drug delivery, cyclodextrins in, 2:454
- Nasal thermistor sensor, 6:336
- Nasopharynx temperature monitoring, 6:317
- National Academy/Board on Radiation Effects Research (BRER), 2:154
- National Bureau of Standards (NBS), on biofeedback instrumentation, 1:168
- National Committee on Clinical Laboratory Standards (NCCLS), automated system guidelines of, 1:24
- National Council on Radiation Protection and Measurement (NCRP), 2:154
- National Electrical Manufacturers Association (NEMA) standards, 6:603–604
- National Eye Institute, 1:444
- National Fire Protection Association (NFPA), gas system standards, 3:381
- National Institute for Occupational Safety and Health (NIOSH), 1:39
- National Institute of Biomedical Imaging and Bioengineering (NIBIB), 4:315
- National Institutes of Health (NIH), 1:267, 403
- National Institutes of Health Bioengineering Consortium (BECON), 4:315
- Native digital cross-sectional modalities, acquisition of, 5:336
- Natural absorbable polymers, as biomaterials, 1:107
- Natural language processing (NLP) techniques, 1:227
- Naturally derived biomaterials, in tissue engineering, 6:383–384
- Naturally derived collagen matrices (NDCM), 6:196–198
- Natural polymers, 5:389–390
in engineered tissue, 3:194–195
- Natural polymer scaffolds, 1:367–371
- Navigational aids, for the visually impaired, 1:451–453
- Near-field scanning optical microscopy (NSOM), 4:435–448
apertured, 4:439–441
apertureless, 4:441–443
applications of, 4:447f
evaluation of, 4:445–448
light sources for, 4:438
operation of, 4:443
theoretical principles of, 4:436

- Near-field spectroscopy, 4:443–445
light sources for, 4:438
- Near-infrared fibers, 3:303–304
- Near infrared (NIR) region, 6:347
- Near-infrared spectroscopy, in erectile assessment, 6:157–158
- Neck cancer, treatment planning for, 5:538
- Needle electrode field, in electrosurgery, 3:174–175
- Needle electrodes, EMG, 3:103–104
- Needleless injection, 2:503
- Needleman–Wunsch (N–W) algorithm, 1:218–219
- Needs assessment, for augmentative and alternative communication systems, 2:207
- Negative punishment, 1:167
- Negative reinforcement, 1:167
- Negative temperature coefficients (NTC), 6:321. *See also* NTC thermistors
- Nellcor CapnoProbe™ sublingual capnometer, 1:482f
- Nellcor Easy Cap II Pedi-Cap colorimetric CO₂ detector, 1:482f
- Nellcor Microstream® ETCO₂ breath sampling unit, 1:480
- Nellcor monitors, 1:473
- Nellcor OxiMax® NBP-75 handheld capnograph/pulse oximeter, 1:481f
- Neonatal bilirubin monitoring, 5:28–29
- Neonatal blood gas/biochemical measurement, 4:590–591
- Neonatal blood gas/biochemical sensors, continuous intravascular, 4:591–592
- Neonatal blood pressure measurement, 5:26–27
- Neonatal care unit, blood-chemistry parameters monitored in, 4:592t. *See also* Neonatal intensive care unit (NICU); Newborn entries
- Neonatal hemodynamic monitoring, 4:593–594
- Neonatal intensive care unit (NICU), monitoring in, 4:588–589
- Neonatal intracranial hemorrhage monitoring, 5:28
- Neonatal intracranial pressure, monitoring, 5:27–28
- Neonatal life support systems, monitoring, 5:29–30
- Neonatal monitoring, 5:11–32
blood gas measurement in, 5:22–25
cardiac, 5:13
diagnostic recordings and, 5:30
invasive, 4:595
respiratory, 5:13–18
temperature, 5:25–26
transthoracic impedance combined with cardiac monitors, 5:22
- Neonatal pressure measurement, 5:26–28
- Neonatal respiration monitoring, by transthoracic electrical impedance, 5:18–22
- Neonates, high frequency ventilation applications in, 3:505–508
- Neoplasms
electron microscopic diagnosis of, 4:484
esophageal, 3:390
lower GI, 3:392
- Nernst equation, 1:122–123
- Nerve-axon-related hyperemic response, 2:379–380
- Nerve electrodes, implanted, 3:355–357. *See also* Neural electrodes
- Nerve growth factor (NGF), 1:375–377
- Nerve recordings, 3:110–111
- Nerve tissue, engineered, 3:205
- Nervous system, 1:385–386. *See also* Neural entries; Neuro- entries
- Net magnetization, creating, 4:283
- Net Present Value (NPV) calculations, for Mount Sinai total laboratory automation, 1:24–26
- Net Present Value profile, for Mount Sinai total laboratory automation, 1:27
- Networked-Attached Storage (NAS), 5:530
- Networking hardware, 5:350–351
- Networking software, 5:352–353
- Network security, 5:353
- Neural electrodes, 1:149. *See also* Nerve electrodes
encircling, 1:157
- Neural net (NN) systems, 2:319–320
- Neural network approaches, to arrhythmia analysis, 1:76–77
- Neural-network based blood pressure controller, 1:497
- Neural prosthetic systems, experimental, 3:128–129
- Neural signals, 3:111–112
- Neurodermatomal thermatomes, 6:349
- Neuro-electronic interface, 3:110–119
- Neuroendoscopy, 3:185
- Neurofeedback, biofeedback clinical outcome literature related to, 1:181–182
- Neurological effects
of cooling, 3:466
of therapeutic heat, 3:465
- Neurological monitors, 5:32–41
classification of, 5:32–33
common specifications of, 5:39
electrodes in, 5:33–34
main components of, 5:33–34
types of, 5:35–39
- Neurologic procedures, electrosurgery in, 3:159–160
- Neurology, high intensity focus ultrasound in, 4:79
- Neuromagnetic fields, neural origin of, 1:238–239
- Neuromagnetism, 1:242
- Neuromuscular blocking drug, 1:29
- Neuromuscular control, orthotics and, 6:80
- Neuromuscular reeducation applications, biofeedback and, 1:170
- Neurons
axon guidance in, 1:414
pyramidal, 3:63–64
- Neuropeptides, quantitation of, 4:412
- Neuroprosthetic applications, use of
peripheral nerve signals in, 3:125–129
- Neuroprosthetic systems, preclinical, 3:126–127
- Neuroscience, microdialysis sampling in, 4:412
- NEUROS project, 1:157
- Neurostimulation
history of, 1:429–430
of bladder dysfunction, 1:429–443
- Neurostimulator, Medtronic InterStim, 1:432–436
- Neurostimulatory techniques, clinical, 3:26–32
- Neurosurgery, surgical microscope in, 4:524
- Neurosurgical electroanalgesia methods, 3:27
- Neutral thermal environment, 4:148–150
- Neutron activation, theory of, 5:41–43
- Neutron activation analysis (NAA), 5:41–50
applications of, 5:48–49
chemical recovery in, 5:46
equipment and methodology in, 5:43–49
evaluation and quality assurance for, 5:47–48
of gamma spectrum, 5:46–47
irradiation in, 5:44
measurement in, 5:44–45
preparation for, 5:43–44
radiochemical separation in, 5:45–46
sample preparation in, 5:44
sampling in, 5:43
- Neutron beam therapy, 5:50–64. *See also* Neutron radiotherapy
beam characteristics in, 5:55–57
clinical results review for, 5:53–54
facilities for, 5:61–62
origins of, 5:50–51
radiobiological rationale for, 5:51–53
- Neutron dose, 5:55
- Neutron dose distributions, measurement of, 5:493
- Neutron production, 5:57–61
- Neutron radiotherapy, fast and slow, 6:6–8
- Neutron sources
for boron neutron capture therapy, 1:577–579
for medical use, 5:54–55
for radiation therapy, 5:54–57
- Neutron spectra, 5:55
- Neutron therapy centers, list of, 5:56t
- Neutron therapy facility, key requirements for, 5:54t
- Neutron yield, 5:55
- Neutrophils, 2:81
- Newborn heat transfer, infant incubators and, 4:148. *See also* Neonatal entries
- Newborn intensive care units (NICUs), 3:499
high frequency ventilation in, 3:509–511
- New drug development, dosimetry in, 5:571–572
- Newsgroups, in office automation systems, 5:155
- Newsletters, Nuclear Regulatory Commission, 2:171
- Newtonian fluids, 1:500–501, 504–505
- Nickel, from implants, 1:112, 313
in dental prosthetics, 1:325

- Nickel–titanium (Ni–Ti) shape memory alloys, 1:2, 3–5. *See also* Nitinol entries
 medical devices using, 1:8–10
 thermomechanical properties of, 1:4–5
- Niosomal drug carriers, 2:473–477
- Niosome preparation
 components used in, 2:474
 methods of, 2:474–475
- Niosomes
 in complex systems, 2:475
 therapeutic applications of, 2:475–477
 toxicological aspects of, 2:475
- 55-Nitinol, physical and mechanical properties of, 1:3t
- Nitinol (Nickel–Titanium Naval Ordnance Laboratory) alloy, 1:2, 313
- Nitinol stents, 1:272
- Nitric oxide therapy, 3:508
- Nitrogen meter, 5:436
- Nitrogen washout, 5:438
- Nitroglycerine, 2:51
- NM imaging, 5:108–114. *See also* Nuclear medicine (NM)
- Noble metal electrodes, 1:126, 129–131
- Nociceptors, 6:437–438
 sensitization of, 6:439–440
- Noise
 cancellation of, 1:234–236
 at electrode interface, 1:130
 in exercise stress testing, 3:249–250
 in the scanning electron microscope, 4:483–484
- Noise levels, computer tomography, 6:575–576
- Noise reduction, using higher order gradients, 1:235–236
- NOMOS Peacock System, 6:397–398
- Nonbonded pairs, in protein structure prediction, 1:220–221
- Nonbyproduct radiation, materials codes and regulations for, 2:171–177
- Nonbyproduct radionuclides, regulations related to, 2:175–177
- Non-catheter/noninvasive angiography, 2:425–426
- Nonclinical laboratory studies, codes and regulations related to, 2:143
- Nonclinical studies, codes and regulations for, 2:147
- Noncontacting motion sensors, in neonatal respiratory monitoring, 5:16
- Nondispersible infrared gas analyzers, 5:436
- Nonelectrophoretic-based DNA sequencing methods, 2:433
- Non-endoscopic procedures, 4:537–538
- Nonfluid-resistance pneumotachometers, 5:370–371
- Nonfocused radiation fields, 4:63–64
- Nonfusion treatment, for spine stabilization, 3:585–588
- Nonhomogeneous obstructive lung disease, 3:507
- Nonimaging peripheral vascular noninvasive measurements, 5:235–245
- Noninvasive blood gas measurements, 1:465
- Noninvasive blood glucose sensors, 1:17
- Noninvasive blood pressure measurement, 1:486–489
- Noninvasive cerebral oximetry, optical sensors in, 5:167–168
- Noninvasive electric bone treatment, 1:563–564
- Noninvasive electromagnetic devices, 1:564–568
- Noninvasive measurements
 of arterial elasticity, 1:86–87
 of cardiac output, 1:200–201
- Noninvasive optical glucose sensing, 3:396–397
- Non-ionizing radiation (NIR)
 biological effects of, 5:64–72
 exposure limits for, 5:67t
 extremely low frequency radiation, 5:68–69
 infrared radiation, 5:66–67
 protection against, 5:64, 69–70
 radio frequency radiation, 5:67–68
 regulations for, 2:177–184
 safety standards organizations publishing, 2:159t
 serious injury from, 5:69
 spectrum of, 5:65t
 ultraviolet radiation, 5:65–66
 U.S. divisions regulating, 2:157–158
 use in medical diagnosis and therapy, 5:70–71
 visible radiation, 5:66
- Nonlaser heating devices, 3:468–470
- Nonlinear heat transfer modeling, 6:349
- Nonluminous emitters, 3:469
- Nonmedical devices, as a hospital problem, 6:113
- Nonmedical radiation-producing equipment, regulations related to, 2:174–175
- Non-Newtonian fluids, 1:500–501
- Nonobstructive hydrocephalus, 4:9
- Nonparametric histogram analysis, 2:401
- Nonparametric methods, in EEG analysis, 3:69–71
- Nonparametric testing, 6:257–259
- Nonpathogens, 1:113, 114
- Nonporous microparticles, 2:448
- Non-uniform rational B-splines (NURBS), 5:121–122
- Nonunit density tissues, gel simulation of, 5:494
- Nonvariceal upper GI bleeds, 3:388–390
- Nonzero-order release profile, 2:439
- Normal distribution, 6:245–246
- Normalized area of difference (NAD), 1:74–75
- Normalized separation-based neurological monitor, 5:37–38
- Normal pressure hydrocephalus (NPH), 4:3
- Normal tissue complication probability (NTCP), 5:547
- Northern Ontario Remote Telecommunication Health (NORTH) network, 1:47
- Nose and throat diseases, cryosurgical treatment of, 2:374
- Nottingham Physiology Simulator, 5:303
- Novac7 mobile intraoperative radiotherapy unit, 6:19–20
- Nova CCX critical care analyzer, 1:21
- Novacor LVAS (left ventricular assist device), 3:453
- Novel recombinant molecules, creation by polymerase chain reaction, 5:383–384
- Novoste BetaCath IVB system, 1:604–605
- NSOM head, 4:438. *See also* Near-field scanning optical microscopy (NSOM)
- NTC thermistors, 6:333. *See also* Negative temperature coefficients (NTC)
 clinical applications of, 6:333–338
- Nuclear Chicago scintillation camera, 1:52
- Nuclear diagnostics, radiopharmaceuticals applied in, 5:567t
- Nuclear magnetic resonance (NMR), 4:283, 330–331
- Nuclear magnetic resonance (NMR) spectroscopy, 5:72–90. *See also* Magnetic resonance spectroscopy (MRS); MRS entries
 applications of, 5:83–89
 equipment and experiments in, 5:76–83
 theory behind, 5:74–78
- Nuclear medicine (NM)
 computers in, 5:106–124
 education related to, 5:122
 monoclonal antibody molecules in, 4:605–606
 phantom materials in, 5:267
 sodium iodide crystal in, 1:53–54
- Nuclear medicine books/reports, 4:344–345
- Nuclear medicine detectors
 one-dimensional, 5:94–95
 three-dimensional, 5:100–103
 two-dimensional, 5:95–100
- Nuclear medicine instrumentation, 5:90–106
 animal imaging devices, 5:104–106
 hybrid, 5:103–104
 in one-dimensional nuclear medicine detectors, 5:94–95
 in radiation detection systematics, 5:93–94
 in radionuclide production, 5:92
- Nuclear medicine labeling, 5:91
- Nuclear perfusion imaging, 3:254–255
- Nuclear pharmacists, training requirements for, 2:166t
- Nuclear reactors, as neutron sources for BNCT, 1:577
- Nuclear Regulatory Commission (NRC), 2:154–155
 communications from, 2:171
 regulations, 2:160–166
 website, 2:171
- Nuclear techniques, in exercise stress testing, 3:254–255

- Nuclear ventricular function assessment, 3:254
- Nucleic acid amplification procedures, 5:385
- Nucleic acid-based optical technologies, 4:376–378
- Nucleic acid sequence-based amplification (NASBA), 4:377–378
- Nucleic acid species, hybridization of, 4:363–364
- Nucleus implants, spinal, 6:237–238
- Nutrient solutions
essential components of, 5:124–125
formulating, 5:125–126
three-in-one system for, 5:127–128
- Nutrition, parenteral, 5:124–134
- Nutrition home health care devices, 3:533
- Nylons, 1:337–338
- Nystagmus
congenital versus acquired, 5:140
spontaneous, 5:140
- Nystagmus scanpaths, 5:138
- OASIS wound matrix, as a skin substitute, 6:174–175
- Object contrast, in computed tomography, 2:252–253
- Observer studies, 2:300–301
- Obstetric electroanalgesia, 3:31–32
- Obstructive sleep apnea, 6:212
- Obstructive sleep apnea/hypopnea syndrome (OSAHS), 2:329–332
comfort/compliance issues in, 2:335–336
indications for use of continuous positive airway pressure in, 2:334–335
- Occipital-atlantoaxial complex, biomechanics of, 3:554–555
- Occupational Safety and Health Administration (OSHA), 1:456; 2:156.
See also OSHA regulatory standards
- Ocular fundus reflectometry, 5:135–136
clinical applications of, 5:136
- Ocular motility recording, 5:137–149.
See also Eye movements
techniques in, 5:146–147
- Ocular motor recording systems, 5:140–146
- Ocular structures, effect of microemulsions on, 2:463t
- Oddball paradigm, 3:236–237
- Oersted, Hans, 1:429
- Off-hour reading teleradiology model, 6:308
- Office automation systems, 5:149–160
defined, 5:151
digital communication systems in, 5:155–156
groupware systems in, 5:156–158
historical perspective on, 5:150
organizational information systems and, 5:151–152
productivity tools in, 5:152–155
teleconferencing in, 5:158–159
- Offset connectors, for electrodes, 1:140
- Offset instability standards, 1:158–159, 160
- Ohm's law, 1:466
- OLV-5100 ear pulse oximeter, 1:471
- Omitted evoked potentials (OEPs), 3:237
- On-chip image processing, 6:350
- Oncology
automated cytology in, 2:404–405
radiation dosimetry for, 5:465–481
- 1D electrode arrays, 1:155
- One-dimensional nuclear medicine detectors, 5:94–95
- OneDose patient dosimetry system, 5:600–601
- One-half standard deviation rule, 5:451
- One-sample t-test, 6:251–253
- One-way ANOVA
for matched samples, 6:254–256
for unmatched sample, 6:254
- Onset, in rate-based arrhythmia analysis, 1:72
- On-X carbon, 1:302
- Open-cell foam layers, for electrodes, 1:140
- Open chest defibrillation, 2:37
- Open reading frames (ORFs), 1:222
- Open scavenger systems, anesthesia machine, 1:39
- Operant conditioning, 1:166–167
- Operating theater, anesthesia monitoring outside, 4:563–564
- Operative contamination, prevention of, 1:117–118
- Ophthalmic drug delivery, cyclodextrins in, 2:454–455
- Ophthalmic microemulsion application, 2:463
- Ophthalmic photography, 5:291–294
- Ophthalmoscope, scanning laser, 5:294
- Opportunistic pathogens, 1:113–114
- Optacon, 6:294
- Optelec Traveller, 1:444–445
- Optical aberrations, in intraocular lenses, 4:237
- Optical absorption WBC counting technique, 2:412
- Optical beam deflection, in the atomic force microscope, 4:506
- Optical biosensors, in drug discovery, 5:173
- Optical cell measurements, 2:392–396
- Optical coherence tomography, 3:310–311
- Optical components, of fluorescence microscopes, 4:491–492
- Optical computed tomography (OCT), 5:487
- Optical density, in screen-film systems, 6:142
- Optical elements, in optical sensors, 5:164
- Optical eye movement measurement techniques, 3:267
- Optical fiber(s), 5:162. *See also* Fiber optics
construction of, 3:303
light delivery with, 3:179
transmission of images through, 3:307
- “Optical fiber” devices, 3:179–180
- Optical filter, 3:346
- Optical glucose sensing, non-invasive, 3:396–397
- Optical glucose sensors, 5:164–165
implantable, 3:397
- Optically stimulated luminescence (OSL), 5:516
- Optical/magneto-optical disk (OD/MOD), 5:348
- Optical scanning, of polymer gels, 5:487–488
- Optical sensor oxygen analyzers, 5:206–207
- Optical sensors, 3:267–268; 5:160–175
advantages and disadvantages of, 5:163
general principles of, 5:161–163
instrumentation related to, 5:163–164
in vitro diagnostic applications of, 5:172–173
in vivo applications of, 5:164–172
- Optical signal acquisition system, in near-field scanning optical microscopy, 4:437
- Optical spectroscopy, in cellular parameter measurement, 2:393–396
- Optical strain gages, 6:283
- Optical techniques
for characterizing surfaces, 1:352
implementations of, 3:276–283
- Optical technologies, nucleic acid-based, 4:376–378
- Optical tweezers, 5:175–187
alternatives to, 5:182
assays using, 5:185
calibration of, 5:179–182
experimental concerns related to, 5:183
history of, 5:176
motor proteins and, 5:183–185
nonstandard trapping by, 5:185
research related to, 5:182–185
systems of, 5:177–178
thermal force based calibration methods for, 5:181–182
trapping theory and, 5:176–177
- Optic nerve approach, to developing visual prostheses, 6:535
- Optimal point doses, 6:43–44
- Optimization method, joint distribution problem and, 4:216–217
- Optimization tools, in radiotherapy treatment planning optimization, 6:42–43
- Optimizing performance, 1:387–388
- Optode (photochemical-optical) technology, 1:474–475
- Optoisolators, 3:117
- Optokinetic nystagmus (OKN) response, 5:140
- Optokinetic response (OKR), 5:140
- Oral cancer, cryosurgical treatment of, 2:373–374
- Oral cavity temperature monitoring, 6:317
- Oral diseases, cryosurgical treatment of, 2:373
- Oral drug delivery
cyclodextrins in, 2:456–459
microparticles for, 2:446–447
- Orcel, as a skin substitute, 6:177
- Organelle probes, in confocal microscopy, 4:469–470
- Organ function loss, methods to treat, 6:181–183

- Organizational information systems, 5:151–152
- Organizations, medical engineering, 4:311–321
- Organ motion, EPID use and, 4:96
- Organs
 elemental compositions of, 5:254t
 local heating of, 1:190
 tissue layers in, 6:180–181
- Organs at risk (OAR), 6:33
- Organ synthesis, induced, 6:183
- Organ weights, Wistar rat and human, 5:572t
- Orthodontic arch wires, nickel–titanium shape memory alloy, 1:8
- Orthokeratology, contact lenses for, 2:327
- Orthopaedic polymers, common properties found in, 1:257t
- Orthopedic applications
 of nickel–titanium shape memory alloys, 1:9–10
 of porous biomaterials, 5:402
- Orthopedic braces, 6:231
- Orthopedic devices, 5:187–192. *See also* Orthopedics
 biocompatibility issues in, 5:191–192
 biomaterial surfaces of, 1:342–343
 design issues in, 5:190–191
 factors influencing, 5:188
 materials issues in, 5:188–190
- Orthopedic implants, 1:255–256
 history of, 1:255
 resorbable, 1:255–256
- Orthopedic materials, material properties of, 1:304t
- Orthopedic prostheses
 materials in, 1:317–320
 titanium alloys in, 1:312
 wear of, 1:315, 316
- Orthopedics
 cemented fixation in, 5:194
 osseointegration in, 5:194–19S
 prosthesis fixation for, 5:192–198
 resorbable implants in, 1:255–256
 revolution in, 1:289–290
- Orthostatic hypotension, 1:16
- Orthotic devices, 6:80–85. *See also* Orthotics
 conventional, 6:84–85
 fabrication techniques for, 6:82f
 lower limb, 6:88t
 materials used in, 6:84
 outcomes for, 6:92
 performance criteria for, 6:80–84
 prefabricated and custom-fabricated components in, 6:83f
 spinal, 6:89t
 upper limb, 6:89t
- Orthotics, 6:80–93
 future of, 6:92
 musculoskeletal evaluation and, 6:85–92
 uses of, 6:80
- Orthovoltage units, for radiation therapy, 6:582
- Orthovoltage X-ray units, requirements for, 2:175t
- Osborne pneumotachometer, 5:370
- Oscillation networks, for audiometers, 1:92
- Oscillators, fluidic, 5:371
- Oscillatory techniques, 1:486–487
- Oscillometry
 blood pressure measurement via, 1:488
 cuff, 1:14
 volumetric, 1:14
- OSHA regulatory standards, 2:171. *See also* Occupational Safety and Health Administration (OSHA)
- Osmetech Microbial Analyzer, 4:381
- Osmotic pumps, for drug delivery, 2:441–442
- Osseointegration, 1:109, 328, 355
 in orthopedics, 5:194–195
- Osteoarthritis (OA), 2:70, 71
 facet joint, 6:234
- Osteochondral injuries, 2:72
- Osteoconductivity, 1:289
- Osteogenesis. *See also* Bone entries
 direct current, 1:560–561
 electromagnetic, 1:561–562
- Osteoligamentous cadaver models, 3:573
- Osteolysis, 1:314, 316
- Osteoporosis, 1:540
 bioceramics and, 1:285
- Osteoproduction, 1:289
- Osteoradionecrosis, hyperbaric medicine and, 4:26
- Otis, Arthur, 5:430
- Otoacoustic emissions (OAEs), 1:101–102
 audiometric threshold prediction/estimation and, 1:102
 clinical applications of, 1:102
- Otolaryngology, 3:185–186. *See also* Ear entries
- Outcome prediction, patient modeling and, 1:48
- Outer Helmholtz plane (OHP), 1:123
- Output formats, in augmentative and alternative communication systems, 2:206–207
- Overactive bladder, pudendal nerve stimulation for, 1:441
- “Overdrive pacing,” 2:47
- Oxidation, 1:122
 of biomaterials, 1:308–309
- Oxide ceramics, as biomaterials, 1:272–273
- OxiMax pulse oximetry, 1:472–473
- Oximeters, 1:469. *See also* Pulse oximeters
 blood, 6:523
- Oximet MET-1471 pulse oximeter, 1:472
- Oximetry, 1:469–471. *See also* Pulse oximetry
 ear, 1:471
 intrapartum fetal pulse, 1:475–476
 optical sensors in, 5:165
 pulmonary artery, 5:214–215
 transcutaneous, 4:20
 transmission versus reflection, 1:470
 versus cooximetry, 1:470–471
- Oxyapatite, 1:287. *See also* Apatite entries; Hydroxyapatite entries
- Oxycardiorespirogram, neonatal, 5:30
- Oxygen. *See also* PO₂ electrode
 CPR and, 2:39–40
 in anesthesia delivery, 1:31
 in arterial blood, 5:210–212, 212–213
 in blood, 1:465
 in continuous positive airway pressure, 2:335
 in tissue, 5:213–214
 in venous blood, 5:214–215
- Oxygen alarm, anesthesia machine, 1:36–37
- Oxygen analysis, capacitive coulometry, 5:204
- Oxygen analyzers, 5:198–209, 436
 electrochemical, 5:202–206
 galvanic, 5:202–203
 history and relevance of, 5:198–199
 magnetodynamic (dumbbell or autobalance), 5:201
 magnetopneumatic (differential pressure), 5:201–202
 optical sensor, 5:206–207
 paramagnetic, 5:199–202
 polarographic, 5:203–204
 thermomagnetic (magnetic wind), 5:200–201
- Oxygenation
 of arterial blood, 6:518–519
 CPR and, 2:40
- Oxygen balance, myocardial, 4:163–164
- Oxygen carrying capacity (ctO₂), 2:14
- Oxygen-carrying colloids, 1:515–516
- Oxygen consumption (VO₂), 2:13, 14, 15–16
 direct measurement of, 2:15–16
- Oxygen flush, anesthesia machine, 1:37
- Oxygen–hemoglobin dissociation curve, 2:41
- Oxygen measurement
 in blood, 6:522–524
 in gas, 6:524–525
- Oxygen monitoring, 5:209–216. *See also* Oxygen transport
 continuous intraarterial pO₂ measurement, 5:212–213
 pulmonary artery oximetry, 5:214–215
 pulse oximetry and, 5:210–212
- Oxygen polarography, 1:466–467
- Oxygen pump, in solid electrolyte cell oxygen analyzers, 5:205
- Oxygen Ratio Monitor Controller (ORMC), 1:37
- Oxygen saturation, monitoring, 6:523
- Oxygen sensitivity, of polymer gels, 5:494
- Oxygen sensors, gaseous, 5:207–208
- Oxygen tension, 6:523
 expected, 6:105
- Oxygen therapy home health care devices, 3:534–535
- Oxygen transport
 cardiopulmonary resuscitation and, 2:40–41
 in blood gas physiology, 1:467–468
 in the human body, 5:209–210
- Oxygen uptake, calculation of, 5:433–434
- Oxyhemoglobin (HbO₂), 1:467; 2:40
- Ozone hazards, 6:488
- Pacemaker codes, international, 5:218t
- Pacemaker programming, 1:202–203

- Pacemakers, 1:151–152; 5:217–224
 clinical implantation of, 5:218–219
 clinical use of, 5:217
 components of, 5:220
 dual-chamber, 1:77
 features of, 5:219–220
 first, 1:150–151
 future of, 5:224
 lead features in, 5:222–223
 market for, 5:219
 physiological function of, 5:218
 problems in, 5:223
 pulse generators and, 5:220–222
 special devices for, 5:223–224
 stimulation thresholds in, 5:223
 temporary external, 5:218
- Pacing electrode, 1:144f
- Pacing equipment, early, 1:143, 144f
- PACS-interfaced cross-sectional modalities, 5:336t
- Paddle electrodes, 1:143
- PageWriter Touch electrocardiograph, 3:42–47
 ECG data analysis program in, 3:46–47
 ECG data flow and storage in, 3:45–46
 system description of, 3:42–45
 technical specifications for, 3:51–53
- Pain
 anesthesia in controlling, 1:44
 burn depth and, 6:171–172
 defined, 6:437
 morphine analgesia for, 3:32
 physiology of, 6:437–439
 psychological aspects of, 6:440
 theories of, 6:440–441
- Pain relief
 cooling and, 3:466
 in systemic hyperthermia, 4:58
 therapeutic heat and, 3:465
- Paint and draw programs, in office automation systems, 5:154–155
- Pain thresholds, 6:296
 during ramp inflation, 1:65
- Pair bonds, in protein structure prediction, 1:220
- Paired samples
 sign test to compare, 6:250, 258
 Wilcoxon test for, 6:258
- Paired *t*-test, to compare paired data samples, 6:253
- Paired/unpaired samples, 6:247–248
- Pair production, 6:596–597
- Pair-wise sequence alignment, 1:218–219
- Palacos R bone cement, 1:546
- Palliation, in systemic hyperthermia, 4:58
- Pancreas, effects of parenteral nutrition on, 5:131. *See also* Artificial pancreas
- Pancreatic tissue, engineered, 3:202
- Paper-based medical record systems, 4:351–352
- Paracelsus, 2:35
- PARAD+ algorithm, 1:77–78
- Paraffin wax baths, 3:467–468
- Parallel capacitance (C_{SP}), of skin, 1:132, 134
- Parallel-column model, bioimpedance and, 1:209–210
- Parallel-column thoracic cavity model, 1:200
- Parallel plate capacitor principle, 2:1
- Parallel plate flow channel, 1:508–509
- Parallel resistance (R_{SP}), of skin, 1:132–133, 134–135
- Paralysis, orthotics and, 6:80
- Paramagnetic analyzers, 6:524
- Paramagnetic gas sensors, 4:331
- Paramagnetic oxygen analyzers, 1:43f; 5:199–202
- Parameter extraction, EEG, 3:76–77
- Parametrical hypothesis testing, on continuous variables, 6:251
- Parametric histogram analysis, 2:401–402
- Parametric methods, in EEG analysis, 3:71–72
- Parasitic capacitances, 2:7
- Parenchyma, 6:101
- Parenteral drug administration, cyclodextrins in, 2:459
- Parenteral nutrition, 5:124–134
 comparing methods of, 5:131–132
 complications of, 5:128–130
 development of, 5:124
 essential components of, 5:124–125
 formulating, 5:125–126
 glucose system and, 5:126–127
 home, 5:133
 hyperlipidemia and, 5:132–133
 indications for, 5:131
 lipid system and, 5:127
 non-nutritional effects of, 5:130–131
 three-in-one system and, 5:127–128
 trace element and vitamin requirements in, 5:125t
- Partial gas pressures, 6:104
- Partial pressure (*P*), 1:465
- Partial volume artifact, in computed tomography, 2:255
- Particle accelerators, regulations related to, 2:175
- Particle dose distributions, measurement of, 5:493
- Particle drug carriers, 2:480–486
 preparation methods and formulative aspects of, 2:480–481
 therapeutic applications of, 2:481–484
- Particle fluence, 5:465
- Particle image velocimetry (PIV), 3:335–340
- Particle therapy, 6:34
- Particle tracking velocimetry (PTV), 3:335, 339
- Particle transport simulation, 5:535
- Particulate debris, from implants, 1:111
- Passivating films, 1:311
- Passive fixation devices, for electrodes, 1:152
- Passivity, of metallic biomaterials, 1:309, 310–311
- Pathogenic organisms, characteristics of, 4:373–374t
- Pathogens, classification of, 1:113–114
- Patient anatomy dependent perturbations, 1:611–613
- Patient cable, in neurological monitors, 5:34
- Patient-controlled analgesia (PCA), 1:44
- Patient-controlled analgesia pump drug infusion systems, 2:500
- Patient data acquisition, using three-dimensional CT, 2:268
- Patient dose, from mammographic X-ray equipment, 6:573
- Patient dosimetry, quality control of, 6:577–578
- Patient event button, 1:13
- Patient marking lasers, 5:530
- Patient modeling, outcome prediction and, 1:48
- Patient motion, in computed tomography, 2:255
- Patient participation, increased, 3:542–543
- Patient positioning
 for EGG, 3:88
 improvement of, 4:92–96
- Patient preparation, for automated analytical methods, 1:19
- Patient programmer, 1:435–436
- Patient records, electronic, 4:311
- Patient restraint/repositioning devices, 5:587–594
- Patients
 EGG in, 3:92–93
 as a hospital problem, 6:114
- Patient safety
 enhancing, 6:119–120
 medical gas sensors and, 4:334
 ultraviolet radiation and, 6:488
- Patient safety movement, 6:109
- Patient's Bill of Rights, 1:456
- Patient simulators, 1:45–46
- Patient-specific morphological finite element models, 4:219–220
- Patient State Index (PSI), 4:560
- Patient table top indexing, 6:577
- Pattern electroretinogram (PERG), 3:150, 152
 clinical applications of, 3:155
 visual evoked potential and, 3:155
- Patterning methods, in biosurface engineering, 1:409–414
- Pattern recognition
 in automated cytology, 2:402–403
 in neonatal respiration monitoring, 5:21–22
- Pavlovian conditioning, 1:166
- Payback period, for Mount Sinai total laboratory automation, 1:27
- p-Be (proton-beryllium) reaction, in neutron production, 5:59–60
- p*CO₂ electrode, 1:467, 468f; 6:525
- p*CO₂ sensors, optical, 5:170
- ¹⁰³Pd (palladium), physical characteristics of, 5:419–420
- PDMS molding, 1:413
- Peacock System, 6:397–398
- Peak detector, in neonatal respiration monitoring, 5:21
- Peak flow meters, 5:439
- Peak kilovoltage calibration, in X-ray equipment, 6:568–569

- Peak velocity, in eye movements, 5:138
- Pediatric cardiopulmonary resuscitation (CPR), 2:57–58
- Pediatric emergency simulator (PediaSim-ECS), 1:46
- Pediatric patients, EGG in, 3:94
- Pedicle screws, 6:235
- PEG Hemoglobin, clinical trial of, 1:520
- Pelvic floor applications, biofeedback and, 1:170
- Pelvic floor stimulation, 1:430
- Pelvic nerve stimulation, 1:431
- Penile plethysmograph, 6:159–160
- Penile prosthetics, for erectile dysfunction, 6:158–159
- Penile tumescence measurement, 6:156–157
- Pennes model, of bioheat transfer, 1:189
- Penumbra, geometric, 2:131
- Penumbra effect, pulse oximetry and, 5:212
- People, as a hospital problem, 6:113–114
- Peptic ulcer disease, 3:388–390
- Peptide and protein immunoassay, for microdialysis samples, 4:411–412
- Percentage depth dose (PDD), 2:130; 6:586–589
- Percutaneous coronary interventions, evaluation of, 3:258
- Percutaneous electrical nerve stimulation (PENS), 3:27
- Percutaneous extension hardware, 1:434
- Percutaneous implants, 1:436–438
- Percutaneous neurostimulation testing (PNE), 1:433–434
- Percutaneous transluminal angioplasty (PTA), 1:615
- Percutaneous transluminal coronary angioplasty (PTCA), 1:601; 2:349 with stent placement, 4:542
- Perflubron, 1:515
- Perfluorochemical emulsions, as blood substitutes, 1:513
- Perfluorochemicals (PFCs) clinical studies with, 1:515–516 liquid ventilation with, 1:516 oxygen content of, 1:513–514
- Performance criteria for gas systems, 3:380–381 for vacuum systems, 3:383
- Performance improvement, in a total hospital safety program, 6:117
- Performance maximization, 1:387
- Performance optimization, 1:387–388
- Perfusion measures, in shock assessment, 6:167
- Perfusion MR, 5:246
- Perfusion ports, in anorectal manometry, 1:62
- Pericellular matrix (PCM), 2:65
- Perimeter effects, with metal electrodes, 1:146–148
- Perimetry, 6:529–530
- Periodontal ligament (PDL), 6:410, 413–414
- Periodontum, 6:413–414
- Periontogenic illness, as a hospital problem, 6:114
- Periosteum, soft tissue grafts with, 2:72–73
- Peripheral arterial occlusive disease (PAOD), 5:234 assessment of, 5:242
- Peripheral auditory mechanism, 1:94f
- Peripheral blood lymphocytes (PBL), 4:111
- Peripheral blood mononuclear cells (PBMC), processing and collection from whole blood, 1:461–464
- Peripheral blood samples, in neonatal blood gas measurement, 5:23
- Peripheral hemodynamics, impedance plethysmography and, 4:127–128
- Peripheral nerve diseases, electron microscopic diagnosis of, 4:486
- Peripheral nerve interfaces, long-term, 3:119–125
- Peripheral nerve signals, use in neuroprosthetic applications, 3:125–129
- Peripheral nerve stimulation (PNS), 6:225
- Peripheral nerve studies, biomagnetic measurements and, 1:247–248
- Peripheral nervous system, 3:109–110
- Peripheral nervous system applications, for porous biomaterials, 5:404
- Peripheral quantitative computed tomography, 1:553
- Peripheral vascular disease, vascular graft prostheses and, 6:493
- Peripheral vascular noninvasive measurements, 5:234–252 nonimaging methods for, 5:235–245
- Peripheral vasculature (PV), 5:234
- Peritoneoscopy, 3:185
- Permaceel tape, 1:356
- Permanent prosthetic devices, for organ function loss, 6:182
- Permanent skin substitutes, 6:175–178 properties and uses of, 6:176t
- Peroxidase enzyme, 2:411
- Persistent pulmonary hypertension of the newborn (PPHN), 3:507, 508
- Personal computers. *See also* Computers basics of, 2:313–314 interfacing with thermistors, 6:332–333
- Personalized health, wearable electrodes for, 1:141–142
- Personnel dosimeters, parameters of, 5:517t
- Perturbation factors, IVB device, 1:609–613
- PET–CT hybrid imagers, 5:104
- PET–CT scanners, 2:244. *See also* PET scanners; PET scanning
- PET–CT simulator, 5:531–532
- PET–CT systems, 5:109f. *See also* PET systems; Positron emission tomography (PET) computers in, 5:119–121 development of, 5:416–417
- PET H₂¹⁵O compartment model, 6:431–432
- PET imagers, animal, 5:105–106
- PET imaging three-dimensional, 5:102–103 two-dimensional, 5:101–102
- PET instrumentation, current status and future aspects in, 5:415–417
- PET scanners, 5:411 performance characteristics of, 5:415t
- PET scanning, 2:244 procedure for, 5:414–415 risks associated with, 5:415
- PET systems, 5:101
- PGA nonwoven sheet scaffold, 1:378
- pH hemoglobin oxygen affinity and, 2:41 measurement in blood, 6:525 vaginal, 6:153
- Phagocytes, 1:113
- Phagocytosis, biomaterial failure and, 1:280
- Phantom(s) cobalt unit calibration in, 2:129–130 gel, 5:264 geometric, 5:264–265 humanoid, 5:265–266 Lucy 3D precision, 5:265 in orthovoltage X-ray beam dosimetry, 6:585 Quasar, 5:265 solid, 5:263–264 tissue equivalent, 5:262–263
- Phantom limb pain, 6:437
- Phantom materials applications of, 5:262–267 defined, 5:252 in diagnostic imaging, 5:266–267 formulation procedures for, 5:255–256 future of, 5:267–268 historical background of, 5:252–253 in nuclear medicine, 5:267 physics background of, 5:253 in radiation dosimetry, 5:262 in radiation therapy, 5:262–266 radiological equivalence of, 5:253 in radiology, 5:252–269 in simulated tissues and critical tissue elements, 5:254–255 types of, 5:256–258
- Pharmaceuticals, radiolabeled, 5:565–566
- Pharmacodynamics (PD), 5:269–270 exposure-effect link in, 5:272–273 role of modeling and simulation in, 5:275–276 statistical variation and, 5:273–275
- Pharmacogenomics, use of microarrays in, 4:371
- Pharmacokinetics (PK), 5:269–270 concentration–time curve and, 5:270–272 microdialysis sampling in, 4:412–413 role of modeling and simulation in, 5:275–276 statistical variation and, 5:273–275
- Pharmacological models, physiologically based, 5:303–305
- Phase Bode Plot, 5:375
- Phase diagram, of nickel–titanium shape memory alloys, 1:3–4
- “Phase-locking,” 1:244
- Phase-separation method, scaffold fabrication via, 1:378

- Phase-transformation temperature range (TTR), 1:1
 chemical composition effect on, 1:6
 heat treatment effect on, 1:6
 mechanical deformation effect on, 1:6
- Phasic responses, electrodermal activity and, 1:173
- pH electrode, 1:467
- PHEMA polymer, 1:277. *See also* Poly(hydroxyethyl methacrylate) (PHEMA)
- Phenomenological joint models, 4:216
- pH gradient loading method, 2:469
- Philips PageWriter Touch 12-lead ECG system, 3:42–47
- Phlebotomists, role of, 1:456
- Phlebotomy, 1:455, 458–459
- Phlebotomy techniques, for blood collection, 1:19
- Phonocardiography, 5:278–290. *See also* Electronic stethoscope
 evaluation of, 5:289
 hardware and software for, 5:289
 heart sound processing and, 5:287–288
 heart sounds/murmurs and, 5:279–282
 heart vibrations and, 5:282–283
 process of, 5:283–287
- Phosphate ions, bioactive glasses and, 1:286
- Phosphenes, 6:539–540
- Phosphors, physical properties of, 6:140t
- Phosphorus oxides, as biomaterials, 1:273
- Photoacoustic spectrography, 1:479
- Photobleaching, in confocal microscopy, 4:472–474
- Photocathodes, 5:514
- Photochemical–optical technology, 1:474–475
- Photochemotherapy, psoralen, 6:483
- Photodermatoses, 6:481t
- Photodetectors, for optical sensors, 5:164
- Photodisintegration, 6:597
- Photodynamic therapy, fiber optics in, 3:314
- Photoelectric effect, 1:469–470; 6:594–595
- Photoelectric reflectivity pattern (limbus) trackers, 3:276–277
- Photographic detectors, 5:518–519
- Photography. *See* Medical photography; Photomicrography
- Photoionization detector (PID), 4:325–326
- Photolithographic protein patterning, 1:411
- Photolithography, in microfabrication, 4:425
- Photolithography-based cellular patterning techniques, 1:409, 410–411
- Photoluminescence (PL), 5:516
- Photomicrography, 5:296–297
- Photomicroscope, 5:296–297
- Photomultiplier tubes (PMTs), 1:53, 54; 5:109, 110–111. *See also* PMT array
- Photon attenuation, 6:591–593
- Photon beams, beam quality specifiers for, 5:475–476
- Photon beam scattering, coherent, 6:593–594
- Photon DR, 1:78
- Photon-emitting remote afterloaders, regulations related to, 2:162–165
- Photon fluence, 6:591
- Photon interaction modes, 6:591
- Photon scattering, Compton, 6:595–596
- Photopatch testing, 6:481–482
- Photopheresis, 6:486
- Photoplethysmograph (PPG) sensors, 1:175
- Photoplethysmography, 5:239–241
 vaginal, 6:150–152
- Photoresist-based methods, of 3D patterning, 1:413
- Photosensitivity investigations, 6:481–482
- Photostimulable phosphor systems, 6:557
- Phototherapy
 blue light, 6:486
 calibration of meters for, 6:480
 home, 6:486
 risks of, 6:483–484
 ultraviolet, 6:482
- pH-sensitive hydrogels, 5:391
- pH sensors, optical, 5:169–170
- Phylogenetic trees, 1:220
- Physical activity, importance of, 1:388. *See also* Exercise entries
- Physical impairments, assistive devices for, 2:222
- Physical–motor assessment, for augmentative and alternative communication systems, 2:207
- Physical Parameters Transformation, 1:395
- Physical segregation cellular patterning techniques, 1:409–410
- Physical training, 1:393–394
 principles for, 1:391–393
- Physician programmer, 1:435
- Physicians, radiation-related training requirements for, 2:167t
- Physiological dead space Bohr equation, 5:431–432
- Physiological heat balance, in infant incubators, 4:146
- Physiologically based pharmacological models, 5:303–305
- Physiological research, strain gages in, 6:288
- Physiological systems modeling, 5:299–331. *See also* Physiome project
 action potentials, 5:320–322
 cardiovascular–circulation, 5:307–310
 cardiovascular regulation, 5:310–313
 cerebral, 5:313–316
 coronary, 5:316
 endocrine, 5:319–320
 geriatric, 5:323
 microcirculation, 5:313
 pulmonary–respiratory, 5:316–318
 regional circulation and autoregulation in, 5:305–307
 renal, 5:318–319
 resources related to, 5:301
 thermal, 5:322–323
 uses for, 5:323–324
- Physiologic monitors, in delivering anesthetics, 1:29
- Physiology, tactile, 6:291–293
- Physiome project, 5:324–326
- Picture Archiving and Communication System (PACS), 5:331–359, 549–550. *See also* Computed Radiography systems; Enterprise level PACS; PACS entries
 architectures used in, 5:334
 components and features of, 5:335–356
 historical overview of, 5:331–334
 standards in, 5:332–334
 teleradiology and, 6:303–304, 308–309
 trends and future prospects in, 5:357–358
- Pi (π) electrons, in graphite, 1:297
- Piezoelectric conversion, 4:432
- Piezoelectric crystal, 6:456
- Piezoelectric generator, 4:260
- Piezoelectric genosensors, 5:365–366
- Piezoelectric sensors, 5:359–367
 applications of, 5:363–366
 equipment and experiments for, 5:362–363
 as immunosensors, 5:363–365
 theory behind, 5:360–362
- Piezoelectric transducers, 4:529
- Piezoresistive devices, 2:1
- Piezo scanner, 4:437
- Pills, endoscopic wireless, 1:423–424
- Pinhole images, 6:574
- Pin-on-disk wear geometry, 1:316
- Piped vacuum systems, 3:382–383
- Piping, gas system, 3:380
- Piston prover, 5:374
- Pit and fissure sealants, 6:98
- Pit corrosion cells, 1:310
- PITV ratio, 5:581–582
- Pixel size, in computed tomography, 2:251
- PLA scaffolds, 1:379
- Planar radiometers, 1:233–234
- Planar joint motion, 4:207
- Planar sound waves, 6:453–454
- Planar transducers, 4:63–64
- Plane of maximum curvature, 6:124–125
- Plane radiography, for scoliosis, 6:126
- Planning target volume (PTV), 5:462–463, 545, 546, 547; 6:32, 33
- Plantibodies, 4:601
- Plaque morphology perturbation, 1:612
- Plasma-based modifications, to biomaterials, 1:345
- Plasma immersion ion implantation (PIII), 1:347
- Plasma polymerization, 1:345–346
- Plasma polymerization monomers, 1:346t
- Plasma processing, 1:461
- Plasma viscosity, 1:501
- Plastic deformation, of bone, 1:530–531
- Plastic dental amalgam, 1:323
- Plastic scintillators, 5:482–483, 513–514
- Plastic vertebra, 3:573
- Plate electrodes, 1:138, 139
- Platelet aggregation, collagen-induced, 1:107
- Platelet count (PLT), 2:87

- Plating, in medical microbiology, 4:372–375
- Platinum electrodes, 1:129–130
- Platinum-iridium coil electrode, 1:151
- Plethysmograph, penile, 6:159–160
- Plethysmographic blood pressure measurement, 1:489
- Plethysmography, 5:237–241. *See* Impedance plethysmography
- air, 5:239
- functional residual capacity by, 5:438
- impedance, 5:238–239
- inductive, 4:131
- magnetic susceptibility, 4:131
- respiratory inductive, 5:378
- strain gauge, 5:237–238
- Pleural pressure measurements, 6:520–521
- PLGA scaffolds, 1:379
- PMT array, 1:54–55, 56, 58. *See also* Photomultiplier tubes (PMTs)
- Pneumatic artificial heart, 3:457
- Pneumatic ventilators, 6:506
- Pneumatic ventricular assist devices, 3:451–452
- Pneumohydraulic pump, in anorectal manometry, 1:62
- Pneumolarynges, 4:231
- Pneumotachography, in neonatal monitoring, 5:14
- Pneumotachometers, 5:367–379. *See also* Spirometers
- calibration of, 5:372–374
- correction of, 5:376–377
- design specifications for, 5:377–378
- dynamic characteristics of, 5:374–376
- Fleish, 5:369
- fluid-resistance, 5:369–370
- nonfluid-resistance, 5:370–371
- Osborne, 5:370
- ultrasound-acoustic, 5:370
- Pneumothoraces, upper airway, 3:507
- pO_2 electrode, 1:466
- pO_2 measurement, continuous
- intraarterial, 5:212–213. *See also* Transcutaneous PO_2 entries
- pO_2 sensors, optical, 5:169
- POCO AXF 5Q synthetic graphite, 1:299
- Point dose verification, 5:601
- Point doses, optimal, 6:43–44
- Point of care devices, 1:23t
- Point of care testing (POCT), 1:18, 22–23
- Point-of-gaze, measuring in the presence of head motion, 3:273–275
- Point REsolved SpectroScopy (PRESS), 5:80–81
- Point-to-point video conferencing, 5:159
- Poiseuille, Jean Louis Marie, 1:504
- Poiseuille's law, 1:504
- Poisoning, activated charcoal as antidote for, 1:302
- Poisson distribution, 6:245
- Poisson's ratio, of bone, 1:529
- Poland, infrared imaging in, 6:353–354
- Polarization
- electrode-electrolyte interface and, 1:125–126
- of metallic biomaterials, 1:309–311
- Polarization-based oxygen sensor, 5:206
- Polarization microscopy, 4:501–502
- Polarization resistance (R_p), 1:311
- Polarographic gas analyzers, 6:524
- Polarographic oxygen analyzers, 5:203–204
- Polarography, oxygen, 1:466–467
- Polyacetal, 1:338
- Polyacrylamides, 1:339
- Polyacrylates, 5:388
- Poly(α -esters), in tissue engineering, 6:385
- Poly(α -hydroxy ester)s, in tissue engineering, 1:372
- Polyamides, 1:337–338; 5:388
- Poly(anhydrides), 1:260
- in tissue engineering, 1:372–373; 6:385–386
- Polycarbonates, 1:259–260, 338; 5:388
- Polycrystalline bioceramics, 1:284
- Poly(dimethylsiloxane) (PDMS), 1:422, 509; 5:388
- as biomaterial, 1:106
- stamp, 1:411, 412f
- Poly(dioxanone), 1:260
- Polydispersity index (PI), 1:331
- Poly(DTE carbonate), 1:259–260
- Polyesters, 1:338
- Polyethylene (PE), 1:333–335
- as biomaterial, 1:106–107
- wear and, 1:316–317
- Polyethylene-based tissue substitute manufacture, 5:257
- Poly(ethylene glycol) (PEG), 1:276–277
- in tissue engineering, 6:386
- Poly(ethylene oxide) (PEO), in tissue engineering, 1:373
- Polyethylene terephthalate (PET), 1:338
- Poly(glycolic acid) (PGA), 1:258–259, 260
- as biomaterial, 1:107
- Polyglycolide (PGA), 1:339–340, 371–372
- Poly(glycolide-co-lactide) (PLGA), 1:339, 340, 371–372
- Poly(glycolide lactide), 5:388
- Polygraph examinations, 6:161
- Polyhedral boranes, 1:573–574
- PolyHeme, clinical trial of, 1:519
- Poly(hydroxy acids), 1:258–259
- Poly(hydroxyalkanoates), in tissue engineering, 1:371
- Polyhydroxybutyrate (PHB), 1:371
- Poly(hydroxybutyrate-co-hydrovalerate) (PHBV), 1:371
- Poly(hydroxyethyl methacrylate) (PHEMA), 1:339
- Polyimide substrates, for thin-film electrodes, 1:156–157
- Poly(lactic acid) (PLA), 1:258–259, 260
- Poly(lactide) (PLA), 1:339, 371–372
- Poly(lactide-co-glycolide) (PLG), 1:258–259
- Poly L-lactic acid (PLLA), as biomaterial, 1:107
- Polymerase chain reaction (PCR), 4:106–107; 5:380–387. *See also* Quantitative polymerase chain reaction (QPCR)
- creation of novel recombinant molecules by, 5:383–384
- degenerate, 5:384
- as a detection system, 5:384
- DNA, 5:380–381
- length limitations in, 5:382–383
- mutations and, 5:382
- nucleic acid amplification procedures and, 5:385
- real-time, 4:376–377
- sensitivity and contamination of, 5:381–382
- Polymer gels, 5:485–486
- MR imaging of, 5:486–487
- optical scanning of, 5:487–488
- ultrasound imaging of, 5:488
- vibrational spectroscopic imaging of, 5:488
- X-ray CT scanning of, 5:488
- Polymer grafting, 1:347
- Polymer/hydrogel molding, 1:413
- Polymeric biomaterials, 1:329–342
- ASTM standards for, 1:334t
- composition, structure, and properties of, 1:331–333
- degradable, 1:279
- evaluation of, 1:341
- properties of, 1:334t
- structures and trade names of, 1:333t
- Polymeric drug delivery implants, 2:441
- Polymeric drug delivery systems, biodegradable, 2:504
- Polymeric materials, 5:387–392
- as biomaterials, 1:105–107
- in biomedical engineering applications, 5:388–391
- chemical and physical properties of, 5:388
- for drug delivery, 5:390–391
- used in tissue engineering, 5:388–390
- Polymeric particles, as drug carriers, 2:480
- Polymeric scaffold fabrication techniques, 5:390
- Polymerization, 1:330–331
- of dental resin composites, 1:323
- Polymerization heat, of acrylic bone cement, 1:543–544
- Polymer microfluidic devices, 4:425–426
- Polymer photolithography, 1:410–411
- Polymer powder-based tissue substitute manufacture, 5:257
- Polymers, 1:330
- absorbable, 1:107
- biodegradable synthetic, 1:339–340
- as biomaterials, 1:274–278
- biomedical applications of, 1:333–341
- in cardiovascular applications, 1:276
- chemical structures of, 1:368–369f
- condensation, 1:330t
- crystallinity of, 1:257–258
- in engineered tissue, 3:194–196
- medical uses of, 1:314
- molecular weight of, 1:331
- morphology of, 1:257
- natural, 5:389–390
- resorbable, 1:257–260
- structure of, 1:257, 274, 275f
- synthetic, 5:388–389
- for tissue engineering, 1:276

- Poly(methyl methacrylate) (PMMA), 1:267, 277, 335–336; 5:387
 applicators, 6:21
 as biomaterial, 1:106
 in bone cement, 1:540, 541, 542, 546
 in dentistry, 1:327
- Poly(methyl methacrylate)-rigid gas permeable contact lens design, 2:322–323
- Poly-*n*-isopropylacrylamide (PNIPAM), 4:382
- Polyolefins, 1:333–335; 5:388
- Poly(oxyethylene), 1:338
- Polyphosphazene, in tissue engineering, 1:373
- Polypropylene, 1:335
- Poly(propylene fumarate), in tissue engineering, 1:373
- Poly(propylene oxide) (PPO), in tissue engineering, 1:373
- Polysomnogram assay (PSGA), 6:215–218
- Polysomnographic data, computer analysis of, 6:214–219
- Polysomnographic recording, 6:209–211
 scoring and interpreting, 6:211–212
- Polysomnography
 ambulatory measures in, 6:212
 digital, 6:212–213
 integration of computers into, 6:214
 neonatal, 5:30
- Poly(tetrafluoroethylene) (PTFE), 1:335; 5:388
 vascular graft prostheses, 6:495
- Poly(trimethylene-carbonate) (PTMC), 1:260
- Polyurethane-based tissue substitute manufacture, 5:257
- Polyurethanes, 1:336–337; 5:388
 in tissue engineering, 1:373–374
- Poly(vinyl alcohol) (PVA), 1:339
 in tissue engineering, 1:372; 6:386
- Poly(vinyl chloride) (PVC), 1:335; 5:388
- Pop-off valve, 1:33
- Population variation, in pharmacokinetics and pharmacodynamics, 5:273–275
- Porcelain
 composition of, 1:326
 in dentistry, 1:326
- Porcelain-fused-to-metal (PFM)
 restorations, 1:325
- Porogen leaching methods, scaffold fabrication via, 1:378
- Porosity
 of acrylic bone cement, 1:545–546
 of bioceramic scaffolds, 1:375, 379
- Porous biomaterials. *See also* Porous materials
 for cartilage applications, 5:403
 for central and peripheral nervous system applications, 5:404
 fabrication methods for, 5:400–401
 future directions in the design of, 5:401–404
 mechanical properties and degradation byproducts of, 5:398–399
 next generation of, 5:393–400
 orthopedic applications for, 5:402
- porosity, pore size, and interconnectivity of, 5:399–400
- protein interactions with, 5:393–397
 for vascular and bladder applications, 5:403
- Porous biomaterial surfaces
 design of, 5:397–398
 properties of, 5:393
- Porous electrodes, 1:152, 153
- Porous implant surfaces, 1:130
- Porous layers
 for electrodes, 1:140–141
- Porous materials, biological applications for, 5:392–406. *See also* Porous biomaterials
- Porous melt-derived bioactive glasses, 1:291–292
- Porphyrins, boron-containing, 1:575
- Portable ECG recorder, 1:13
- Portable reading aids, 1:445–446
- Portal vein thrombosis, after liver transplantation, 4:271–272
- Positive airway pressure, bilevel, 2:332
- Positive beam limitation, in X-ray equipment, 6:566–567
- Positive end-expiratory pressure (PEEP), 3:500; 6:508, 511–512, 514
- Positive punishment, 1:167
- Positive reinforcement, 1:166–167
- Positive temperature coefficient (PTC), 6:321
- Positron emission tomography (PET), 5:108, 406–418. *See also* PET entries
 applications of, 5:412–414
 history of, 5:407–408
 image interpretation in, 5:414
 in peripheral vascular noninvasive measurements, 5:246–247
 physical principles of, 5:408–412
 radiopharmaceutical manufacture and, 5:412
 risks associated with, 5:415
- Postanesthesia care units (PACUs), 1:44
- Posterior lumbar interbody fusion (PLIF) surgery, 6:235
- Posterior plating fusion techniques, 3:577–578
- Posterior spinal instrumentation, 3:579–580
- Posterior tibial nerve stimulation (PTNS), 1:430
- Postero-lateral gutter fusion surgery, 6:236
- Postexercise period, in exercise stress testing, 3:252–253
- Postmarket rules, for medical devices, 2:148
- Postoperative hypertension, 1:491
- Postoperative pain, abolition of, 3:31
- Potassium hydroxide (KOH), in micromachining, 2:3
- Potential
 electrical, 1:466
 electrode-electrolyte, 1:122–123
 at electrode-skin interface, 1:121
- Potential distribution, 3:110–111
- Potential motion artifact, of skin, 1:133–134
- Potentiometer circuit, 6:285
- Pourbaix diagram, 1:7
- Power Doppler, 6:465
- Power Doppler imaging (PDI), 5:245
- Power flow imaging (PFI), 5:245
- Power generation, with ambient vibration, 4:431–433
- Power generator, micro, 4:430–433
- Power ratio, EGG, 3:90
- Power spectra analysis, EEG, 3:70
- Power spectrum estimation
 using the autoregressive model, 3:77–79
 using the Welch method, 3:75–77
- Power supply, for electrophoresis, 3:138
- Power switching, 3:210–212
- Preamplification, in phonocardiography, 5:286–287
- Preclinical neuroprosthetic systems, 3:126–127
- Pre-effector T cells, tumor-reactive, 4:111–112
- Pregelged foil electrodes, 1:144
- Pregnancy, hypertension in, 1:16
- Premarket “510(k)” notification, for medical devices, 2:145–146
- Premarket approval (PMA), of medical devices, 2:146
- Premarket notification exemptions, for medical devices, 2:145
- Premature ventricular contractions (PVCs), 1:70
- Preplanning, for prostate seed implants, 5:419
- Preprocessing, in computer-assisted detection/diagnosis, 2:289–293
- Presbyopia, contact lens designs for, 2:326–327
- Presentation of data, 1:396
- Present Value Interest Factor (PVIF) table, 1:24
- Preservation, by freezing, 1:192
- Preservation injury, after liver transplantation, 4:272
- Pressure, “therapeutic,” 2:333
- Pressure alarms, anesthesia machine, 1:37
- Pressure–diameter–axial force arterial elasticity measurement, 1:86
- Pressure–diameter relations, for arterial elasticity, 1:87–88, 89
- Pressure drop
 during guidewire diagnostics, 2:355–357
 heart valve prostheses and, 3:415–416
- Pressure measurement, neonatal, 5:26–28
- Pressure monitoring systems, 2:8–9
- Pressure ratio–distension ratio, for arterial elasticity, 1:87–88
- Pressure recording, in continuous positive airway pressure, 2:332–333
- Pressure-reducing valve (PRV), 6:506
- Pressure regulators, anesthesia machine, 1:35
- Pressure sensors, optical, 5:172
- Pressure support ventilation (PSV), 6:510
- Pressure transducers, in neonatal hemodynamic monitoring, 4:593
- Pressure–volume loops, 3:477

- Pressure–volume relations, for arterial elasticity, 1:89
- Pressurized gas systems, 3:377–381
- Presurgical functional mapping, 1:244–245
- Preventive maintenance (PM), 3:223
- Preventive maintenance interval, selecting, 3:224–225
- Preventive maintenance procedures, 3:225 sample, 3:226f
- Preventive maintenance program effective, 3:224 evaluating, 3:225–227
- Preventive maintenance stickers, 3:228
- Preview digital radiograph, 2:237
- Primary disease, recurrence following liver transplantation, 4:273
- Primary graft nonfunction (PNF), after liver transplantation, 4:271
- Primary medical physics journals, 4:335–336
- Primary pathogens, 1:113–114
- Primary radiation barrier, 6:570
- Principal Component Analysis (PCA), 1:241; 6:261
- Privacy issues, related to augmentative and alternative communication systems, 2:210
- Probability, 6:243–244 relative frequency and, 6:243–244
- Probability density function (PDF), 1:71 for continuous variables, 6:244
- Probability distributions, 6:244 characteristics of, 6:246
- Probability mass function, 6:244
- Probe configurations, in optical sensors, 5:162–163
- Probe cryosurgery technique, 2:366–367
- Probe geometry, in microdialysis sampling, 4:402–403
- Probe insertion trauma, in microdialysis sampling, 4:408–409
- Probe/junction materials, in thermocouples, 6:344
- Probe materials, in microdialysis sampling, 4:403–404
- Probe placement, in anorectal manometry, 1:63
- Probes. *See also* Microdialysis probes cytochemical, 2:390–392 extrinsic, 4:497–498 fiber-optic, 6:358–359 fluorescent, 4:494–498 genetic expressible, 4:498 intrinsic, 4:497 in nuclear medicine detectors, 5:95 organelle, 4:469–470 solid-state, 1:63 thermistor, 6:333
- Proctalgia, 1:68
- Proctological diseases, cryosurgical treatment of, 2:374–375
- Product development protocol, 2:146
- Production condition optimization, microbioreactors for, 4:393
- Productivity, with Mount Sinai total laboratory automation, 1:26–27
- Productivity tools, in office automation systems, 5:152–155
- Professional organizations, radiation-related, 2:154. *See also* American entries; International entries; National entries; Society entries
- Professional pathogens, 1:113–114
- Prognostics, use of microarrays in, 4:370–371
- Progressive resistance exercises, 1:391
- Projection data, in CT reconstruction methods, 2:246
- Projection radiography, acquisition of, 5:336
- PRO-MED AG SmartDose, 2:504
- Promoter prediction and detection, 1:222
- Pro Osteon, 1:261–262, 263
- Propagation modes, in fiber optics, 3:302–303
- Proportional-integral-derivative (PID) blood pressure controllers, 1:48, 491–492, 497, 498
- Proposal evaluation form, 3:218t
- Prostate ablation, 6:375–376
- Prostate cancer, high intensity focus ultrasound for, 4:77–78
- Prostate gland diseases, cryosurgical treatment of, 2:375
- Prostate seed implants, 5:418–429 analysis following, 5:427 dose escalation to prostate sub-volumes, 5:427 dosimetry in, 5:420–424 ¹²⁵I and ¹⁰³Pd in, 5:419–420 isotope selection and dose prescription for, 5:419 preplanning/intraoperative planning for, 5:419 problems and remedies related to, 5:427–428 secondary dose verification in, 5:427 treatment planning in, 5:424–427 ultrasound-guided, 5:424
- Prostatic hypertrophy, treatment of, 4:542
- Prostheses. *See also* Visual prostheses bone cement, 1:540–541 cochlear, 2:133–141 heart valve, 3:407–426 hip, 5:195–196 intervertebral disk (IVD), 5:197 knee, 5:196 pyrolytic carbons in, 1:302–304 titanium alloys in, 1:312 upper extremity and IVD, 5:196–197 vascular graft, 6:491–505 wear of, 1:314–315, 316
- Prosthesis designs, fixation of, 5:195–197
- Prosthesis fixation cemented, 5:194 evaluation and future strategies in, 5:197 for orthopedics, 5:192–198
- Prosthesis-related infection, 1:542
- Prosthetic devices. *See also* Prosthetics laryngeal, 4:229–234 for organ function loss, 6:182
- Prosthetic feet, 4:552–553
- Prosthetic heart valves, 3:437–449. *See also* Bioprosthetic heart valves; Heart valve prostheses; Mechanical heart valves anticoagulation and durability in, 3:447 challenge of, 3:437–438 clinical performance of, 3:438–439 demographics and etiology related to, 3:438 design concepts for, 3:442–443 design elements for, 3:441 future directions in, 3:447 hemodynamic evaluation of, 3:439–441 materials technology related to, 3:441–442 patient data for, 3:439t
- Prosthetic heart valve technology, 3:428–430
- Prosthetic heart valve testing data analysis for, 3:431–432 design-specific, 3:434–435 Doppler ultrasound in, 3:432–434
- Prosthetic intervertebral nucleus (PIN) device, 3:586–588
- Prosthetic motion, effects of, 1:111
- Prosthetic restorative materials in dentistry, 1:325–329
- Prosthetics, lower extremity, 4:552–554. *See also* Prosthetic devices
- Protein(s) adsorption at surfaces, 1:344–345 bone bonding and, 1:110 in complement system, 1:112 diverse properties of, 5:395t DNA, RNA and, 1:217 fluorescent probes for, 2:392t folding, simulation, and structure prediction of, 1:220–221; 4:512–513 interactions with porous biomaterials, 5:393–397 patterning, 1:410–411, 412f
- Protein adsorption, biomaterial failure related to, 1:279–280
- Protein detection, nanoparticles in, 5:7
- Protein detection immunoassay, for microdialysis samples, 4:411–412
- Protein-mediated cell adhesion, 5:396–397
- Protein network, 1:225
- Protein sequences, in bioinformatics, 1:217–220
- Proteoglycans (PGs), 2:63, 64 in tendons and ligaments, 4:242
- Protocol export, from a computed tomography simulator, 2:273
- Proton-beam radiotherapy, 6:8–10
- Proton resonance frequency (PRF), 6:360
- Proton therapy beams, 5:576
- Provocation testing, 6:481
- Provox valve, 4:234
- Proximal catheter, for treating hydrocephalus, 4:9–10
- Proximal interphalangeal (PIP) finger joint replacements, 1:304
- Proximity telemetry, in temperature measurement electronics, 6:330

- Pseudoelasticity, 1:2. *See also* "Rubber-like behavior"
- Pseudomonas aeruginosa*, 1:114, 319
- Psoralen photochemotherapy (PUVA), 6:483
- Psoriasis clearance, using UVR, 6:482
- Psychiatric disorders, cognitive training for persons with, 6:72–73
- Public access defibrillation (PAD), 2:49, 56
- Public education books/reports, 4:350
- PubMed, 4:309
- Pudendal nerve stimulation, for overactive bladder, 1:441
- Pulmonary artery (PA) catheters, 2:29–30
placement of, 4:572–573
- Pulmonary artery oximetry, 5:214–215
- Pulmonary artery pressure, measurement of, 4:572–573
- Pulmonary blood flow (PBF), 2:16–18
- Pulmonary capillary wedge pressure (PCWP), 6:166
- Pulmonary circulation, 2:41–42; 4:567
- Pulmonary closing volume, 5:439
- Pulmonary diffusion/diffusing, 5:432–433
- Pulmonary disease, parenteral nutrition for, 5:132–133
- Pulmonary edema detection, 1:213
- Pulmonary function, complex models of, 6:103
- Pulmonary function testing/tests, 5:430, 436
future of, 5:440–441
physiological principles underlying, 5:431
results of, 5:433
standardization of, 5:440
- Pulmonary interstitial emphysema (PIE), 3:506–507
- Pulmonary physiology, 5:429–442. *See also* Pulmonary function testing/tests
CPR and, 2:38–40
instrumentation related to, 5:434–435
post-1940s and the gold age of, 5:430
pre-1940s, 5:429–430
terminology related to, 5:441
- Pulmonary resistance, 6:516–517
- Pulmonary–respiratory systems modeling, 5:316–318
- Pulmonary–systemic (P/S) flow ratio, 2:17–18
- Pulmonary vasculature, 6:103
- Pulmonary wedge pressure (PWP), 1:206
- Pulsatile flow ventricular assist devices, 3:451
- Pulsed Doppler (PW), 3:8, 10
ultrasound, 3:330
- Pulse detection, in blood pressure measurement, 1:488
- Pulsed-laser accelerators, 6:13
- Pulsed magnetic fields (PMF), 1:560, 564–565
- Pulsed short-wave diathermy (PSWD), 3:471–472
- Pulse duration, short vs. long, 1:128
- Pulsed wave Doppler (PW), 3:1
- Pulse–echo method, in ultrasonic imaging, 6:455–456
- Pulse generators, 5:220–222
implantable, 6:225–226
- Pulse interval generator, 5:221
- Pulse inversion imaging mode, 6:469–471
- Pulseless electrical activity (PEA), 2:47–48
- Pulseless ventricular tachycardia, 2:45–46
- Pulse oximeters, 3:535f; 6:523–524
accuracy and limitations of, 1:474
clinical uses for, 1:473–474
future directions for, 1:474
- Pulse oximetry, 1:469, 471–474; 5:210–212
impact of, 2:117
in neonatal blood gas measurement, 5:24–25
new technologies in, 1:472–473
optical sensors in, 5:166–167
in sleep laboratory, 6:212
- Pulse pileup, Anger camera, 1:59–60
- Pulse rate home health care devices, 3:528–529
- Pulse repetition frequency (PRF), 6:463
- Pulse volume recording (PVR)/pulse wave analysis (PWA), 5:243
- Pulse wave Doppler (PW Doppler), 6:463–464
- Pulse-wave velocity (PWV), 1:489; 5:243
- Pulse wave velocity model, 6:408
- Pump failure, left ventricular, 4:164
- Pumps, microbioreactor, 4:390
- Punishment, in operant conditioning, 1:167
- Pupillometry, sexual arousal and, 6:160
- Pupil recognition, 3:269–270
- Pure tone audiometry, 1:93–96, 97
- Purification techniques, for molecules, 4:603
- Purkinje fibers, 2:43
- Pursuit eye movements, 5:139
- Purulent infection, implant-related, 1:116
- PUVA treatment, 6:483
- p*-values, 6:246
- Pyramidal neurons, activity of, 3:63–64
- Pyridoxilated hemoglobin polyoxyethylene (PHP), clinical trial of, 1:520
- Pyrolite, 1:302
- Pyrolytic carbons (PyCs), 1:299–300
medical applications of, 1:302–306
- Pyrosequencing, 2:433
- QRS complex, 2:44, 45, 46
- Quadripolar extension, 1:435
- Quality, X-ray beam, 6:583–584
- Quality measurements, in radiation dosimetry, 5:476–479
- Quality assurance (QA)
dosimeter characteristics required for, 5:482
EPID use and, 4:96
in high-dosage-rate brachytherapy, 1:599
in the hospital safety program, 6:119
in intraoperative radiotherapy, 6:23
in neutron activation analysis, 5:47–48
in radiation therapy, 5:542–549
linear accelerator, 5:547
simulation, 5:543–545
treatment delivery, 5:548
treatment planning, 5:545–547
for X-ray therapy equipment, 6:589
- Quality assurance devices, 5:601
- Quality assurance procedures, dosimetry systems and, 5:481–482
- Quality assurance/quality control (QA/QC), for screen-film systems, 6:147–148
- Quality control. *See also* X-ray quality control program
for Gamma Knife, 3:375–376
in phototherapy equipment, 6:484–485
in tissue substitute manufacture, 5:257–258
- Quality of life, as a "soft" endpoint, 5:451
- Quality-of-life measures, 5:443–454. *See also* Change; Clinical significance
clinical significance of, 5:444–445
historical background of, 5:443
- Quality system regulations, FDA, 2:150–152
- Quantitative colorimetric measurement, 2:192–194
- Quantitative computed tomography (QCT), 1:553
peripheral, 1:553
- Quantitative Gibson–Cooke Pilocarpine Iontophoresis Sweat Test (GCST/QPIT), 2:385–386
potential successor for, 2:387
- Quantitative polymerase chain reaction (QPCR), 5:385
- Quantitative ultrasonometry (QUS), 1:553–556
- Quantum dots, in confocal microscopy, 4:470–471
- Quantum patch technology, 2:386, 387
- Quantum yield, 3:346
- Quartz crystal thermometry, 6:315–316
- Quasar phantom, 5:265
- Quenching, in confocal microscopy, 4:472–474
- Questionnaires, in exercise stress testing, 3:252
- Quick-Prep Applicator, 1:136
- Racing, wheelchair, 4:549
- Radial basis function (RBF), 1:494
- Radial gradiometers, 1:233, 236
- Radiation. *See also* Ionizing radiation
coupled with hyperthermia, 4:67–68
effects linear no-threshold model of, 4:183
indirectly ionizing, 5:503
into input receptor(s), 6:571–572
interaction with matter, 5:253
medical, 5:253
monochromatic, 6:478
professional organizations related to, 2:154
quantities and units of, 5:504–505
sources of, 5:504
standards for protection against, 2:160–162
types of, 5:503–504
use in medical diagnosis and therapy, 5:71
web sites with information about, 2:160t

- Radiation barrier, primary, 6:570
- Radiation beam, characteristics of, 2:127–128
- Radiation biology, 4:181
books/reports, 4:348–349
- Radiation byproduct material, medical use of, 2:162–166
- Radiation codes/regulations, 2:153–186
acronyms and definitions related to, 2:184–185
for byproduct material, 2:158–171
for nonbyproduct and machine-produced radiation, 2:171–177
for nonionizing radiation, 2:157–158, 177–184
organizations involved in, 2:153–158
- Radiation delivery, in three-dimensional conformal radiotherapy, 6:34–35
- Radiation detectors, 5:506, 507, 518
solid-state, 5:514–518
- Radiation dose(s). *See also* Radiation dosimetry
absorbed, 5:466–467
in computed tomography, 2:244–246
effects on cancer risk, 2:260
from screening CT, 2:260–261
- Radiation dose planning, computer-aided, 5:455–463
- Radiation dosimetry. *See also* Dosimetry; Gel dosimetry; Radiation dose(s)
determining absorbed dose and air kerma, 5:467–471
measurement and quantities in, 5:465–467
Monte Carlo techniques for, 5:471
for oncology, 5:465–481
phantom materials in, 5:262
relative dosimetry and quality/verification in, 5:476–479
secondary dosimeter calibration, 5:474–476
three-dimensional, 5:481–500
- Radiation exposure, effect of age at, 2:260
- Radiation exposure limits, microwave and radiofrequency, 2:183t
- Radiation fields
focused, 4:65
nonfocused, 4:63–64
- Radiation heating, 3:468–470
- Radiation leakage, 6:564–565
- Radiation levels, scattered, 6:572–573
- Radiation measurements, books/reports on, 4:348
- Radiation oncology, imaging for, 6:30–32
- Radiation oncology physics books/reports, 4:338–341
- Radiation physics books/reports, 4:350–351
- Radiation-producing equipment, nonmedical, 2:174–175
- Radiation protection, books/reports on, 4:346–348
- Radiation protection instrumentation, 5:500–520
availability of, 5:501
choice of, 5:501–503
common features of, 5:505–508
distributors and manufacturers of, 5:501t
Geiger-Müller counters, 5:511
ionization chambers, 5:508–510
major features of, 5:502
proportional counters, 5:510–511
scintillation detectors, 5:511–514
solid-state radiation detectors, 5:514–518
- Radiation protection products, medical, 5:501t
- Radiation reflection, in neonatal respiratory monitoring, 5:17
- Radiation Research Society, 4:320
- Radiation safety, in intraoperative radiotherapy, 6:23–24
- Radiation safety officers, training requirements for, 2:165t
- Radiation therapy
applications of, 5:91–92
devices for, 6:581–583
intensity modulated, 2:273–274; 5:520–525
linear accelerator quality assurance in, 5:547
neutron sources for, 5:54–57
phantom materials in, 5:262–266
quality assurance in, 5:542–549
simulation QA in, 5:543–545
treatment delivery quality assurance in, 5:548
- Radiation therapy dose calculation, applications in, 5:535–538
- Radiation therapy imaging, 5:526
- Radiation therapy simulator, 5:525–533
conventional, 5:527–529
technology overview for, 5:527
- Radiation therapy treatment planning, 2:243–244; 5:534–542
quality assurance in, 5:545–547
- Radiation thermometry, 6:360–361
- Radiation tissue damage, hyperbaric medicine and, 4:26
- Radiation treatment, breathing motion problem in, 5:593
- Radiation waveform, quality control of, 6:563–564
- Radioactive byproduct material, regulation of, 2:158–171
- Radioactive decay
rates of, 5:561–562
types of, 5:559–561
- Radioactive ionization detectors, 4:331–332
- Radioactive labels
limitations of, 5:91
strategies for using, 5:91
- Radioactive material
disposal regulations for, 2:170–171
transport regulations for, 2:166–169
- Radioactive medical devices, FDA regulation of, 2:155–156
- Radioactive nuclei, 5:562
- Radioactive source
for Guidant Galileo IVB system, 1:606
for high dosage rate brachytherapy, 1:590–591
for Novoste BetaCath, 1:604
- Radioactive source ribbon, in IVB devices, 1:603
- Radioactive tracers, 1:51
- Radiochemical separation, in neutron activation analysis, 5:45–46
- Radiochromic film, 5:477–478
measurements with, 1:614–615
- Radio frequency ablation, 6:363–364, 370–371, 373–374, 376, 378
- Radio frequency (rf) devices. *See also* rf entries
clinical studies with, 4:37
interstitial hyperthermia and, 4:35–36
therapeutic applications of, 1:190
- Radio frequency glow discharge (rfgd), 1:345
- Radio frequency exposure limits, 2:183t
- Radio frequency fields, 5:508
- Radio Frequency Identification (RFID) tags, 1:453; 4:358
- Radio frequency (RF) radiation, 5:67–68
use in medical diagnosis and therapy, 5:71
- Radio frequency (rf) SQUIDS, 1:231
- Radiographic film, 5:477, 483, 599
replacement of, 4:92
- Radiography, computed, 4:302
- Radiography-based densitometry, 1:550–553
- Radiography (imaging) units, regulations related to, 2:174
- Radioimmune-guided surgery (RIGS), 5:95
- Radioimmunotherapy (RIT), 5:95
dosimetry in, 5:570–571
- Radioiodine therapy. *See also* ¹²⁵I (iodine-125); ¹³¹I-MIBG therapy
dosimetry for thyrotoxicosis, 5:568–569
- Radiological equipment, control and supervisory logic in, 6:557–558
- Radiological physics, books/reports, 4:337–338, 349
- Radiological Society of North America, 4:320–321
- Radiologic image receptors, 6:556–557
- Radiology, phantom materials in, 5:252–269. *See also* Teleradiology
- Radiology information systems (RIS), 5:549–559
interfacing with hospital information system, 5:335
advanced feature set for, 5:552t
core feature set for, 5:551t
functionality of, 5:551–553
future of, 5:556–558
interfacing systems for, 5:553–554
options for integrating, 5:554–556
role in radiology workflow, 5:550–551
selecting, 5:556
- Radiology report, sample portion of, 5:557t
- Radiology workflow, 5:550–551
- Radiometers, UV, 6:479–480
- Radiometer TCM 4 transcutaneous blood gas monitor, 1:478f
- Radiometric quantities, 6:478

- Radionuclide identification, 5:508
 Radionuclide production, 5:559, 562–564
 Radionuclides
 exemptions from regulation control for, 2:180t
 nonbyproduct, 2:175–177
 production of, 5:92
 use in positron emission tomography, 5:408t
 Radiopharmaceutical dosimetry, 5:565–574
 MIRD schema for, 5:566–567
 in new drug development, 5:571–572
 for radionuclide therapy, 5:567–571
 Radiopharmaceuticals
 effective dose values per unit activity for, 5:567t
 manufacturing of, 5:412
 PET, 5:408
 physician training requirements for, 2:168t
 regulation of, 2:155
 Radiosurgery
 delivery techniques for, 5:578–580
 Gamma Knife, 3:368
 stereotactic, 2:275; 4:542; 5:574–585
 stereotactic frame application and, 6:269
 Radiosurgery systems, early, 5:576
 Radiosurgery treatment plans, tools for evaluating, 5:580–583
 Radiotherapy. *See also* Cobalt-60 units; Heavy ion radiotherapy; Intraoperative radiotherapy (IORT); Three-dimensional conformal radiotherapy (3DCRT)
 adaptive, 6:401
 electron-beam, 6:4–6
 fast and slow neutron, 6:6–8
 proton-beam, 6:8–10
 systemic hyperthermia and, 4:58
 Radiotherapy accelerator, requirements for, 2:176t
 Radiotherapy accessories, 5:585–602
 field-shaping, shielding, and dose modifying devices, 5:594–599
 patient restraint and repositioning devices, 5:587–594
 treatment verification and quality assurance devices, 5:599–601
 tumor localization and treatment simulation devices, 5:585–587
 Radiotherapy treatment planning, 6:40–41
 Radiotherapy treatment planning optimization, 6:38–48
 assessing benefits and risks in, 6:38–39
 biological-probabilistic approaches to, 6:44–46
 examples of, 6:39–40, 41–42
 need for quantitative optimization tools in, 6:42–43
 using physical/geometrical criteria, 6:43–44
 Radiotherapy units, regulations related to, 2:174
 Rahn, Herman, 5:430
 Raised intracranial pressure, 4:580–582
 Raman laser spectroscopy, 4:332
 Raman microscopy, 2:98–99
 Raman spectrography, 1:479
 Raman spectroscopy, 3:312–313
 in multiple gas monitoring, 5:207
 Ramp inflation method, 1:64–65
 Ramp testing, 3:252
 Randomization methods, 6:260–261
 Randomized Evaluation of Mechanical Assistance for the Treatment of Congestive Heart Failure (REMATCH) study, 3:453
 Random sampling, 5:534
 Range of motion, in scoliosis, 6:126
 Rank correlation test, Spearman's, 6:259
 Rapid eye movement (REM) sleep, 6:214, 215, 218
 Rapid prototyping, for porous biomaterial fabrication, 5:400
 Rapid shallow breathing index (RSBI), 6:518
 RAPTOR™ fiber-optic fluoroimmunoassay system, 4:378
 Rate-based analysis, for arrhythmia detection, 1:71–73
 Rate modulation, in functional electrical stimulation, 3:348–350
 Rate stability, in rate-based arrhythmia analysis, 1:72–73
 Rayleigh scattering, 6:593–594
 Raynaud's disease, biofeedback clinical outcome literature related to, 1:181
 Ray nucleus device, 3:585–586
 Ray spacing, in computed tomography, 2:251
 Reaction kinetics, in microbioreactors, 4:387
 Reactive airway disease, 3:507
 Reactive biomaterials, 1:104
 Reactor modifications, for BNCT, 1:577
 Reactor radionuclide production, 5:92
 Reading aids
 for the blind, 1:445–447
 low vision, 1:444–445
 Reading machine, components of, 1:446f
 Reagents, for electrophoresis, 3:138
 Real-time PCR (RT-PCR), 4:376–377
 Real-time Position Management (RPM) system, 5:593
 Receptor-ligand adhesion measurements, 4:510–512
 Recessed electrodes, 1:139
 Reciprocity law, 6:145
 Recombinant genetic engineering, 4:600–601
 Recombinant molecules, creation by polymerase chain reaction, 5:383–384
 Reconstruction techniques
 analytical, 2:247
 CT, 2:246–250
 direct Fourier, 2:247
 iterative, 2:246–247
 tomographic, 5:114–117
 Record and verify (R&V) system, 6:35
 Recording equipment, EEG, 3:86–87
 Recordings, with EEG biofeedback instrumentation, 1:171
 Recording systems, for lung sounds, 4:279–280
 Recreation mobility aids, 4:548–550
 Recruitment, in functional electrical stimulation, 3:348–350
 Rectal compliance, 1:65
 Rectal distension, intermittent, 1:64
 Rectal drug delivery, cyclodextrins in, 2:456, 458t
 Rectal sensory function, 1:64
 Rectal temperature monitoring, 6:318
 Rectilinear scanners, 1:51; 5:95–96
 Rectoanal inhibitory reflex, 1:64
 Rectoanal pressure, changes during attempted defecation, 1:64
 Rectum, anatomy and physiology of, 1:62
 Recursively applied and projected MUSIC (RAP-MUSIC), 1:241
 Red blood cell count, 2:87
 Red blood cells (RBCs), deformability of, 1:503–504. *See also* Erythrocyte entries
 Red cell distribution width (RDW), 2:87
 Redox paste, 1:136
 Reduction method, joint distribution problem and, 4:216
 Redundant array of inexpensive disks (RAID), 5:347–348
 Reference electrodes, in ion-sensitive field-effect transistors, 4:193–194
 Reference FETs (REFETs), 4:194–195
 Reflectance spectroscopy, fiber optics in, 3:311
 Reflection oximetry, versus transmission oximetry, 1:470
 Reflectivity pattern movement, eye orientation as a function of, 3:273
 Reflex measurement, acoustic, 1:101
 Refractive lenses
 in intraocular lenses, 4:238–239
 optical quality of, 4:239
 Refractory osteomyelitis, hyperbaric medicine and, 4:26
 Regenerating electrodes, 3:119
 Regeneration
 implantable devices and, 1:104
 tissue, 1:108–109
 Regenerative ability, of cartilage, 2:71–72
 Regenerative medicine, bioceramics in, 1:290
 Regional pain syndromes, chronic, 6:229
 Regressions, 6:256–257
 Regulation. *See also* Regulations of biomaterials, 1:268–270 of engineered tissue, 3:202
 Regulations, for nonionizing radiation, 2:177–184. *See also* Medical device codes/regulations; Radiation codes/regulations; Regulation
 Regulatory network, 1:225
 Regulatory standards, OSHA, 2:171
 Regulatory Summaries, Nuclear Regulatory Commission, 2:171
 Regurgitation, heart valve, 3:416–417
 Rehabilitation
 cognitive, 6:71–79
 muscle, 6:62–71

- orthotics in, 6:80–93
 virtual reality as a computer-generated technology for, 6:74
- Rehabilitation Engineering & Assistive Technology Society of North America (RESNA), 4:321
- Rehabilitation exercises, 1:398
- Rehabilitative biotelemetry microsystems, 1:424–427
- Rehau system, 4:579–580
- Reinforcement, of behavior, 1:166–167
- Relative biological effectiveness (RBE), 1:572; 6:3
- Relative dosimetry measurements, 5:476–479
- Relative frequency, probability and, 6:243–244
- Relative power change, EGG, 3:90
- Relaxation, biofeedback and, 1:178
- Reliability, of electrophoresis, 3:140–141
- Remote afterloaders (RALs), 1:590. *See also* High-dose-rate remote afterloaders
 high-dose-rate, 1:590–594
- Remote-control defibrillators, standards for, 1:160–161
- Remote video-based eye tracking systems, 3:280–283
- Renal systems modeling, 5:318–319
- Repassivation, 1:310–311
- Repeat-fixation stereotactic radiotherapy, 5:491
- Repetitions, exercise, 1:392, 399
- Reporting
 codes and regulations related to, 2:149–150
 human factors and, 3:541–542
- Resampling methods, 6:259–261
- Research, in murine cardiac ventricular cells, 3:145–146
- Research and development, contraception-related, 2:347–348
- Research-based software, for communication disorders, 2:212
- Reservoir bag, 1:33
 in anesthesia delivery, 1:30–31, 32
- Reshaping cuff electrodes, 3:124–125
- Residual stress, of acrylic bone cement, 1:545–546
- Residual voltage output, in linear variable differential transformers, 4:255
- Resin-based composites (RBCs), 1:322; 6:93–99
 classification and physical properties of, 6:97t
 clinical application of, 6:97–98
 fabrication of, 6:93–95
 phases in fabricating, 6:94–95
 physical and mechanical properties of, 6:95–97
- Resin composites, as dental fillings, 1:323–324
- Resin materials, prosthetic, 1:327. *See also* Resins
- Resin-modified glass ionomers (RMGIs), 1:324
- Resins, unfilled acrylic, 6:93–94. *See also* Resin materials
- Resistance. *See* Skin resistance activity (SRA)
- Resistance calibration, 1:194, 195
- Resistance strain gages, 6:284
- Resistance temperature detectors (RTDs), 6:356–357, 358
- Resistance thermometers, 6:313–314
- Resistive electrodes, 1:148–149
- Resistive training devices, with computerized feedback, 1:398
- Resistivity, of wet gels and hydrogels, 1:135–136
- Resolution
 in electron microscopy, 4:480
 lateral, 4:463–464
 in linear variable differential transformers, 4:255
- Resolution/contrast, in confocal microscopy, 4:461–462
- Resorbable bioceramics, 1:285
- Resorbable calcium ceramics, 1:260–262
- Resorbable composites, 1:262–264
- Resorbable implants
 function of, 1:256–257
 orthopedic applications of, 1:255–256
- Resorbable polymers, 1:257–260
- Respiration (RSP), versus ventilation, 6:509
- Respiration sensors/amplifiers, 1:174–175
- Respiration therapy home health care devices, 3:534–535
- Respiration training, 1:174
- Respiratory control, visceral neural signals in, 3:129
- Respiratory disturbance index (RDI), 6:220, 221
- Respiratory flow, devices for measuring, 5:367–369
- Respiratory inductive plethysmography (RIP), 5:378
- Respiratory measurements, thermistors in, 6:335–336
- Respiratory mechanics, 6:99–103
 in disease states, 6:106
 role of cardiovascular system in, 6:105–106
- Respiratory neonatal monitoring, 5:13–18
- Respiratory sinus arrhythmia (RSA)
 instruments, 1:174
- Respiratory sounds, 5:378
- Respiratory system
 dynamic events in, 6:102–103
 effects of parenteral nutrition on, 5:131
 gas transport and exchange in, 6:103–106
 static mechanics of, 6:100–102
 structure of, 6:100
- Respirometers, 5:370
- Response, stimulus and, 1:166, 167
- Restenosis
 epidemiology and clinical trials related to, 1:601–602
 mechanisms of, 1:601
 prediction with exercise test, 3:258–259
- Resting membrane potential, 3:141–142
- Restoration algorithms, smart, 6:331
- Restorative materials, directly placed, 1:322–325
- Resuscitation, after gastrointestinal hemorrhage, 3:385. *See also* Cardiopulmonary resuscitation (CPR)
- Retardation, mental, 2:211
- Reticulated alumina substrate, 1:355–356, 357, 358, 359, 360
- Reticulocyte count, 2:88
- Retinal prosthesis, 1:155
- Retinotopy, 6:531–532
- Return on investment (ROI), for Mount Sinai total laboratory automation, 1:27
- Reusable electrodes, 1:148
- Reverse iontophoresis, 3:397
- Reverse-phase evaporation vesicles (REVs), 2:468
- Reversible potential, 1:122–123
- Review standard, for medical devices, 2:147–148
- Rewarding, of behavior, 1:166–167
- rf generators, for electrosurgery, 3:161–164
- RF thermoablation, 6:345
- Rheological properties
 of blood, 1:500–502
 of blood cells, 1:506–507
 of fluids, 1:504–506
 techniques for measuring, 1:504–506
- Rheoscope, 1:507–509
- Rhythms, sensorimotor, 1:243, 244
- Ribonucleic acid (RNA). *See also* RNA entries
 isolation of, 4:362
 labeling of, 4:362–363
- Richards, Dickinson Woodrow, 5:430
- Ridley, Harold, 1:267
- Right-to-left intracardiac shunt, 2:16
- Rigid gas permeable contact lenses, 2:322–323
 fitting philosophy for, 2:323
- Rigid metal plate electrodes, 1:144
- R/IR ratio, in pulse oximetry, 1:472
- Risk analysis, for Gamma Knife, 3:376
- Risk assessment, in radiotherapy treatment planning optimization, 6:38–39
- Risk factors, for implant-related infections, 1:114–115
- Risks, nanoparticle-related, 5:10
- RITA Medical ablation system, 6:373
- RNA. *See also* Ribonucleic acid (RNA)
 DNA, proteins, and, 1:217
 in genome annotation, 1:222
 scanning tunneling microscopy and, 4:517–519
- RNA/DNA staining, 2:411
- Robotic medicine, strain gages in, 6:288
- Robotic surgery systems, 4:525–526
- Robots, arraying, 4:367–369
- Roche Diagnostics artificial pancreas, 5:228
- Roche/Hitachi Modular™ analytic system, 1:20f
- Roentgenographic studies, in diagnosing implant-related infection, 1:116

- Rollators, 4:547–548
 Root-mean-square (rms) value, 3:107–108
 Rotary chair testing, vision-related, 5:139
 Rotate–rotate computed tomography, 2:231–232
 Rotational testing, vision-related, 5:139
 Rotator cuff tendon, reconstruction of, 1:256
 Roughening, of electrodes, 1:152–153
 Routine threshold audiometry, 1:94
 Rubber bellows sensors, 1:175
 Rubber electrodes, 1:148–149
 “Rubber-like behavior,” 1:1, 2
 Rugby, wheelchair, 4:549–550
 Rule-based blood pressure controller, 1:496–497
- Saccades, 5:137–139
 Saccadic intrusions/oscillations, 5:140
 Sacral nerve test stimulator, 1:433f
 Sacral root stimulation, 1:431–432
 Sacrococcygeal teratoma, 4:172–173
 Safar–Elam–Kouwenhoven studies, 2:37
 Safety. *See also* Hospital safety program of anesthesia, 1:28
 of biofeedback instrumentation, 1:168
 in blood collection/processing, 1:456
 contact lens, 2:325–326
 hyperbaric chamber facility, 4:27
 medical gas analyzer, 4:334
 systems perspective on, 6:110
 for wheelchair users, 4:550–552
 Safety appliances, in electrosurgical units, 3:163–164
 Safety committee, hospital, 6:115
 Safety features, in high-dose-rate remote afterloaders, 1:594–596
 Safety monitors, anesthesia machine, 1:39–40
 Safety officer, in a total hospital safety program, 6:116–117
 Safety systems, anesthesia machine, 1:36–37
 Sagittal laser, 5:530
 Sagittal plane, scoliosis and, 6:122
 Salt concentration, skin electrodes and, 1:135
 SAM alignment tool, 1:222
 Sample holder/stage, in near-field imaging, 4:438
 Sample preparation, in neutron activation analysis, 5:44
 Samples. *See also* Sampling
 paired/unpaired, 6:247–248
 in statistical methods, 6:240–241
 Sampling
 in neutron activation analysis, 5:43
 replacement in, 6:248
 Sandwich electrode designs, 1:149
 Sanger DNA sequencing method, 2:428–431
 Sanitation systems, as a hospital problem, 6:113
 Saxtorph, M. H., 1:430
 Scaffold fabrication
 methods of, 3:196t
 in tissue engineering, 3:196–198
 Scaffold matrices, in tissue engineering, 1:367
 Scaffolds
 bioceramic, 1:374–375
 calcium phosphate, 1:261–262
 in engineered tissue, 3:194–200
 fabrication and characterization of, 1:377–379
 ideal, 1:290–291
 natural polymer, 1:367–371
 sol–gel-derived bioactive glass foam as, 1:292–293
 sterilization methods for, 1:380–381
 surface modification of, 1:379
 synthetic polymer, 6:500–501
 in tissue engineering, 1:366
 Scaling effects, in microbioreactors, 4:385–386
 Scan field of view (SFOV), 5:529
 Scanned focus ultrasound systems (SFUs), 4:73
 Scanner lasers, 5:530
 Scanners
 microarray, 4:369
 rectilinear, 1:51; 5:95–96
 Scanning acoustic microscopy (SAM), 1:525–530
 Scanning control, of electronic aids to daily living, 3:214–215
 Scanning electron microscope (SEM), 4:478, 482–486
 Scanning electron microscopy (SEM), 1:356–357, 360–362
 Scanning force microscopy, 4:503–516. *See also* Atomic force microscope (AFM)
 Scanning laser ophthalmoscope (SLO), 5:294
 Scanning probe microscopy (SPM), 4:478
 Scanning tunneling microscopy, 4:516–523
 biological samples and, 4:519–521
 in biology and medicine, 4:517–519
 imaging artifacts and data restoration using, 4:522
 Scanning tunneling microscopy
 interaction, basic physics of, 4:519
 Scanning Ultrasound Reflector Linear Array System (SURLAS), 4:72
 Scanpaths, eye movement, 5:138
 Scar tissue, implantable devices and, 1:104
 Scatter diagram analysis, in separating ventricular fibrillation from tachycardia, 1:79
 Scattered radiation levels, 6:572–573
 Scatter radiation, in computed tomography, 2:256
 Scavenger systems, anesthesia machine, 1:39
 Scenario Editor, with human patient simulators, 1:45
 Schafer–Emerson–Ivy ventilation method, 2:36
 Schedulers, in office automation systems, 5:154
 Scherrer equation, 1:359
 Scholander volumetric technique, 6:525
 Scintillation camera, 1:51, 52. *See also* Anger camera
 Scintillation crystals, characteristics of, 5:410t
 Scintillation detectors, 2:235–236; 5:409f, 411, 511–514
 Scintillators, plastic, 5:482–483, 513–514
 Scleral search coil technique, 3:265–266; 5:144
 Scoliosis, 6:231–232. *See also* Idiopathic scoliosis etiology
 axes and coordinate systems in, 6:124–125
 biomechanics of, 6:122
 classification of, 6:123–124
 conservative treatment of, 6:129–131
 effect on rib cage and back surface shape, 6:123
 functional anatomy of, 6:125
 kinematics of the spine and, 6:126
 measurement of curve magnitude in, 6:125
 progression during growth, 6:129
 ribcage and costovertebral articulations in, 6:125
 surgical planning for, 6:132
 surgical treatment of, 6:131–134
 terminology and morphology of, 6:122–125
 trunk deformity documentation in, 6:126–128
 Screen-film cassette, 6:145
 Screen-film mammography, 4:300–301
 Screen-film systems, 6:138–149
 film contrast and latitude in, 6:143–144
 film in, 6:140–145
 film processing for, 6:145–147
 intensifying screen in, 6:138–140
 latent image formation in, 6:141–142
 optical density in, 6:142
 quality assurance/quality control for, 6:147–148
 reciprocity law and, 6:145
 spectral emission and spectral sensitivity in, 6:144–145
 speed and resolving power in, 6:145
 Secondary dosimeters, calibration of, 5:474–476
 Secondary ion mass spectrometry (SIMS)
 surface analysis, 1:350–351
 Secondary medical physics journals, 4:336–337
 Second-order gradiometer, 1:233
 Security, network, 5:353
 Security issues, related to computer-based patient records, 4:357–358
 Seebeck effect, 6:357–358
 Segmental instrumentation, 6:133
 Segmental Systolic Pressure (SSP) testing, 5:242–243
 Seizure detection, reasons for, 3:73–74
 Seizure response, electroconvulsive therapy and, 3:59
 Seizures, mechanisms of, 3:72–73
 Selection methods, in augmentative and alternative communication systems, 2:205–206

- Selection set, in augmentative and alternative communication systems, 2:204–205
- Self-assembled monolayers (SAMs), 1:347, 410, 412; 4:388
- Self-assembled nanoparticles, 5:3–4
- Self-control conditions, of biofeedback response, 1:178
- Self-etch adhesives, 1:324
- Self-heating
in thermistors, 6:321–322
in thermistors, 6:357
- Self-tuning regulator (STR), 1:492–494
- Semiautomatic blood pressure monitoring, 1:489
- Semiclosed circle system, 1:31–32
components of, 1:32–33
- Semiconductor diodes, 5:476
- Semiconductor materials, used for radiation detection, 5:516–517
- Semiconductor strain gages, 6:283–284
- Senographe full field digital mammography systems, 4:304
- Senoscan full field digital mammography system, 4:304–305
- Sensitivity, in automated arrhythmia analysis, 1:70
- Sensitization, 6:439–440
- Sensor attachment, to microdialysis probes, 4:411
- Sensor design, in solid electrolyte cell oxygen analyzers, 5:205
- Sensorimotor rhythms, 1:243
- Sensor materials, thermocouple, 6:344
- Sensors. *See also* Optical sensors;
Piezoelectric sensors
in arterial tonometry, 6:403–404
biomaterial surfaces of, 1:343
ECG, 1:175
glucose, 1:16–17
heart rate, 1:175
respiration, 1:174–175
SQUID, 1:231–232, 233, 237, 238, 247
temperature, 1:170
- Sensory assessment, for augmentative and alternative communication systems, 2:207–208
- Sensory evoked potentials, 3:235–236
- Sensory testing, for anorectal manometry, 1:64–63
- Separation, in microbioreactors, 4:390–391
- Separation/detection methods, electrophoresis with, 2:106–107
- Separation/purification microbioreactor components, 4:390–391
- Separations-based methods, for microdialysis sample quantitation, 4:409–410
- Sepsis, implant-contiguous, 1:117–118
- Septic shock, management of, 6:168
- Sequence alignment methods. *See also* Multiple sequence alignments
in bioinformatics, 1:217–220
heuristic, 1:219
- Sequential tomotherapy, 5:323
- Serial tomotherapy, 6:397–398
- Serological assays, 4:375–376
- Serological probes, monoclonal antibodies as, 4:601–604
- Serous inflammation, implant-related, 1:116
- Serum processing, 1:460–461
- Serum proteins, biomaterial failure and, 1:280
- Servorecorders, translational, 6:58–59
- Sets, exercise, 1:392
- “7-D” mnemonic, 2:52–53
- Severe angiographic CAD, multivariate techniques to predict, 3:257–258
- Severe angiographic disease, predicting, 3:256
- Sevoflurane, 1:35
- Sexual arousal, deviant, 6:159–161
- Sexual deviance
instruments and measurement of, 6:155
treatment of, 6:161–162
- Sexual dysfunction, biofeedback clinical outcome literature related to, 1:183
- Sexual instrumentation, 6:149–163
for erectile dysfunction, 6:158–159
for female sexual behavior assessment, 6:150–153
for female sexual behavior treatment, 6:153–155
for male human sexual behavior, 6:155–162
- Shape memory alloys (SMAs), 1:1–12
in dental prosthetics, 1:325
manufacturing methods for, 1:5–8
mechanical processing of, 1:5
nickel–titanium, 1:2, 3–5
- Shape memory effect (SME), 1:1
- Shape memory programming, of nickel–titanium shape memory alloys, 1:5–6
- Shape-memory suture needle, 1:1, 2f
- Shape of a tumor dose-response curve (S_{tumor}), 6:44–45
- Shaping, in operant conditioning, 1:167
- Shear deformation, subatomic, 1:1
- Shear modulus (G), of bone, 1:529
- Shear properties, of cartilage and meniscus, 2:69–70
- Shear rate, effect on blood viscosity, 1:501
- Shear stress, in fluids, 1:505
- Sheathed thermocouples, 6:342–343
- Shichiri Group artificial pancreas, 5:227–228
- Shielded rooms, noise attenuation of, 1:235f
- Shielded safe, 1:590
- Shielding, in high-dosage-rate brachytherapy, 1:598–599
- Shielding devices, 5:594–599
- Shivering, thermoregulation and, 1:191
- Shock
assessment of, 6:165–167
cardiogenic, 6:164
etiology of, 6:163–164
future of diagnosis and management of, 6:168
hypovolemic, 6:164
management of, 6:167–168
pathophysiology of, 6:164–165
stages of, 6:164–165
treatment of, 6:163–168
- Shock wave lithotripsy (SWL), 4:258. *See also* Lithotripsy
- Shock waves, generation and focusing of, 4:259–260
- Short-term enzyme electrode glucose sensors, based on conductive polymers, 3:399–400
- Short-time spectral analysis, 3:77
- Short-wave diathermy (SWD), 3:470–472
- Shoulder, stability of, 4:223–224
- Shoulder retractor, 5:591
- Shunt blood flow, 2:42
- Shunt material, in hydrocephalus, 4:12–13
- Shunts
intracardiac, 2:14–15, 16–18
malfunction of, 4:13–14
- Sickle cell disease, 1:500
- Sidestream sampling techniques, 1:479–480
- Sieve electrodes, 3:119–120
- Sigma (σ) electrons, in graphite, 1:297
- Signal(s)
in engineered tissue, 3:198–200
in the scanning electron microscope, 4:483–484
- Signal accuracy, in radiation protection instrumentation, 5:506
- Signal amplification, in hemodynamic monitoring, 4:572–573
- Signal detection
in automated differential counts, 2:413–414
EMG, 3:105–107
- Signal Extraction Technology, in pulse oximetry, 1:472
- Signal filtering, in sleep laboratory, 6:211
- Signal generation, in magnetic resonance imaging, 4:283–285
- Signal interpretation, 1:238–242
- Signal noise, filtering techniques for, 1:202
- Signal preprocessing, in phonocardiography, 5:286–287
- Signal processing
fluorescence detectors and, 4:492–493
in neonatal respiration monitoring, 5:20
in optical sensors, 5:164
ultrasound, 3:10–13
- Signal rectification, 3:107
- Signal space projection (SSP), 1:236, 241
- Signal-to-noise ratio (SNR)
with gradiometers, 1:233–234
in magnetic resonance imaging, 4:288
- Significance, 6:253
- Significant test, 6:246
- Sign test, 6:250
for paired samples, 6:258
- SIG ventricular electrogram analysis, 1:75
- Silicon
in microsensor fabrication, 2:2–3, 4f
wet etching of, 2:3
- Silicon-based electrodes, 3:122–123
- Silicone rubber electrode, 1:148f
- Silicone rubbers, 1:337
- Silicones, as biomaterials, 1:106
- Silicon etching, 1:413
- Silicon membranes, nanoporous, 2:450

- Silicon microprobes, *1:155, 156*
- Silicon micropump, high-performance, *2:504*
- Silicon-on-insulator (SOI) technologies, *2:4–5*
- Silicon piezoresistive devices, *2:1*
- Silicon surface-barrier diode counters, *5:517*
- Silicon tonometer sensors, *6:406t*
- Silver coatings, *1:347*
- Silver–silver chloride electrodes, *1:130–131, 139, 140f*
for electrodermal biofeedback, *1:174*
- SimMan human patient simulator, *1:46f*
- Simplex P bone cement, *1:546*
- Simulated tissues, *5:254–255*
- Simulation
quality assurance of, *5:543–545*
role in pharmacokinetics and pharmacodynamics, *5:275–276*
virtual, *5:455–457*
- Simultaneous vision bifocal lenses, *2:327*
- Single-beam infrared CO₂ analyzer, *1:479f*
- Single breath nitrogen washout, *5:431*
- Single-channel microstimulator, *1:424–427*
- Single-diaphragm tonometer sensors, *6:405–406*
- Single-element tonometer sensors, *6:405, 407*
- Single-element transducers, *4:65*
- Single-energy densitometry, *1:551*
- Single-frequency method, of whole-body bioelectric impedance measurement, *1:212*
- Single nucleotide polymorphisms (SNPs), in genome analysis, *1:221*
- Single-photon absorptiometry, *1:551*
- Single photon emission computed tomography (SPECT), *2:277–284; 5:108. See also SPECT entries*
clinical applications in, *2:282–283*
data acquisition and processing in, *2:278–280*
instrumentation for, *2:280–281*
- Single-point laser probe, *2:379*
validation of, *2:381*
- Single-trial analysis, of evoked potentials, *3:241–242*
- Single voxel (SV) MRS, *5:79*
- Sintering aids, *1:357, 358–359, 362*
- SIS/PLGA scaffolds, *1:379*
- Size exclusion chromatography, *2:104*
- Skeletal ligaments, *4:243–245*
- Skeletal muscle diseases, electron microscopic diagnosis of, *4:485–486*
- Skeletal muscle fiber types, *3:349t*
- Skiing, adaptive, *4:550*
- Skin. *See also Electrode-skin interface*
anisotropy of, *6:205*
bioelectrodes and, *1:131–137*
blood gas diffusion through, *1:476–477*
effect of UV exposure on, *6:475–476*
electrical properties of, *1:131–132*
engineered, *3:204–205*
functions of, *6:169*
hair follicles in, *1:133*
microstructure of, *6:203*
nanoparticle penetration of, *5:9–10*
nonlinear elasticity of, *6:204–205*
normal properties of, *6:169–170*
parallel capacitance of, *1:132*
parallel resistance of, *1:132–133*
potential motion artifact of, *1:133–134*
structure of, *1:131*
viscoelasticity of, *6:204*
- Skin abrasion, for electrodes, *1:136–137*
- Skin biomechanics, *6:202–208*
collagen in, *6:203–204*
elastin in, *6:204*
ground substance in, *6:204*
- Skin cancer, associated with ultraviolet radiation, *6:475–476*
- Skin conductance activity (SCA), *1:173*
- Skin diseases
cryosurgical treatment of, *2:373*
laser Doppler flowmetry for, *2:382–383*
- Skin disease therapy, using ultraviolet radiation, *6:482–485*
- Skin equivalent, living, *6:191–192*
- Skin grafts, hyperbaric medicine and, *4:25*
- Skin impedance, *1:131–137*
diurnal and seasonal variations in, *1:133*
- Skin loss, current treatment of, *6:189–190*
- Skin mechanical properties, measurement of, *6:205*
- Skin microcirculation, techniques for assessing, *2:379t*
- Skin preparation
for EGG, *3:87–88*
for electrode application, *1:136–137*
- Skin Rasp, *1:136*
- Skin replacement, technologies for, *6:190–198*
- Skin resistance activity (SRA), *1:173*
- Skin stripping, for electrodes, *1:136–137*
- Skin substitutes
available, *6:172–178*
for burns, *6:169–179*
categories of, *6:172t*
permanent, *6:175–178*
role of, *6:172*
temporary, *6:173–175*
- Skin temperature, *6:320*
electrodes and, *1:135*
- Skin temperature measurement, in incubators, *4:154–155*
- Skin temperature monitoring, *6:316–317*
- Skin tissue engineering, *6:179–202. See also Skin replacement*
extracellular matrix analogs in, *6:184–187*
tissue triad structure/function and, *6:187–190*
- Skin ulcer, laser Doppler flowmetry for, *2:382–383*
- Skin ulceration, diabetic, *6:206–207*
- Slab-gel DNA sequencing, *2:431–432*
- Sleep acquisition/analysis/management system, prototypical, *6:219–221*
- Sleep apnea, *6:220. See also Obstructive sleep apnea/hypopnea syndrome (OSAHS)*
obstructive, *6:212*
- Sleep-disordered breathing (SDB), *6:211–212*
- Sleep-disordered breathing diagnosis, thermistors in, *6:336–337*
- Sleepiness, daytime, *6:272*
- Sleep laboratory, *6:208–213*
future trends in, *6:212–213*
polysomnographic recording in, *6:209–211*
- Sleep microstructure, computer analysis of, *6:219*
- Sleep stages, *6:211, 214–215*
computer identification of, *6:215–219*
- Sleep studies, computer analysis of, *6:213–224. See also Polysomnography*
- Sleep study summary, *6:222f*
- Slide diagram, *1:20f*
- Slip-cast all-ceramic materials, *1:327*
- Slowly penetrating interfascicular electrode (SPINE), *3:124*
- Slow-neutron therapy, *6:7–8*
- Slow wave coupling, EGG, *3:90*
- Small intestine submucosa (SIS), in tissue engineering, *1:371*
- Small-field optokinetic response, *5:140*
- SmartDose, *2:504*
- SMARTeR nitinol stent, *1:272f*
- Smart image enhancement/restoration algorithms, *6:351*
- Smart image processing, *6:351*
- Smart polymers, *1:340*
- Smokers, low dose CT screening for, *2:263–264*
- Smoothing transformation, *1:395–396*
- Smooth muscles, local blood flow and, *1:190*
- Smooth pursuit eye movements, *5:138–139*
- Snap fasteners, with electrodes, *1:139, 148–149*
- Snap Gauge, *6:156–157*
- Soap bubble calibration, *5:374*
- Societies, medical engineering, *4:311–321*
- Society for Biological Engineering of the American Institute of Chemical Engineers, *4:321*
- Society for Biomaterials, *4:321*
- Society for Biomolecular Screening, *4:321*
- Society for Modeling and Simulation International, *4:321*
- Society of Interventional Radiology, *4:321*
- Society of Nuclear Medicine (SNM), *4:321; 5:122*
- Society of Rheology, *4:321*
- Sodium borocaptate (BSH), *1:572*
- Sodium iodide crystal, in nuclear medicine, *1:53–54*
- Sodium ions, bioactive glasses and, *1:286*
- Sodium nitroprusside (SNP), *1:491, 493–499*
- Soft contact lenses
design of, *2:323*
fitting, *2:323*
- Soft-read workstations, in computed tomography, *2:239*
- Soft tissue engineering, *1:294*
- Soft tissue fixation, *1:256*
- Soft tissue grafts, with periosteum, *2:72–73*

- Software
 in computer-based instrument systems, 2:314–316
 for computerized exercise feedback, 1:399–400
 DNA sequence analysis, 2:432t
 medical image display, 5:355–356
 networking, 5:352–353
 phonocardiography, 5:289
 sleep study, 6:220
 statistical, 6:263–264
 text-to-speech, 1:446–447
 virtual simulation, 5:532
- Solar simulators, 6:477–478
- Solar ultraviolet radiation, 6:476
- Sol-gel-derived bioactive glasses, 1:287–289
 biological responses to, 1:293–294
- Sol-gel-derived bioactive glass foams, 1:292–293
- Solid ankle, cushion heel (SACH) prosthetic foot, 4:552–553
- Solid conductive adhesive electrodes, 1:141
- Solid electrolyte cell, oxygen analyzers, 5:204–206
- Solid gels, for electrodes, 1:141
- Solid lipid nanoparticles (SLN), 2:484–485
- Solid phantom, 5:263–264
- Solids, surface energy of, 1:343
- Solid-state detectors, 5:476–477
- Solid-state electrosurgical generators, 3:162–163
- Solid-state gas sensors, 4:332–333
- Solid-state probe, in anorectal manometry, 1:63
- Solid-state radiation detectors, 5:514–518
- Solid tumors, treatment of, 4:606
- Sols, 1:288
- Soluble factor delivery, in tissue engineering, 6:389
- Solvent drying, nanoparticle fabrication via, 5:3
- Somatic afferent feedback, in control of joint movement, 3:128
- Somatosensory evoked fields (SEFs), biomagnetic measurements and, 1:242–243
- Sonic Guide, 1:449–450
- Sound
 as a hospital problem, 6:111
 quantifying, 6:454
- Sound analysis, 4:280–281
- Sound measurement, in neonatal respiratory monitoring, 5:15
- Sound transducers, 4:278–279
- Sound waves, 6:453–454
- Source lens, in the transmission electron microscope, 4:481
- Source lumen eccentricity, 1:610
- Source stepping, 1:610–611
- Source-to-skin distance, in X-ray equipment, 6:567
- Space, as a hospital problem, 6:113
- Spark gap oscillators, 3:161–162
- Spatial encoding, in magnetic resonance imaging, 4:285–286
- Spatial filtering, 3:106
- Spatial frequency processing, 5:341
- Spatial linearity, Anger camera, 1:58–56, 60
- Spatial resolution, Anger camera, 1:57f, 60
- Spatial resolution measurements, in X-ray systems, 6:572
- Spearman's rank correlation test, 6:259
- Specific adsorption, 1:123–124
- Specificity, in automated arrhythmia analysis, 1:70
- Specimen collection, in automated analytical methods, 1:19
- Specimen manipulation system
 in the scanning electron microscope, 4:483
 in the transmission electron microscope, 4:481
- SPECT/CT hybrid imagers, 5:104
- SPECT imaging, 5:100–101
- Spectral edge frequency, neurological monitor, 5:38
- Spectral emission, in screen-film systems, 6:144–145
- Spectral imaging microscopy, 2:94–95
- Spectral measurements, 4:498–500
- Spectral sensitivity, in screen-film systems, 6:144–145
- Spectrofluorimeters, 3:347, 343–344
- Spectrogram-based neurological monitor, 5:36–37
- Spectrometry, radiation detection and, 5:93–94. *See also* Atomic absorption spectrometry; Flame atomic emission spectrometry; Spectroscopy
- Spectrophotometry, 1:469
- Spectroradiometry, 6:480
- Spectroscopy. *See also* Impedance spectroscopy; Mass spectroscopy; Near-field spectroscopy; Nuclear magnetic resonance (NMR) spectroscopy
 evanescent-wave, 5:161–162
 fiber optics in, 3:311–313
 fluorescence, 4:489–490
 fluorescence correlation, 2:97–98
 force, 4:505–506
 infrared/optical, 4:327–329
 Raman laser, 4:332
- Speech
 assessment of, 2:215
 cueing of, 2:218
 esophageal, 4:230
- Speech analysis programs, 2:213–214
- Speech audiometry, 1:96–97
 suprathreshold, 1:97
- Speech awareness threshold (SAT), 1:97
- Speech detection threshold (SDT), 1:97
- Speech disorders, 2:210–211. *See also* Speech impairments
 new directions in, 2:224
- Speech feedback, 2:218
- Speech impairments, assistive devices for, 2:222–223. *See also* Speech disorders
- Speech intervention, 2:218
- Speech movement indicators, 2:218
- Speech recognition threshold (SRT), 1:96–97
- Speech therapy, thermistors in, 6:336
- Speed of sound (SOS), 1:554
 for bone measurement, 1:555
- Spermicides, vaginal, 2:340
- Sphenomandibular ligament, 6:417
- Spherical sound waves, 6:453–454
- Sphygmomanometer, 1:486
- Spinal cord stimulation (SCS), 1:431; 3:27; 6:224–229. *See also* Spinal cord stimulator
 equipment for, 6:225–226
 future uses of, 6:229
 mechanisms of, 6:227
 outcomes of, 6:228–229
 patient selection for, 6:226–227
 studies of, 6:228–229
- Spinal cord stimulator, implantation
 technique for, 6:227–228
- Spinal disorders, 6:231–234
 treatment options for, 6:238t
- Spinal fusion, 6:235–237
 techniques for, 3:575–579
- Spinal implants, 6:229–240
 terminology related to, 6:238–239
 treatment options and medical devices related to, 6:234–238
- Spinal instability
 biomechanics of, 3:558–568
 role of environmental factors in, 3:558–562
- Spinal instrumentation
 anterior and posterior, 3:579–581
 testing of, 6:132–133
 types of, 6:132
- Spinal ligaments, 4:245
- Spinal motion changes
 due to degeneration-trauma, 3:562–566
 due to surgical procedures, 3:566–568
- Spinal motion segment, 6:125
- Spinal nerves, peripheral distribution of, 6:448f
- Spinal orthotic devices, 6:89t
- Spinal stenosis, 6:233–234
- Spine
 active stabilizers in, 3:553–554
 coupling of rotations in, 6:123
 dynamic stabilizers in, 3:553
 human, 6:230
 passive elements of, 3:553
 scoliosis and, 6:125
 stability of, 4:222–223
- Spine anatomy, 3:547–552
- Spine biomechanics, 3:547–598. *See also* Spine anatomy
 of the normal spine, 3:552–558
- Spin echoes, in magnetic resonance imaging, 4:288
- Spine deformities, implants for, 6:234
- Spine stabilization procedures
 animal models of, 3:581–582
 biomechanics of, 3:568–591
 cervical spine, 3:573–579
 finite element models of, 3:585
 human clinical models of, 3:582–585
 implant-bone interface in, 3:569–572
 nonfusion treatment alternatives to, 3:585
 recent and future initiatives related to, 3:588–591

- Spin-lattice relaxation time, 4:287
- Spiral electrodes, 1:157
- Spirometers, 2:39
 anesthesia machine, 1:39
 commercially available, 5:372
 volume-displacement, 5:372
- Spirometry, 5:436–437
- Splints, 6:81f
- Spondylolisthesis, 6:234
- Sponge, contraceptive, 2:341–342
- Spontaneous nystagmus, 5:140
- Spontaneous otoacoustic emissions (SOAEs), 1:101
- Sports medicine, strain gages in, 6:288
- Sports mobility aids, 4:548–550
- Spray cryosurgery technique, 2:366
- Spreadsheets, in office automation systems, 5:153
- Square wavelet, 1:75
- Squeeze sphincter pressure, 1:64
- SQUID Array for Reproductive Assessment (SARA) system, 1:247
- SQUID electronics, 1:232. *See also* Superconducting quantum interference device (SQUID)
- SQUID gradiometers, 1:235, 238
- SQUID magnetometers, 1:234, 235
- SQUID sensors, 1:231–232, 233, 237, 238, 247. *See also* MicroSQUID systems
- Squire, J. R., 1:469
- Stable cells, tissue regeneration and, 1:109
- Staffing requirements, for biomedical equipment maintenance, 3:228
- Staining, in medical microbiology, 4:375
- Stainless-plate electrode, 1:138
- Stainless steel(s)
 corrosion resistance of, 1:311–312
 as biomaterials, 1:105
 in dental prosthetics, 1:325–326
- Stainless steel alloys, as biomaterials, 1:270, 271
- Stamping, 1:409, 411, 412f
- Stamp test, for erectile dysfunction, 6:157
- Standalone instruments, in biofeedback, 1:170
- Standard deviation (σ , SD), 6:253, 254
- Standard error (SE), estimating, 6:253
- Standard Erythematous Dose (SED), 6:478
- Standard exercise test, ACC/AHA guidelines for prognostic use of, 3:259–262
- Standard hydrogen electrode (SHE), 1:122
- Standards
 audiometric, 1:93
 for biomaterials, 1:281–282
 biosignal monitoring electrode, 1:158–160
 for blood collection/processing, 1:455–456
 for blood pressure measurement devices, 1:489–490
 electrode, 1:158–162
 gas system, 3:380–381
 medical device conformance with, 2:142
 for polymeric biomaterials, 1:334t
 for protection against radiation, 2:160–162
 vacuum system, 3:383
- Standard temperature and pressure, dry (STPD), 2:14, 15
- Staphylococcus aureus*, 1:114, 115, 320
- Staphylococcus epidermidis*, 1:114
- State laser regulations, 2:158t
- States, radiation regulation by, 2:156
- Static, 5:68–69
 use in medical diagnosis and therapy, 5:71
- Static acoustic immittance, 1:99–100
- Static calibration, in blood pressure monitoring, 4:570
- Static single-head gamma cameras, 5:99
- Static tracers, 6:429–430
- Stationary method, for anorectal manometry, 1:63
- Station pull-through technique, 1:63
- Statistical analysis, in bioinformatics, 1:223–224
- Statistical computing, 6:263–264
- Statistical inference, 6:246–248
- Statistical inference tests, uses for, 6:247t
- Statistical methods, 6:240–264
 data sample and experimental design in, 6:240–241
 descriptive statistics, 6:241–243
 multivariate methods, 6:261–262
 probability, random variables, and probability distributions in, 6:243–261
- Statistical Package for the Social Sciences (SPSS), 6:263
- Statistical test for lifetime variables, 6:262
- Statistical variation, in pharmacokinetics and pharmacodynamics, 5:273–275
- Statistics books/reports, 4:349
- Stator, in X-ray tubes, 6:605
- STEAM protocol, 5:80–81
- Steel, implant-grade, 1:311–312. *See also* Stainless steel(s)
- Steep dose gradient, diffusion of monomer in, 5:495
- Steganography, 4:359–360
- Stem cells
 engineered tissue and, 3:193–194
 for organ function loss, 6:182
 in tissue engineering, 6:382–383
- Stenosis, spinal, 6:233–234
- Stent-grafts, endovascular, 6:495–496
- Stent perturbation, 1:612–613
- Stent placement, in percutaneous transluminal coronary angioplasty, 4:542
- Stents
 biomaterials for, 1:278
 CYPHER, 1:278
 in microsurgery, 4:532–533
 nickel–titanium shape memory alloy, 1:9
 nitinol, 1:272
- Step-down sensitization, in hyperthermia, 4:49
- Step growth polymerization, 1:330
- Stepping-source remote afterloader, 1:590
- Stereo fundus photos, 5:293
- Stereotactic biopsy, 6:269
- Stereotactic body radiation therapy (SBRT), 5:594
- Stereotactic Body Frame, 5:594
- Stereotactic breast biopsy mammography, 4:302–303
- Stereotactic craniotomy, 6:269
- Stereotactic frame-based procedures, 6:268–271
- Stereotactic imaging/localization, 5:577–578
- Stereotactic radiosurgery, 4:542; 5:574–585
 computed tomography simulation for, 2:275
 gels in, 5:490–491
- Stereotactic radiotherapy, repeat-fixation, 5:491
- Stereotactic surgery, 6:265–273. *See also* Stereotactic radiosurgery
 applications of, 6:271
 future directions in, 6:271
 principles of, 6:266–267
- Stereotactic surgery planning, computed tomography in, 2:244
- Stereotactic targeting systems, early, 5:576
- Stereotactic technique, frameless, 6:269–270
- Stereotaxis
 based on CT, 6:267
 based on MRI, 6:267–268
 functional, 6:268
- Sterilants, chemical, 6:279t
- Sterility assurance level (SAL), 6:274
- Sterilization. *See also* Sterilization methods
 bilateral tubal, 2:346–347
 of biologic scaffold materials, 6:273–282
 ethylene oxide, 6:276–277
 gamma, 6:277–278
 heat, 6:275–276
 ionizing radiation, 6:277–278
 of microbioreactors, 4:388
 moist heat, 6:275–276
 of scaffolds, 1:380–381
 versus disinfection, 6:274
- Sterilization methods
 alternative, 6:278–280
 for implant-related infection, 1:117
 summary of, 6:279t
 validation of, 6:274–275
- Sterilization standards, 6:274t
- Steroid-eluting electrode, 1:154
- Stethoscope, 4:278
 electronic, 5:288–289
- Stewart, G. N., 2:25
- Stewart–Hamilton method, 2:25
- Stiffness matrix, 6:126
- Stimulated muscle force assessment, 6:66–70
- Stimulation algorithm, for visual prostheses, 6:541
- Stimulation electrodes
 implanted, 1:121
 standards for, 1:160–161
- Stimulation threshold, 1:130
- Stimulation thresholds/limits, in visual prostheses, 6:540–541

- Stimulator case, in visual prostheses, 6:538–539
- Stimuli-responsive hydrogels, 5:391
- Stimulus frequency otoacoustic emissions (SFOAEs), 1:101
- Stimulus, response and, 1:166, 167
- Stimulus words, in audiology, 1:96–97
- Stokes, George Gabriel, 1:469
- Stokes–Adams–Morgagni syndrome, 2:47
- Stomach
 - effects of parenteral nutrition on, 5:130
 - electrophysiology of, 3:84–85
- Stone expulsion, improving, 4:264
- Stone fragmentation
 - improving, 4:264
 - in lithotripsy, 4:260–261
- Stone localization, during lithotripsy, 4:260
- Storage Access Networks (SAN), 5:350
- Stow, Richard, 2:110f
- Strain, relationship to stress, 6:282–283
- Strain energy equations, for arterial elasticity, 1:88
- Strain gage displacement sensors, in neonatal respiratory monitoring, 5:16
- Strain gages, 6:282–290. *See also* Strain gage entries
 - applications for, 6:286–288
 - signal preparation and amplification in, 6:285–286
 - types of, 6:283–285
- Strain gauge, 4:578. *See also* Strain gage entries
- Strain Gauge measurement, 6:156–157
- Strain gauge plethysmography, 5:237–238
- Strain tensor, 1:88
- Strand displacement amplification (SDA), 4:378
- Strategic planning, hospital technology changes and, 6:117
- Stratum corneum, 1:131; 6:169
 - electrical properties of, 1:131–132
- Stress
 - biofeedback training and, 1:167–168
 - relationship to strain, 6:282–283
- Stress concentrations, of bone, 1:531
- Stress echo echocardiographic examination, 3:17–18
- Stress electrocardiography, 3:47–48
- Stress induced martensite (SIM), 1:1, 4
- Stress reduction, biofeedback clinical outcome literature related to, 1:178
- Stress–strain relations, for arterial walls, 1:88–89
- Stress testing
 - cardiopulmonary, 5:440–441
 - of skin electrodes, 1:134
- Stroke, 2:52–53
 - biofeedback clinical outcome literature related to, 1:182–183
- Stroke volume (SV), 2:12
- Structural tissue, engineered, 3:203–204
- Structure identification, using a computed tomography simulator, 2:270
- Stylomandibular ligament, 6:417
- Subatomic shear deformation, 1:1
- Subcellular biology/chemotaxis, use of multiple laminar streams to study, 4:394–395
- Subcutaneous artificial pancreas, 5:226
- Subcutaneous layer, 1:131
- Sublingual capnometry, 1:481–482
- Subretinal implants, 6:534
- Subretinal stimulating arrays, 1:156
- Substrates, in enzymatic reactions, 2:194–195
- Subzero effects, 1:192
- Suction electrodes, 1:138–139
- Sudden cardiac death, 1:69
- Sudden injection method, 2:25
- Sunbed lamps, 6:485
- Sunbeds, 6:485–486
- Superconducting accelerators, 6:12
- Superconducting quantum interference device (SQUID), 1:230–231. *See also* SQUID entries
- Supercritical fluid chromatography, 2:104
- Superelastic behavior, 1:1
- “Superelasticity,” 1:1
- Superficial second degree burns, 6:170
- Superficial units, for radiation therapy, 6:582–583
- Superficial zone, of articular cartilage, 2:64–65
- Superheated drop detectors (SDD), 5:518
- Support, in speech intervention, 2:218
- Supporting media-based enhanced resolution techniques, in electrophoresis, 3:135–137
- Suppurative infection, implant-related, 1:116
- Supramolecular aggregates
 - for drug delivery, 2:460–464
 - lecithin organogel, 2:463–464
 - microemulsions, 2:460–463
 - surfactants, 2:460
- Suprathreshold speech audiometry, 1:97
- Supraventricular tachycardia (SVT), 1:69
- Surface analysis techniques, 1:348–351
- Surface-barrier detectors, 5:517
- Surface chemistry, 1:343–344
 - in microbioreactors, 4:388
- Surface electrodes
 - EMG, 1:169; 3:102–103
 - in functional electrical stimulation, 3:352–353
- Surface electromyographic (EMG) activity, biofeedback training and, 1:168
- Surface energy, 1:343, 344–345
- Surface engineered metal-on-metal hip joint bearings, 3:522
- Surface force apparatus (SFA), 1:351
- Surface force measurements, 1:351
- Surface mechanical properties, 1:343–344
- Surface micromachining, 2:3–5, 6f
- Surface modeling, of joints, 4:214–215
- Surface modification, of scaffolds, 1:379
- Surface plasmon resonance, 5:162
 - for characterizing surfaces, 1:352
- Surface properties, of porous biomaterials, 5:393. *See also* Biomaterial surfaces; Biosurface engineering
- Surface replacement hip joint bearings, 3:522
- Surface topology, 1:343, 344
- Surfactants, 1:292–293
 - characteristics of, 2:460
- Surgery. *See also* Surgical procedures; Surgical techniques
 - cardiac, 3:459–461
 - eye, 4:530–532
 - fetal, 4:533
 - minimally invasive, 4:525
 - stereotactic, 6:265–273
- Surgical applications, fiber optics in, 3:313–314
- Surgical Information Systems (SIS), 1:45
- Surgical maneuvers, analytical simulation of, 6:133–134
- Surgical microscope
 - history of, 4:523
 - use of, 4:523–526
- Surgical navigation, 4:538–539
- Surgical planning, computed tomography in, 2:242
- Surgical procedures
 - anesthesia in, 1:42–43
 - spinal motion changes due to, 3:566–568
 - using electrosurgery, 3:164–166
- Surgical techniques, intrauterine, 4:171–181
- Surgical technology, minimally invasive, 4:535–544
- Surgical treatment of scoliosis, 6:131–134
- Suspended animation, 1:192
- Suspended cells, impedance analysis of, 4:135–137
- Suspension method, of biomaterials testing, 1:355–356, 357–358
- Sustained drug delivery, 2:439
- Suture material
 - nylons as, 1:338
 - polypropylene as, 1:335
- Sutures, polymers in, 1:274–276
- Swan–Ganz catheters, 2:26, 30
- Sweating, thermoregulation and, 1:191
- Sweat test, cystic fibrosis, 2:384–388
- Swelling properties, of cartilage and meniscus, 2:70
- Switched capacitor interfaces, 2:7
- Symptoms, of implant-related infections, 1:116
- Synchronized intermittent mandatory ventilation mode (SIMV), 6:510
- Syncope, 1:16
- Syndiotactic polypropylene, 1:335
- Synovial fluid, 6:417
- Synovial joints, 2:63; 4:199, 200–201f
- Synovial-like tissue, from prosthetic motion, 1:111
- Synteny blocks/groups, 1:223
- Synthetic absorbable polymers, as biomaterials, 1:107
- Synthetic aperture magnetometry (SAM), 1:241
- Synthetic bone grafts, 6:236
- Synthetic carbons, 1:299–300
- Synthetic hydroxyapatite (HA), 1:284, 285

- Synthetic materials, in tissue engineering, 6:385–386
- Synthetic polymers, 1:329–330; 5:388–389
 biodegradable, 1:339–340
 in engineered tissue, 3:195–196
 in tissue engineering, 1:371–374
- Synthetic tissue implants, 1:355
- Synthetic tissue transplants, 1:366
- Syringe pump drug infusion systems, 2:499–500
- Syringe spirometer calibration procedure, 5:372–373
- Sysmex XE-2100 Hematology System, 2:418–419
- System architecture, of computer-based patient record systems, 4:355–356
- Systematic Analysis of Language Transcripts (SALT), 2:216
- Systemic blood flow (SBF), 2:16–18
- Systemic circulation, 4:567
- Systemic electroanalgesia, 3:24–34
 history of, 3:24–26
- Systemic errors, in computed tomography, 2:257
- Systemic hyperthermia, 4:42–62
 biological effects of, 4:46–51
 cancer and, 4:56–59
 chemotherapy and, 4:56–58
 clinical experience with, 4:56–59
 clinical toxicities of, 4:51
 clinical trials of, 4:54–55t
 disease treatment using, 4:58–59
 future of, 4:59
 historical background of, 4:42
 induction of, 4:45
 metabolic therapy and, 4:58
 pain relief in, 4:58
 physics of, 4:42–45
 radiotherapy and, 4:58
 temperature measurement in, 4:52–56
 temperature probes for, 4:53t
 thermal dose in, 4:51–52
 websites related to, 4:60
- Systemic hyperthermia centers, clinical academic/ regional, 4:43t
- Systemic perfusion, real-time noninvasive measures of, 6:167
- Systemic sepsis, as a complication of parenteral nutrition, 5:128–129
- Systemic vascular resistance (SVR), 2:19
- Systems analysis, hospital, 6:114
- Systems models. *See also* Physiological systems modeling
 reduction versus simplification of, 5:299–301
 types of, 5:301–303
- Systolic pressure, 1:485
 segmental, 5:242–243
- Tab solid adhesive electrodes, 1:141
- Tachycardia(s), 6:370
 pulseless ventricular, 2:45–46
 ventricular, 1:76, 77, 78, 79
- Tachycardia detection interval (TDI), 1:72
- Tactaid, 6:293
- Tacticity, of thermoplastics, 1:332
- Tactile physiology, 6:291–293
 Tactile shape displays, 6:295
 Tactile stimulation, 6:291–302
 applications of, 6:291
 future directions of, 6:298–299
 techniques and applications of, 6:293–297
- Tactile stimulators, mechanical, 6:293–295
- Tactors, 6:293
- Tafel behavior, 1:309
- Taggart, W. H., 1:325
- Talking computers, 1:445, 446–447
- Talking Signs, 1:451, 452f
- Tangent screen exam, 6:529
- Tanning, 6:485–486
- Target concentration, 1:48–49
- Target controlled infusion (TCI) systems, 2:506
- Target discovery, use of microarrays in, 4:370
- Targeted drug delivery, 2:438–439
- Targeted reconstructions, in computed tomography, 2:251–252
- Target-tip electrode, 1:154
- Target volumes, in three-dimensional conformal radiotherapy, 6:32–33
- Task-dependent morphological finite element models, 4:219–220
- TAXUS Express² Paclitaxel-eluting stent, 1:278
- T cell administration, use of immune adjuvant with, 4:113
- T-cell depleting agents, 4:274
- T cells
 effector, 4:112–113, 113–114, 116–117
 tumor-reactive pre-effector, 4:111–112
- T cell subsets, antitumor reactivity of, 4:114–115
- t* distribution, 6:252t
- Teaching software, for communication disorders, 2:212
- Technetium-99m, 1:52
- Technologies
 chromatography, 2:103–108
 prosthetic heart valve, 3:428–430
- Technology management, medical device, 6:117
- Teeth, 1:523
 classification of, 6:411–412
 elastic properties of, 1:526
 stiffness of, 1:525t
 structure of, 6:413
- Teflon, in cardiovascular applications, 1:276
- Teleconferencing, in office automation systems, 5:158–159
- Telemedicine, 5:295–296
 in anesthesia, 1:47
 growth of, 3:542
 screening instrument, 5:295–296
 trends in, 6:311
- Telemonitoring, ECG signal, 3:51
- Teleradiology, 6:302–311
 combined with grid computing and enterprise level PACS, 6:309
 combined with Picture Archiving and Communication System, 6:303–304, 308–309
 components of, 6:304–308
 important issues in, 6:309–310
 models of, 6:303t
 need for, 6:304
 operation models for, 6:308–309
 trends in, 6:311
 web-based, 6:308
- Teleradiology system, user friendliness of, 6:306
- Teleradiology technologies, 6:309
- Telerobotic technologies, 6:291
- Telethermometry, dynamic area, 6:352
- TEMED catalyst, 1:291
- Temperature. *See also* Heat; Skin temperature; Thermal entries; Thermo- entries
 baseline, 2:29
 capacitive microsensors and, 2:5–6
 defined, 6:312
 glass-transition, 1:331
 measurement of, 6:341
 melting, 1:331
 viscosity and, 1:501
- Temperature amplifiers, 1:170
- Temperature biofeedback instrumentation, 1:170–171
- Temperature-change devices, 3:466–475
- Temperature coefficient of resistance, 6:322
- Temperature dependence, during polymer gel irradiation, 5:494
- Temperature displays, 1:171
- Temperature gradients, 6:343, 345
- Temperature increases, metabolic effects of, 4:48
- Temperature measurement(s)
 electronics in thermistors, 6:327–333
 errors in, 6:343–344
 sexual arousal and, 6:161
 in systemic hyperthermia, 4:52–56
- Temperature measurement electronics, low cost, 6:329
- Temperature measuring systems
 specifications of, 6:356
 types of, 6:356–361
- Temperature monitoring, 6:311–320
 chemical phase changes, 6:315–316
 clinical applications of, 6:316–319
 definitions related to, 6:312
 electrical property changes, 6:313–316
 emitted thermal radiation changes, 6:316
 esophagus, 6:318
 intravascular, 6:318
 nasopharynx, 6:317
 neonatal, 5:25–26
 oral cavity, 6:317
 rectum, 6:378
 sites of, 6:316
 skin, 6:316–317
 thermometer types and, 6:313
 tympanic membrane, 6:317
 urinary bladder, 6:318
- Temperature probes, for systemic hyperthermia, 4:53t
- Temperature regulation, body, 6:318–319
- Temperature scales, 6:312–313

- Temperature-sensitive hydrogels, 5:391
- Temperature sensors, 1:170
integrated-circuit, 4:157–162
in neonatal respiratory monitoring, 5:14–15
- Temperature-to-frequency converters, in temperature measurement electronics, 6:328–329
- Temperature-to-time interval converters, in temperature measurement electronics, 6:328–329
- Template matching by CWA, 1:76
- Template-based algorithms, in arrhythmia analysis, 1:74–76
- Templates, in cellular patterning, 1:409–410
- Temporal-spectral imaging, in peripheral vascular noninvasive measurements, 5:247–249
- Temporary external pacemakers, 5:218
- Temporary skin substitutes, 6:173–175
ideal properties of, 6:173t
- Temporomandibular joint (TMJ), 6:410, 415–417
- Temporomandibular (TM) ligament, 6:417
- Tendons, 4:201, 245–246
collagen in, 4:241–242
components of, 4:241–243
elastic fibers in, 4:242–243
failure mechanisms of, 4:247–248
fiber–matrix interactions in, 4:243
measuring the properties of, 4:246–247
properties of, 4:241–252
proteoglycans in, 4:242
repair of, 4:248
- Tennis, wheelchair, 4:550
- TENS electrode system, 1:146, 148
- Tensile behavior, of arterial walls, 1:87
- Tensile modulus, of cartilage and meniscus, 2:66–67
- Tensile properties, of cartilage and meniscus, 2:66–67
- Tension
biofeedback clinical outcome literature related to, 1:178–179
biofeedback procedures for, 1:176–178
biofeedback training and, 1:167–168
- Tension headache, biofeedback clinical outcome literature related to, 1:178–179
- Tenth value layer (TVL), 6:592
- Terahertz radiation, use in medical diagnosis and therapy, 5:70–71
- Teratoma, sacrococcygeal, 4:172–173
- Tergitol TMN10, 1:291
- Terminal arrhythmia, 2:47
- Terminal units (station inlets)
in gas systems, 3:380
for vacuum systems, 3:383
- Tertiary structure, of proteins, 1:344
- Testicular shield, 5:597
- Testing
of heart valve prostheses, 3:430–435
point of care, 1:22–23
of tissue substitutes, 5:258–262
- Tetraethoxyl orthosilicate (TEOS), 1:292
- Tetramethylsilane (TMS), 5:77–78
- Tetrapolar electrode configuration, 1:198–199
- Text-to-speech software, 1:446–447
- Thallium, in nuclear medicine, 1:53–54, 55
- Theoretical analysis, in joint mechanics, 4:213–214
- Theoretical dosimetry, 1:613–614
- Theranostics, use of microarrays in, 4:371
- Therapeutic agents, monoclonal antibodies as, 4:604–605
- Therapeutical applications, microbioreactors for, 4:394
- Therapeutic biotelemetry microsystems, 1:424
- “Therapeutic” pressure, in continuous positive airway pressure, 2:333
- Therapy
electroconvulsive, 3:53–62
use of nonionizing radiation in, 5:70–71
- Thermal ablation, 6:362–367
- Thermal anemometry, 3:329–330
- Thermal burns, hyperbaric medicine and, 4:26
- Thermal conduction, 6:343
- Thermal conductivity (k), 1:193
measurement of, 1:192–196
- Thermal conductivity detector (TCD), 4:325, 326
- Thermal damage, cellular, 4:48
- Thermal deposition method, of biomaterials testing, 1:356–357, 358–360
- Thermal detectors, cooled versus uncooled, 6:349–350
- Thermal diffusivity (α), 1:193
measurement of, 1:192–196
- Thermal dilution method, fundamental equations for, 2:24–25
- Thermal dose
in interstitial hyperthermia, 4:34
in systemic hyperthermia, 4:51–52
- Thermal-electric analogue, 6:348
- Thermal energy, in living systems, 1:188
- Thermal energy balance, in human body, 1:191
- Thermal environment, neutral, 4:148–150
- Thermal environmental conditions, 6:348
- Thermal expansion, 6:320
- Thermal imagers, development of, 6:350–351
- Thermal infrared (TIR) region, 6:347
- Thermal injury, 1:192
- Thermally induced martensite (TIM), 1:1
- Thermally induced phase separation (TIPS), 5:400
- Thermal models, human, 6:347–348
- Thermal probes, 1:195–196
- Thermal properties, defined, 1:193
- Thermal shunting, 6:343
- Thermal systems modeling, 5:322–323
- Thermal therapy
interstitial ultrasound, 6:365
laser interstitial, 6:365–366
- Thermal time constant (TC), in thermistors, 6:322
- Thermal tissue injury, principles of, 6:368–369
- Thermal tolerance, 4:48–49
- Thermatomes, 6:348–349
- Thermistor clip, labial temperature and, 6:152
- Thermistor flow meters, 6:521
- Thermistor probes, 6:333
- Thermistors, 1:170, 193–194; 5:370; 6:314, 320–340. *See also* Temperature measurement
calibrating, 1:195, 196
clinical applications of, 6:333–338
commercially available, 6:333
fabrication of, 6:357
linearization of, 6:325–327
negative temperature coefficient, 6:333–338
resistance–temperature characteristics of, 6:323
temperature measurement electronics in, 6:327–333
terminology in, 6:321–323
thermal characteristics of, 6:324–325
voltage–current characteristics of, 6:323–324
- Thermoablation, 6:345
- Thermocouples, 6:314–313, 340–346
connection errors in, 6:344
in cryosurgery, 2:369–370
design of, 6:342–343
electromagnetic interference in, 6:344
material defects and ageing in, 6:344
measurement errors using, 6:343–344
in medicine and biomedical engineering, 6:345–346
response time in, 6:343–344
self-heating errors in, 6:344
sheathed, 6:342–343
temperature measurement using, 6:341–342
theory behind, 6:341–342
types of, 6:342
- Thermomodulation cardiac output
measurement, 2:25–35
application of the theory of, 2:28–29
equipment calibration for, 2:31
equipment for, 2:29–32
performance of, 2:32–34
theory of, 2:28
- Thermodynamics, laws of, 6:341
- Thermoelectric generator, 4:430
- Thermoelectricity, 6:357–358
- Thermogram patterns, abnormal, 6:348–349
- Thermography, 6:346–355
asymmetry analysis using, 6:351–352
in cryosurgery, 2:372
international activities using, 6:353–354
new generation, 6:349–353
pathophysiological-based understanding of, 6:347–349
- Thermoluminescence (TL), 5:514
- Thermoluminescent dosimeters (TLDs), 5:473–474
- Thermoluminescent dosimetry (TLD)
sheets/plates, 5:483–484
- Thermoluminescent materials, properties of, 5:515t

- Thermomagnetic (magnetic wind) oxygen analyzers, 5:200–201
- Thermometers
characteristics of, 6:313t
electrical resistance, 6:356–357
liquid-in-glass, 6:356
types of, 6:313
- Thermometry, 6:355–362
MRI, 6:359–360
radiation, 6:360–361
specifications in, 6:356
types of, 6:356–361
- Thermoplastics, 1:331–332
- Thermoregulation, in mammals, 1:190–192
- Thermosets, 1:331, 332
- Thermotherapy, history of, 3:463
- Thermotherapy devices, 3:466–474
- Theta waves, EEG, 3:66
- Thin-film electrodes, 1:156–157
- Thin-film-transistor (TFT) array, 5:343
- Thin-layer evaporation (TLE), 2:468
- Third degree (full thickness) burns, 6:171
- Thoracic cage, 6:102
- Thoracic cavity, parallel-column model of, 1:200
- Thoracolumbar region, degeneration–trauma in, 3:564
- Thoracotomy, for attaching heart electrodes, 1:151–152
- Thoratec HeartMate IP, 3:452
- 3D electrode array, 1:155, 157
- Three-dimensional conformal radiotherapy (3DCRT), 5:593; 6:30–38
beam determination in, 6:33–34
clinical consequences of, 6:37
dose calculation in, 6:34
dose prescription in, 6:33
imaging for, 6:30–32
radiation delivery in, 6:34–35
target volume definition in, 6:32–33
using particle radiation, 6:34
verification of radiation delivery in, 6:35–36
- Three-dimensional CT patient data acquisition, 2:268
- Three-dimensional detectors, 5:478–479
- Three-dimensional dosimetry, detectors for, 5:482–484
- Three-dimensional echocardiographic reconstruction, 3:22
- Three-dimensional histograms, 2:401
- Three-dimensional imaging (3D), 3:1; 6:465
- Three-dimensional nuclear medicine detectors, 5:100–103
- Three-dimensional PET imaging, 5:102–103
- Three-dimensional radiation dosimetry, 5:481–500
- 3D patterning methods, 1:413
- 3D reconstruction, for scoliosis, 6:126–127
- 3D surface imaging, in computed tomography, 2:239–242
- Three-function blood gas analyzer, 2:113
- Three-in-one nutrient system, 5:127–128
- 316L stainless steel, 1:271
- Threshold training, 1:176
- Throat diseases, cryosurgical treatment of, 2:374
- Thrombosis. *See also* Clotting; Thrombus
hepatic artery, 4:271
portal and hepatic vein, 4:271–272
- Thrombus. *See also* Clotting
acute coronary events and, 2:50–51
formation of, 1:503
stroke and, 2:52–53
- Thyroid cancer, radioiodine therapy
dosimetry for, 5:569
- Thyrotoxicosis, radioiodine therapy
dosimetry for, 5:568–569
- Ti6Al4V alloy, 1:271
- Tibial fracture orthosis (AFO), 6:82f
- Tickle Talker speech perception device, 6:296
- Tidal volume(s)
anesthesia machine, 1:38, 40
lower limit of, 3:498
lung, 2:39
- Tilt board, 5:590
- Time accuracy, in X-ray equipment, 6:567
- Time constants
long, 1:128
for skin parallel resistance, 1:135
- Time-delay image integration technique, 6:351
- Time-domain data analysis, EGG, 3:88–89
- Time-of-flight SIMS (TOF–SIMS) surface analysis, 1:350–351
- Timer reproducibility, in X-ray equipment, 6:567
- Time varying elastance, 3:478–480
- Tined lead implant, surgical technique for, 1:437–438
- Tin-stannous chloride, in electrodes, 1:131
- Tissue(s). *See also* Engineered tissue
effect of freezing on, 2:362–363
elemental compositions of, 5:254t
evaluation of thermal properties of, 6:337–338
oxygen in, 5:213–214
phantom materials in, 5:254–255
propagation of ultrasound in, 4:66–67
radiological equivalence of phantom materials to, 5:253
response to biomaterials, 1:108–109
response to bone implants, 1:109–110
ultrasonic properties of, 3:3t
- Tissue ablation, 6:362–379
chemical, 6:367–368
clinical applications and devices for, 6:362, 369–378
laser, 6:365–366
physical principles of, 6:362–368
- Tissue air ratio (TAR), 2:130
- Tissue birefringence, 3:166f
- Tissue compatibility, *in vivo* assessment of, 1:110
- Tissue damage, transient response and, 1:126–128
- Tissue Doppler imaging, 3:13f
- Tissue engineering, 1:366; 6:379–395. *See also* Engineered tissue; Soft tissue engineering
biomaterials for, 1:367–375; 6:383–387
biomaterial surfaces and, 1:343
bioreactors in, 6:387–388
case studies in, 6:389–394
cell sources for, 6:382–383
components of, 6:382–388
cytokine-release system for, 1:375–377
defined, 6:379–380
history of, 3:189–190; 6:381
inductive approaches to, 6:388–389
motivation for, 6:380–381
nanoparticles in, 5:5, 6
natural polymers in, 1:367–371
polymers for, 1:276; 5:388–390
scaffold matrices in, 1:367
synthetic polymers in, 1:371–374
in vascular graft prostheses, 6:500–501
- Tissue equivalence, criteria for, 5:255
- Tissue equivalent phantom, 5:262–263
- Tissue fluid extraction, glucose sensors, 3:397–398
- Tissue fusion, electrosurgical, 3:167–169
- Tissue heart valves (THVs), 3:429–430
- Tissue heating, 6:345
- Tissue heterogeneities, evaluation of, 5:493–494
- Tissue impedance, 1:198, 208–210; 4:135–137
- Tissue injury
with clinical lithotripsy, 4:261–263
from freezing, 6:369
from heating, 6:368–369
- Tissue injury/repair
cooling and, 3:466
thermotherapy and, 3:465–466
- Tissue layers, in organs, 6:180–181
- Tissue phantom ratios, 2:130
- Tissue regeneration
critical cell path length in, 6:185–186
implants and, 1:256–257
living environment parameters during, 6:183–184
for organ function loss, 6:183
- Tissue regeneration scaffolds
design principles for, 6:184
template pore structure of, 6:186–187
template residence time for, 6:184–185
- Tissue resistivities, 1:199t
- Tissue response, assessing acceptability of, 1:110
- Tissue sensing, in microsurgery, 4:529–530
- Tissue substitutes
basic data method of formulating, 5:256
classification and testing of, 5:258–262
effective atomic number method of formulating, 5:256
elemental compositions of, 5:261t
formulation procedures for, 5:255–256
materials and method of manufacture of, 5:256–258
types of, 5:258–259t
- Tissue transplants, 1:355
- Tissue triad structure, 6:180–181, 187–190
- Titanium, in dental implants, 1:328
- Titanium alloy coatings, 1:347
- Titanium alloys, 1:312–313
as biomaterials, 1:105, 270–271
- TITRATOR blood pressure regulator, 1:491
- T lymphocytes, implants and, 1:112, 113.
See also T cell entries

- Tomographic imaging, 6:559
- Tomographic reconstruction, computers in, 5:114–117
- Tomography
 electrical impedance, 6:361
 fast X-ray computed, 5:246
 optical coherence, 3:310–311
 phantom materials in, 5:266–267
 positron emission, 5:246–247
 quantitative computed, 1:553
 for scoliosis, 6:127
- Tomotherapy, 5:523; 6:395–401
 computed tomography simulation for, 2:274–275
 development of, 6:396–397
 helical, 6:398–399
 image guidance and adaptive radiotherapy in, 6:401
 serial, 6:397–398
- Tonometer sensors, design of, 6:404–406
- Tonometric blood pressure, measurements
 accuracy of, 6:408
- Tonometry, applanation, 5:236–237. *See also* Arterial tonometry
- Tooth/jaw biomechanics, 6:410–429. *See also* Masticatory system
- Tooth replacement modalities,
 biomechanical properties of, 6:424–427
- Tooth restorations
 extracoronaral, 6:425
 intracoronaral, 6:424–425
- Top hat housing, for electrodes, 1:139
- Topographical (3D) patterning methods, 1:413
- Topography control, in microbioreactors, 4:388
- “Torsade de Pointes,” 2:46–47
- Torsional eye movement, measurement of, 3:275
- Torsion angles, in protein structure prediction, 1:220
- Total body water (TBW) resistance, in dialysis patients, 1:212
- Total disk replacement, 6:237
- Total hospital safety program, 6:115–117
- Total internal reflection, in fiber optics, 3:302
- Total internal reflection microscopy (TIRM), 4:494
- Total joint replacement, 2:73; 4:540–541
- Total laboratory automation, 1:24, 25f
- Total lung capacity (TLC), 6:100
- Total parenteral nutrition, comparing methods of, 5:131–132
- Total parenteral nutrition regimen, selecting, 5:132
- Touch thresholds, 6:296
- Toxicity, of whole-body hyperthermia, 4:51
- Toxicological aspects of niosomes, 2:475
- Toxicological profile, of cyclodextrins, 2:454
- Toxicological testing, microbioreactors for, 4:393
- Trabecular bone, 1:533–534
- Trabecular bone measurement, broadband ultrasound attenuation in, 1:555
- Trace element requirements, in parenteral nutrition, 5:125t
- Tracer kinetics, 6:429–436
 arterial input function and, 6:435–436
 compartment modeling in, 6:431–434
 estimating kinetic parameters in, 6:434–435
 future issues in, 6:436
 of gadolinium, 6:432–433
 improving/automating analysis of kinetic tracer curves, 6:435
 model-free methods in, 6:433–434
 nomenclature related to, 6:430
- Tracers
 dynamic, 6:430
 static, 6:429–430
- Tracheoesophageal puncture (TEP), 4:232
- Tracheostoma breathing valve, 4:233
- Tracking systems, in microsurgery, 4:530
- Training
 for augmentative and alternative communication systems, 2:208–209
 medical device, 6:119
- Training to criterion, 1:177
- Transabdominal Doppler ultrasound, fetal, 3:290–291
- Transaction Processing System (TPS), 5:151
- Transcutaneous acupoint electrical stimulation (TAES), 3:27
- Transcutaneous bilirubin measurement, optical sensors in, 5:171–172
- Transcutaneous blood gas monitoring (TCM), 1:476–478
- Transcutaneous blood gas tension measurement, in neonatal monitoring, 5:24
- Transcutaneous blood oxygen monitoring
 history and theory of, 2:113
 methods development in, 2:113–114
- Transcutaneous CO₂, 2:114–116
- Transcutaneous cranial electrical stimulation (TCES), 3:27
 clinical usage of, 3:30–32
- Transcutaneous cranial electrical stimulators, using limoge currents, 3:29–30
- Transcutaneous electrical nerve stimulation (TENS), 3:26–27; 6:225, 437–452. *See also* Pain
 electric amplitude-frequency selection in, 6:449–450
 electrode placement in, 6:446–448
 electrodes in, 6:443–446
 indications for, 6:450t
 theories regarding analgesic effect of, 6:441
 treatment plans related to, 6:450–451
 warnings and contraindications concerning, 6:451
- Transcutaneous electrical nerve stimulation units, 6:443
 comparisons of, 6:444–445t
- Transcutaneous electrical nerve stimulators (TENS), 3:25–26
- Transcutaneous electrical stimulation, 6:296
- Transcutaneous mass spectrometry, in neonatal blood gas measurement, 5:24
- Transcutaneous oxygen, 5:213–214
- Transcutaneous oxygen tensions, 6:523
- Transcutaneous oxymetry, in hyperbaric medicine, 4:20
- Transcutaneous PCO₂ monitoring, 1:477
- Transcutaneous PCO₂ sensor, 1:477–478
- Transcutaneous PO₂ measurement (tcPO₂), 1:477
- Transcutaneous PO₂/PCO₂ monitoring, clinical applications of, 1:478
- Transcutaneous PO₂ sensor, 1:477
- Transcutaneous technology, applications of, 2:116–117
- Transcyte, as a skin substitute, 6:175
- Transdermal contraceptive patch, 2:343–344
- Transdermal drug delivery
 cyclodextrins in, 2:455–456, 457t
 microneedles for, 2:442–446
- Transdermal microemulsion application, 2:462–463
- Transducer interface, universal, 2:7
- Transducers
 airflow and volume, 5:433–436
 in biotelemetry systems, 1:418
 extravascular and intravascular, 1:485
 indicator-mediated, 5:163
 multielement, 4:65
 in phonocardiography, 5:283–286
 planar, 4:63–64
 single-element, 4:65
 ultrasound, 3:2; 6:456–459
- Transducer systems
 catheter-tip, 4:579
 fluid-filled catheter, 4:578–579
- Transducer technology, 1:486
- Transesophageal echocardiographic examination, 3:16–17
- Transesophageal echocardiography, 3:22
- Transesophageal transducers, 3:2
- Transfer functions, use of, 3:491–492
- Transfer tubes, in high-dose-rate remote afterloaders, 1:593–594
- Transforaminal lumbar interbody fusion (TLIF), 6:235
- Transformation, in joint biomechanics, 4:209
- Transformers, linear variable differential, 4:252–257
- Transfusion syndrome, twin-to-twin, 4:178–179
- Transient ischemic attack (TIA), 2:52–53
- Transient otoacoustic emissions (TOAEs), 1:101
- Transient response, tissue damage and, 1:126–128
- Transistors, ion-sensitive field-effect, 4:185–198. *See also* Immunologically sensitive field-effect transistors (IMFETs)
- Transit time volume flow meters, 3:325–327
- Transition temperature range (TTR), 1:325
- Transition temperatures, methods of measuring, 1:6–7
- Translate-rotate computed tomography, 2:231

- Translating bifocal contact lenses, 2:327
 Translation, in joint biomechanics, 4:209
 Translational servorecorders, 6:58–59
 Transmission, in biotelemetry systems, 1:420–421
 Transmission electron microscope (TEM), 4:478
 design of, 4:481–482
 Transmission oximetry, versus reflection oximetry, 1:470
 Transmission unit, in visual prostheses, 6:537–538
 Transpedicular screws, 3:570–571
 Transplantation
 liver, 4:266–277
 for organ function loss, 6:182
 Transplants, implants versus, 1:283–284.
 See also Tissue transplants
 Transponder electronics, 2:9–10
 Transponders, MEMS-based, 1:417–418
 Transport, as a hospital problem, 6:113
 Transportation safety, for wheelchair users, 4:550–552
 Transport regulations, for radioactive material, 2:166–169
 Transretinal approach, to developing visual prostheses, 6:535
 Transthoracic bioimpedance, applications of, 1:202
 Transthoracic echocardiographic examination, 3:14–15
 Transthoracic electrical bioimpedance (TEB), 1:199–204
 Transthoracic electrical impedance, neonatal respiratory monitoring by, 5:18–22
 Transthoracic impedance, combined with cardiac monitors, 5:22
 Transthoracic impedance techniques, 1:205
 Transthoracic transducers, 3:2
 Transvalvular impedance (TVI), 1:207–208
 Transvenous pacing of the heart, 1:150–151
 Transverse plane, scoliosis and, 6:123
 Trapping theory, 5:176–177
 Trauma, vascular graft prostheses and, 6:493
 Traumatic brain injury (TBI), 6:75
 biofeedback clinical outcome literature related to, 1:182–183
 Travel aids, conventional electronic, 1:448–450
 Treadmills, 3:251
 biofeedback and, 1:170
 Treatment control, in high-dose-rate remote afterloaders, 1:594
 Treatment couches, 5:591–593
 Treatment delivery quality assurance, 5:548
 Treatment planning
 dose calculation in, 5:537
 for prostate seed implants, 5:424–427
 quality assurance in, 5:545–547
 radiation therapy, 2:243–244
 radiosurgery, 5:580–583
 Treatment regimes, for UV therapy, 6:483
 Treatment simulation devices, 5:585–587
 Treatment verification devices, 5:599–601
 Tree-building methods, with
 bioinformatics, 1:220
 Trees, phylogenetic, 1:220
 Tricalcium phosphate (TCP), 1:261, 285, 314
 in tissue engineering, 1:374
 Triggered CT scan start, 2:238
 Triphasic mixture theory, 2:70–71
 Tripolar recording configuration, 3:115–116
 Tropocollagen, 1:107
 “Trotting Horse Method,” 2:35–36
t-test
 one-sample, 6:251–253
 paired, 6:253
 unpaired, 6:253–254
 Tubal sterilization, bilateral, 2:346–347
 Tube housing, in X-ray tubes, 6:605–606
 Tube housing cooling rate, in X-ray tubes, 6:607
 Tube housing heat capacity, in X-ray tubes, 6:607
 Tubular liquid-filled strain gage, 1:175
 Tumor ablation, 6:372–375
 Tumor destruction, nanoparticles in, 5:5
 Tumor imaging, using single photon emission computed tomography, 2:282–283
 Tumor localization devices, 5:585–587
 Tumor-reactive pre-effector T cells, induction of, 4:111–112
 Tumors
 cryosurgical treatment of, 2:374, 375
 effect of hyperthermia on, 4:49
 external beam radiotherapy options for, 6:5t
 imaging with monoclonal antibodies, 4:605–606
 redirecting effector T cells to, 4:116–117
 solid, 4:606
 studies of BNCT for, 1:581–582
 Tumor-specific antigens, 4:605
 Tungsten deposition, in X-ray tubes, 6:608
 Tunnel vision, 1:444
 defined, 1:444
 Turbostratic carbons, 1:297, 298f
 12-lead clinical electrocardiography, 3:42
 Twin-to-twin transfusion syndrome (TTTS), 4:178
 Twister brace, 6:86f
 2D electrode arrays, 1:156
 Two-dimensional detectors, 5:477
 Two-dimensional echocardiography, clinical uses of, 3:20
 Two-dimensional nuclear medicine detectors, 5:95–100
 Two-dimensional PET imaging, 5:101–102
 Two-dimensional sector scan (2D), 3:1
 Two-dimensional video sensor array eye tracking systems, 3:278–279
 Two-photon fluorescence microscopy, 4:494
 Two-way ANOVA, for two-factor experiments, 6:256
 Tympanic membrane temperature monitoring, 6:317
 Tympanograms, 1:100–101
 Tympanometry, 1:100–101
 multifrequency, 1:101
 Type I audiometer, 1:92–93
 Type I/II errors, 6:247
 Type II audiometer, 1:93
 Type IV audiometer, 1:93
 Type A audiometer, 1:92
 Type A tympanograms, 1:100–101
 Type B audiometer, 1:92
 Type C audiometer, 1:92
 Type E audiometer, 1:92
 Type HF audiometer, 1:92
 UBTL tests, 1:158, 159
 UHMWPE-on-metal/ceramic hip joints, 3:518–520
 Ulceration, skin, 6:206–207
 Ultradeformable liposomes
 formulative aspects of, 2:479
 therapeutic potentialities of, 2:479–480
 Ultradeformable vesicular drug carriers, 2:479–480
 Ultra-density optical (UDO), 5:348
 Ultrahigh molecular weight polyethylene (UHMWPE), 1:274, 316–317, 331, 333–335. *See also* UHMWPE entries as biomaterial, 1:106–107
 Ultrahigh molecular weight polyethylene hip joints, 3:514, 515, 516
 Ultrahigh vacuum instrumentation, in surface analysis, 1:348, 349
 Ultralow temperature isotropic (ULTI) carbon, as a biomaterial, 1:273
 Ultrasonic attenuation, 3:3
 Ultrasonic devices, therapeutic applications of, 1:190
 Ultrasonic hyperthermia, 4:62–86
 medical applications of, 4:67
 Ultrasonic imaging, 6:453–473. *See also* Acoustic imaging; Contrast imaging; Ultrasound entries
 array design in, 6:457–458
 bioeffects of, 6:471–473
 physical principles of, 6:453–456
 pulse-echo method in, 6:455–456
 Ultrasonic transducers, radiation field of, 4:63
 Ultrasonic wave propagating velocity (UV), 1:554
 for bone measurement, 1:555
 Ultrasonometry, quantitative, 1:553–556
 Ultrasound
 in biotelemetry systems, 1:421–422
 clinical formats of, 3:2–3
 during cryosurgery, 2:372
 exposure to, 5:69
 generation of, 4:63; 6:456–459
 high intensity focused, 6:365
 interstitial, 4:39–40
 principles of, 3:3–5
 propagation of, 4:63, 66–67
 signal processing, display, and management in, 3:10–13
 temperature change estimation with, 6:361

- use in medical diagnosis and therapy, 5:71
 versus electromagnetic radiation, 4:63
 Ultrasound ablation, 6:365, 376
 Ultrasound–acoustic pneumotachometers, 5:370
 Ultrasound contrast agents, modern, 6:466, 467t
 Ultrasound flow measurement techniques, 3:325–327
 Ultrasound-guided prostate seed implants, 5:424
 Ultrasound imaging
 during ablative treatment, 6:374f, 375
 in peripheral vascular noninvasive measurements, 5:245
 of polymer gels, 5:488
 Ultrasound physics books/reports, 4:345–346
 Ultrasound propagation, in biological tissues, 4:66–67
 Ultrasound scanners
 B-mode in, 6:461
 M-mode in, 6:461–462
 Ultrasound therapy devices, 3:473–474
 Ultrasound transducers, 6:456–459
 acoustic fields of, 6:458–459
 UltraStim Snap Electrodes, 1:149
 Ultraviolet fibers, 3:303–304
 Ultraviolet phototherapy, 6:482
 Ultraviolet radiation (UVR), 5:65–66. *See also* UV entries
 biological effects of, 6:474–476
 diagnostic uses of, 6:480–482
 hazard assessment and protection related to, 6:487–488
 measurement of, 6:478–480
 in medicine, 6:473–490
 skin disease therapy using, 6:482–485
 sources of, 6:476–478
 use in medical diagnosis and therapy, 5:70
 Ultraviolet spectrum, 6:474
 Ultraviolet therapy equipment, 6:484
 Umbilical artery/vein catheterization
 complications and risks associated with, 4:594–595
 indications and contra-Indications for, 4:589–590
 procedure for, 4:590
 Umbilical artery/vein monitoring, 4:588–597
 anatomical and physiological aspects of, 4:589
 historical aspects of, 4:589
 Umbilical catheter, removal and maintenance of, 4:595
 Umbilical cord occlusion, bipolar, 4:179
 Unalloyed titanium, as biomaterial, 1:105
 Unconditioned response (UCR), 1:166
 Unconditioned stimulus (UCS), 1:166
 Uncooled thermal detectors, 6:349
 Unfilled acrylic resins, 6:93–94
 Ungerleider electrode, 1:138
 Uniaxial tensile behavior, of arterial walls, 1:87
 Unidirectional valves, 1:33
 Uniformity, Anger camera, 1:58–59
 Unipolar electrodes, 1:151
 Unipolar electrograms, 1:69f
 “Unipolar” ventricular electrograms
 in separating ventricular fibrillation from tachycardia, 1:79
 United Kingdom, infrared imaging in, 6:353
 United Nations Committee on the Effects of Atomic Radiation (UNSCEAR), 2:154
 United States. *See also* American entries;
 National entries
 audiometers in, 1:92
 biomaterials regulation in, 1:268–270
 biomedical engineering education in, 1:403–404
 infrared imaging in, 6:353
 mandated web accessibility in, 1:447
 medical engineering societies and organizations in, 4:316–321
 physical fitness in, 1:388
 visual impairment in, 1:443–444
 Units, radiation, 5:504–505
 Universal-function electrode standards, 1:161
 Universal transducer interface, 2:7
 Unmatched samples
 Kruskal–Wallis test for, 6:259
 one-way ANOVA for, 6:254
 Unpaired samples
 χ^2 test to compare, 6:250–251
 Mann–Whiney *U* test for, 6:258–259
 Unpaired *t*-test, to compare unpaired data samples, 6:253–254
 Ununited bone fracture, 1:558
 Up-Converting Phosphor Technology (UPTTM), 4:382
 Update report requirements, codes and regulations related to, 2:148
 Upper airway fistulas/pneumothoraces, 3:507
 Upper cervical spine
 anatomy of, 3:547–550
 stabilization of, 3:574–575
 Upper cervical spine instability, role of environmental factors in, 3:558–561
 Upper extremity prostheses, fixation of, 5:196–197
 Upper gastrointestinal bleeding, 3:385–390
 nonvariceal, 3:388–390
 Upper limb orthotic devices, 6:89t
 Upper limb technical analysis form, 6:91f
 Ureteroscopy, 4:539
 Urethane linkage, 1:336
 Urinary bladder temperature monitoring, 6:318
 Urinary incontinence, biofeedback clinical outcome literature related to, 1:180, 182
 Urinary tract obstruction, lower, 4:173–175
 Urine component home health care devices, 3:532
 Urine specimens, collecting, 1:19
 Urologic procedures, electrosurgery in, 3:160
 Use-related hazards, of medical devices, 3:538
 User input device, in neurological monitors, 5:34
 User-interface features, anesthesia machine, 1:37
 User testing, effect on human factors and medical devices, 3:540
 Utah probe, 1:156
 Uterine contractions
 direct monitoring of, 3:293–294
 electronic signal processing of, 3:295
 fetal monitoring and, 3:293–296
 indirect monitoring of, 3:294–295
 Uterine electromyogram, 3:295–296
 UVA1 therapy, 6:483. *See also* Ultraviolet radiation (UVR)
 UV absorption, in oxygen detection, 5:207
 UV detectors, physical, 6:479
 UV radiometers, 6:479–480
 UV therapy, treatment regimes for, 6:483
 Vacutainer cell preparation tubes, 1:462
 Vacuum constrictive device (VCD), for erectile dysfunction, 6:158
 Vacuum-form body immobilizer, 5:589f
 Vacuum surface analysis techniques, 1:349
 Vacuum systems, 3:381–383
 components of, 3:382–383
 maintenance of, 3:383–384
 performance criteria and standards for, 3:383
 vacuum sources for, 3:382
 Vacuum tube electrosurgical units, 3:161–162
 Vacuum tubes, electrocardiograms using, 1:137
 Vaginal blood volume (VBV), 6:150, 151
 Vaginal dilators, 6:153–154
 Vaginal electromyography, 6:152
 Vaginal fluid production, 6:153
 Vaginal pH, 6:153
 Vaginal photoplethysmography (VPPG), 6:150–152
 Vaginal pulse amplitude (VPA), 6:150, 151
 Vaginal ring contraceptive, 2:344
 Vaginal spermicides, 2:340
 Vaginal temperature gauge, 6:152
 Valleylab ablation system, 6:374
 Valve homografts, 3:443–444
 Valves
 cerebrospinal fluid drainage, 4:10–12
 microbioreactor, 4:389
 unidirectional, 1:33
 Valve stenosis, 2:18–19
 Valvuloplasty, 4:176
 Vanadium, in biomaterials, 1:105
 Vanadium alloys, 1:312–313
 Vaporizers, anesthesia machine, 1:35–36, 41
 Vapor-phase carbons, 1:300
 Variable-bypass vaporizers, anesthesia machine, 1:35–36
 Variable capacitance displacement sensor, in neonatal respiratory monitoring, 5:17

- Variable resistance exercise, 1:392, 400
 Variable velocity exercise, 1:392, 399
 Variance. *See* ANOVA (analysis of variance)
 Varices, gastroesophageal, 3:385–388
 Vascular access complications, of parenteral nutrition, 5:128
 Vascular applications, for porous biomaterials, 5:403
 Vascular diseases
 studies of, 4:395
 vascular graft prostheses and, 6:493
 Vascular graft prostheses, 6:491–505
 clinical need for, 6:491–493
 current, 6:494–498
 failure mechanisms in, 6:496–498
 new developments in, 6:498–501
 Vascular grafts, healing of, 6:496
 Vascular headache, biofeedback clinical outcome literature related to, 1:178–179
 Vascular imaging, 4:291–292. *See also* Digital angiography
 future of, 2:426
 Vascularization, of engineered tissue, 3:191–192
 Vascular pattern, in colposcopy, 2:199–200
 Vascular replacements, history of, 6:493–494
 Vascular resistance, 2:19
 Vascular unloading, 5:235
 Vasculature, pulmonary, 6:103
 Vector timing and correlation (VTC) algorithm, 1:78
 Vectorcardiography, 3:48–49
 Vehicle control systems, for wheelchair users, 4:550–551
 Vein grafts, 6:494
 Velocimetry
 laser Doppler, 3:332–335
 particle image, 3:335–340
 particle tracking, 3:335, 339
 Velocity flow measurements, 3:329–340
 Venous congestion plethysmography (VCP), 5:243–244
 Venipuncture, best sites for, 1:458
 Venipuncture standards, 1:455–456
 Venitricular function assessment, nuclear, 3:254
 Venous blood, oxygen in, 5:214–215
 Venous occlusion PG (VOP), 5:243–244
 Venous pulsations, pulse oximetry and, 5:212
 Ventilation. *See also* High frequency ventilation (HFV)
 adequacy of, 6:519–520
 alveolar minute, 6:104–105
 as a hospital problem, 6:112
 impedance plethysmography and, 4:128–129
 maximal voluntary, 5:439
 mechanical, 1:29; 6:107
 versus respiration, 6:509
 Ventilator, anesthesia machine, 1:37–38. *See also* Acute medical care ventilators
 Ventilator technology, improvements in, 1:41–42
 Ventilatory drive, 6:518
 Ventilatory gas exchange responses, in exercise stress testing, 3:253–254
 Ventilatory monitoring, 6:514–528
 evolving technology in, 6:527
 gas exchange assessment, 6:518–520
 indications for, 6:514–515
 levels of, 6:515
 methods and devices used in, 6:520–527
 parameters used in, 6:515
 Ventilatory support
 adverse reactions to, 6:512–513
 goals of, 6:509–510
 indications for, 6:509
 modes of, 6:510–511
 Ventilation, spontaneous, 1:29
 Ventricle, stroke volume changes in, 4:567
 Ventricular assist devices (VAD), 3:451–452
 design considerations for, 3:450–451
 electric, 3:452–456
 Ventricular cells, impact of computational modeling in, 3:148
 Ventricular defibrillation, cardiopulmonary resuscitation via, 2:36–37
 Ventricular enlargement, in hydrocephalus, 4:3
 Ventricular fibrillation (VF), 1:69; 2:45, 46
 versus ventricular tachycardia, 1:79
 Ventricular pump failure, 4:164
 Ventricular sensing, 5:221
 Ventricular tachycardia (VT), 1:69, 77, 78, 79; 2:45–46
 depolarization width for detecting, 1:76
 versus ventricular fibrillation, 1:79
 Ventriculography, 6:266
 Ventriculostomy, in hydrocephalus, 4:14–15
 Verigene™ nanoparticle-based bio-barcode system, 4:379–380
 Vertebrae. *See also* Spinal entries; Spine entries
 plastic, 3:573
 spinal, 6:230
 Vertebral compression fracture (VCF), 6:232
 Vertebroplasty, 1:540
 Vertical Expandable Prosthetic Titanium Rib (VEPTR), 6:234
 Vesicles by extrusion technique (VET), 2:469
 Vesicular drug carriers, 2:466–480
 ultradeflatable, 2:479–480
 Vessel closure devices, 2:352–353
 Vessel geometry perturbation, 1:612
 Vessel sealing, electrosurgical, 3:167–169
 Vessel segment, cylindrical model of, 1:199–200
 Vestibulo-ocular response (VOR), 5:139–140
 Veterans Specific Activity Questionnaire (VSAQ), 3:253t
 VF counter (VFCNT), 1:72
 Vibrational spectroscopic imaging, of polymer gels, 5:488
 Vibrotactile elements, 6:294
 Vibrotactile stimulation, 6:154
 Vibrotactile stimulators, 6:294
 Video-based eye tracking systems, for use with fMRI, 3:283
 Video conferencing, in office automation systems, 5:159
 Video dimension analyser (VDA), 4:247
 Video magnifiers, 1:444–445
 Viewing time studies, sexual arousal and, 6:160
 Virtual colonoscopy, 2:242–243, 261–263
 Virtual environments, 1:453–454
 Virtual reality
 for activities of daily living assessment and training, 6:76
 for attention assessment and training, 6:74–75
 as a computer-generated technology for rehabilitation, 6:74
 for cognitive training, 6:74
 for memory assessment and training, 6:75–76
 Virtual reality kitchen environments, 6:76
 Virtual reality mobility environments, 6:76
 Virtual simulation, 5:455–457
 software, 5:532
 Viruses
 from banked blood, 1:513
 electron microscopic diagnostic criteria for, 4:484–485
 Visceral neural signals
 in bladder function control, 3:128–129
 in respiratory control, 3:129
 Visceral tumors, cryosurgical treatment of, 2:375
 Viscoelasticity, of bone, 1:531–532
 Viscoelastic materials, in arterial walls, 1:85
 Viscoelastic profile, of blood, 1:501–502
 Viscometers, 1:505
 Viscosity
 of acrylic bone cement, 1:544–545
 of blood, 1:500–502, 504–505
 temperature and, 1:501
 Viscosity-density effect, in piezoelectric sensors, 5:361
 Visible fibers, 3:303–304
 Visible radiation, 5:66
 Visible speech, 2:218
 Vision rehabilitation, with visual prostheses, 6:542
 Visual analog scale (VAS), 1:44
 Visual evoked fields (VEFs), biomagnetic measurements and, 1:243–244
 Visual evoked potential (VEP), pattern electroretinogram and, 3:155
 Visual field testing, 6:528–530
 future directions in, 6:530
 methods of, 6:528–530
 newer modalities of, 6:530
 Visual impairments, 1:443. *See also* Visually impaired persons
 consequences of, 1:444
 Visualization, in phonocardiography, 5:287
 Visually impaired persons, assistive technology for, 1:443–455

- Visual prostheses, 6:530–549
 advanced applications of, 6:546
 approaches to developing, 6:534–536
 candidates for, 6:542–543
 chemical stimulation approach to developing, 6:534
 cortical approach to developing, 6:535–536
 engineering aspects of, 6:536–542
 ethical aspects of, 6:544–545
 evaluation of, 6:545
 external components of, 6:536–538
 history of, 6:533
 human and medical aspects of, 6:542–545
 hybrid approach to developing, 6:535
 implanted components of, 6:538–539
 issues related to, 6:545–546
 limitations of, 6:545–546
 operation of, 6:539–542
 optic nerve approach to developing, 6:533
 risks associated with, 6:543–544
 surgical methods for, 6:543
 theory of operation of, 6:533–534
 transretinal approach to developing, 6:535
- Visual system, basics of, 6:531–532
 Vital capacity (VC), 6:100
 Vitallium, 1:271, 312
 Vitamin D production, ultraviolet radiation in, 6:487
 Vitamin requirements, in parenteral nutrition, 5:125t
 Vitros 950 reaction slide, 1:20
 Vocabulary selection, for augmentative and alternative communication systems, 2:209–210
 VOCARE bladder system, 1:432, 438–441
 Voice, after laryngectomy, 4:230
 Voice command, in electronic aids to daily living, 3:214
 Voice prostheses, 4:231–234
 removable, 4:232
 Volitional tasks, 1:386
 Volta, Alessandro, 1:429
 Voltage-based enhanced resolution techniques, in electrophoresis, 3:137–138
 Voltage-clamp technique, 3:142
 Voltaic piles, 1:143
 Volume conductance catheter, 1:206–207
 Volume currents, 1:238
 Volume determination, model-based relations for, 4:126–127
 Volume displacement respiratory flow devices, 5:368–369
 Volume-displacement spirometers, 5:372
 Volume flow measurement, 3:324–329
 Volume–pressure curves, 6:517
 Volume transducers, 5:435–436
 Volumetric heaters, 1:190
 Volumetric oscillometry, 1:14
 Volumetric pump drug infusion systems, 2:498–499
 Voluntary muscles, 1:385
 von Bekesy, Georg, 1:93
 Vortex shedding, 6:521
 Vortex shedding flowmeter, 5:370–371
 Vroman, Leo, 1:344
 Vroman effect, 1:344, 345
 VT counter (VTCNT), 1:72
 Vulcanization, 1:337
 Vulvalgesiometer, 6:153
 Vulvar vestibulitis syndrome (VVS), 6:153
- Wafer bonding, 2:3
 Walkers, 4:547–548
 Waller, Augustus, 1:137
 Wallerian degeneration, electron microscopic diagnosis of, 4:486
 Wallis device, 6:236
 Wall lasers, 5:530
 Warburg apparatus, 5:208
 Waste gas scavenger system, in anesthesia delivery, 1:30
 Water, as a hospital problem, 6:113
 Water kerma-based backscatter factors, 6:587t
 Watermarking, 4:359
 Water perfused catheter, 1:62
 Water-powered microdrug delivery system, 2:504
 Water vapor, CPR and, 2:39–40
 Wave intensity analysis, 3:489–490
 Wave propagation/reflection analyzing, 3:490–491
 in the arterial tree, 3:487–490
 Waveform processing, in exercise stress testing, 3:249
 Wavelet transform algorithm for, 1:75
 event-related oscillations and, 3:239–241
 Wear
 biomaterial failure from, 1:278–279
 of biomaterials, 1:308, 314–320
 of cartilage, 2:71
 of pyrolytic carbons, 1:303
 Wearable electrodes, 1:141–142
 Wearable Health Care System (WEALTHY), 1:142
 Wearable monitoring systems, 1:142
 Wearable tactile displays, 6:299
 Wear assessment, 1:315–316
 Wear-corrosion, 1:311
 Wear factor, 1:315
 Wear geometries, 1:316
 Wear resistance
 of titanium alloys, 1:312–313
 of ultrahigh molecular weight polyethylene, 1:316–317
 Wear testing, 1:315
 Web accessibility, for the sight impaired, 1:447
 Web-based teleradiology, 6:308
 Web Content Accessibility Guidelines (WCAG), 1:447
 Wedge filters, 5:596–597
 Weeping lubrication, 6:417
 Weibull theory, 1:302
 Weight-based exercise equipment, 1:397
 Weighted irradiance, 6:478–479
 Weighted least squares process, in kinetic parameter estimation, 6:434
 Welch cup electrode, 1:138
 Welch method, power spectrum estimation using, 3:75–77
 Well counters, 5:94–95
 Wellness programs, corporate, 1:388
 Western Electric 1A audiometer, 1:92
 Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) scale, 5:446
 West Nile virus, 1:513
 Wet chemical techniques, for biomaterial surface modification, 1:347–348
 Wet etching, silicon, 2:3
 Wet gel disposable electrode, 1:140
 Wet gels, 1:134–135, 135–136, 141
 Wettability, of polymers, 1:314
 Wet Test Gas Meter, 5:373–374
 Wheatstone bridges, 4:578; 6:285, 326, 327–328
 Wheelchair basketball, 4:548–549
 Wheelchair racing, 4:549
 Wheelchair rugby, 4:549–550
 Wheelchairs
 manual, 4:546
 powered, 4:547
 powered assist, 4:546–547
 Wheelchair tennis, 4:550
 Wheelchair tie-down and occupant restraint systems (WTORS), 4:551–552
 Wheelchair users, transportation safety and adaptive driving for, 4:550–552
 Whitaker Foundation, 1:404
 Whiteboard, in office automation systems, 5:157
 White blood cell count, 2:88
 White blood cell nuclei, size and shape of, 2:470–471
 White blood cells (WBCs), 1:503, 507
 differential analysis on hematology systems, 2:414–419
 measurable properties of, 2:410–411
 size of, 2:410
 “White coat effect,” 1:485
 White-coat hypertension, 1:15
 Whiteside, George, 1:411
 Whole blood, processing and collection of peripheral blood mononuclear cells from, 1:459–460, 461–462
 Whole-body bioelectric impedance measurement, in dialysis patients, 1:212
 Whole-body bioimpedance spectroscopy, 1:212
 Whole-body hyperthermia
 clinical toxicities of, 4:51
 clinical trials of, 4:54–55t
 fever-range, 4:58
 immune system and, 4:50–51
 nanoparticle therapy and, 4:49t
 Whole-body hyperthermia devices, commercially available, 4:45–46, 47t
 Whole-body models, regional circulation and autoregulation in, 5:305–307
 Whole-body plethysmograph, in neonatal respiratory monitoring, 5:15–16
 Whole-body systems models, 5:302–303
 Whole word lookup, 1:447

- Wide-field deconvolution microscopy, 4:493
- Wide-field fluorescence microscopy, 2:92
- Widefield microscopy, laser and arc-discharge spectral lines in, 4:466t
- Wilcoxon test, for paired samples, 6:258
- Williams, David, 1:281
- Windkessel model, 3:485–486; 6:408
three-element and higher order, 3:486
two-element, 3:485–486
- Wipes, skin electrodes and, 1:136
- Wire electrodes, EMG, 3:104
- Wireless chips, 1:420–421
- Wireless communication, in biotelemetry systems, 1:418–422
- Wireless strain gage, 6:285
- Wire sensors, in thermocouples, 6:342
- Wistar rat organ weights, 5:572t
- Wolff's law, 1:256–257
- Wood, Earl, 1:469
- Word processing, in office automation systems, 5:153
- Work, thermoregulation and, 1:191
- Work cell technologies, 1:24
- Work domain analysis, effect on human factors and medical devices, 3:540
- Workflow, radiology, 5:550–551
- Workflow systems, in office automation systems, 5:157–158
- Work of breathing measurements, 6:517–578
- World Health Organization (WHO), 1:267
- Wound healing, response to biomaterials, 1:108–109
- Wounds, hyperbaric medicine and, 4:23–24
- Wound sepsis, implant-contiguous, 1:117–118
- Wright Peak Flow Meter, 6:522
- Wright Respirometer, 6:522
- Wrist, stability of, 4:224
- Wrist blood pressure monitoring, 1:489
- Wrought stainless steels, in dental prosthetics, 1:325–326
- X-ray absorptiometry, dual-energy, 1:552–553
- X-ray beam energy parameters, 6:599
- X-ray beam hardening, 6:577
- X-ray beams
definition of, 6:554–556
quality of, 6:583–584
weighted mean energy of, 6:599
- X-ray CT scanning, of polymer gels, 5:488
- X-ray detector, in CT scanners, 2:235–237
- X-ray diffraction (XRD), 1:356, 357, 358, 359, 362–363
- X-ray equipment
control and supervisory logic in, 6:557–558
custom applications of, 6:559–560
design of, 6:550–560
patient support structures in, 6:558–559
performance of, 6:562
power components in, 6:550–552
selecting, 6:559–560
specialized tests for, 6:561
- X-ray exposure rates, 6:571
- X-ray field, limitation of, 6:571
- X-ray field-light field alignment, quality control of, 6:565
- X-ray generators, 6:550–552
quality control of, 6:567
- X-ray mammography, 4:298–299
- X-ray markers perturbation, 1:610
- X-ray output, fluoroscopic, 6:571
- X-ray photoelectron spectroscopy (XPS)
surface analysis, 1:349–350, 352
- X-ray production, 6:599–608. *See also* X-ray tubes
- X-ray quality control program, 6:560–580
artifacts in, 6:573, 577
automatic exposure termination in, 6:569
beam alignment in, 6:563
bearing rotation in, 6:563
body-section tomographic equipment in, 6:574
bucky motion in, 6:570
bucky tray in, 6:569
centers alignment in, 6:570
collimator assembly in, 6:565
computed tomography in, 6:574–578
contrast ratio in, 6:572
contrast scale in, 6:575
conversion gain in, 6:572
CT number linearity in, 6:575
digital image receptors in, 6:578–579
electrical inspection in, 6:563
film processors in, 6:578
fluoroscopic equipment in, 6:570
fluoroscopic X-ray output in, 6:571
focal spot sizes in, 6:564
function checks in, 6:569
grid uniformity in, 6:570
half-value layer determination in, 6:566
high contrast spatial resolution in, 6:576–577
image quality in, 6:573–574
interlock test in, 6:570
leakage radiation in, 6:564–565
light localizer illumination in, 6:566
low contrast resolution in, 6:577
mammographic X-ray equipment in, 6:573
maximum entrance exposure rate in, 6:571
mechanical inspection in, 6:563
milliamperage accuracy in, 6:567–568
milliamperage linearity in, 6:568
operator's control panel in, 6:569
patient dose in, 6:573, 577–578
patient entrance exposure rates in, 6:571
peak kilovoltage calibration in, 6:568–569
phototimer tracking in, 6:569
pinhole images in, 6:574
positive beam limitation in, 6:566–567
primary barrier transmission in, 6:570
radiation waveform in, 6:563–564
regulatory checks in, 6:569
scattered radiation levels in, 6:572–573
source-to-skin distance in, 6:567
spatial resolution measurements in, 6:572, 573
spatial uniformity in, 6:576
- table and tube stand in, 6:569
time accuracy in, 6:567
timer reproducibility in, 6:567
tracking test in, 6:570
tray transmission in, 6:574
visual inspection in, 6:570
X-ray output in, 6:563
X-ray tubes in, 6:562–563
- X-ray radiography, phantom materials in, 5:266
- X-ray reproducibility, quality control of, 6:567
- X rays, 5:503
in diagnosing implant-related infection, 1:116
interaction with matter, 6:590–599
types of interactions with matter, 6:597–598
- X-ray source, in CT scanners, 2:234–235
- X-ray therapy equipment
clinical dosimetry and, 6:586–589
dosimetry calibration in, 6:584–586
low and medium energy, 6:580–590
quality assurance for, 6:589
for radiation therapy, 6:581–583
- X-ray tube rating chart, 6:607–608
- X-ray tubes, 6:552–554
construction of, 6:600–606
failure of, 6:608
heat dissipation in, 6:606–608
quality control of, 6:562–563
- X-ray units
analytic, 2:175
diagnostic, 2:176t
- Xenogeneic cells, in engineered tissue, 3:193
- Xenogenic tissue transplants, 1:355
- Xenografts, 1:283; 3:444–445
as a skin substitute, 6:174
- Xenon arc lamps, ultraviolet radiation from, 6:477
- Xerogels, 1:288
- Xtratek electrode tester, 1:160
- Yielding, of bone, 1:530–531
- Young's modulus (E), of bone, 1:528–529
- z -axis resolution, in computed tomography, 2:252
- Zero crossings and turns counting method, 3:108
- Zero-order kinetics, through nanoporous membranes, 2:450–451
- Zero order release profile, 2:439
- Zero-power resistance, in thermistors, 6:322
- Zero-resolution cell measurement systems, 2:402–400
- Ziegler catalysts, 1:106
- Ziegler-Natta catalysts, 1:330, 331, 333, 334
- Zimmer dough, 1:546
- Zirconia, in orthopedic prostheses, 1:317–318
- Zirconium alloys, in metal-on-metal prostheses, 1:317–318
- Zones, of articular cartilage, 2:64–65