Joseph N. Pelton
Scott Madry
Sergio Camacho-Lara
Editors

VOLUME 1

Handbook of Satellite Applications





Handbook of Satellite Applications

Joseph N. Pelton • Scott Madry
Sergio Camacho-Lara
Editors

Handbook of Satellite Applications

With 494 Figures and 78 Tables

Springer

Editors
Joseph N. Pelton
Former Dean
International Space University
Arlington
Virginia, USA

Scott Madry
International Space University
Chapel Hill
North Carolina, USA

Sergio Camacho-Lara
CRECTEALC
Tonantzintla
Puebla, Mexico

ISBN 978-1-4419-7670-3 ISBN 978-1-4419-7671-0 (eBook)
ISBN 978-1-4419-7672-7 (print and electronic bundle)
DOI 10.1007/ 978-1-4419-7671-0
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012952160

© Springer Science+Business Media New York 2013
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

Imagine if all satellite services would close down for even a few hours. The global consequences of such a catastrophic event – even for specialists in the field – would be hard to grasp. We can easily imagine huge disruptions in telecommunications traffic and banking operations occurring within minutes. In time chaos would spread to stock markets, television broadcasts, weather forecasting, and storm alerts, as well as airline travel. By the second hour the problems would have even spread to activities like education, health care, and many other basic services of industry and government. Some years back, a communications satellite failed and the satellite-based pager system for many doctors, surgeons, police, and firemen suddenly went down. For the first time people began to realize just how dependent they were on satellites in their daily lives.

At the 2012 International Space University (ISU) symposium on Space Sustainability, one expert referred to the "Day After" scenario for the possibility when all satellites might fail. Whereas one can survive without a utility like electricity for a short time, the longer-term consequences for a global society would be quite dramatic. The same would happen for a world without satellite services. A world stripped of its application satellites would be set back many decades in its progress. We would suddenly inhabit a world where misinformation could reign again. It is not an overstatement to say that a world without satellites could actually plunge us into war.

In short, space applications today have become a utility, just as in the case of electricity or water. We basically do not wonder where our electricity or our water was produced when we use a power socket or turn on a water faucet. We just assume it will be available with good quality in a sustainable way. Today, in a world with rising population and climate change we are becoming more and more concerned about long-term availability of resources, and when we do so we also need to reflect on the availability of space resources. Just think about the consequences of a strong solar storm such as that occurred in 1859. This quite unusual Coronal Mass Ejection (CME) or solar eruption, now called the Carrington Event, managed to set telegraph offices on fire and brought the aurora borealis as far south as Cuba and Hawaii for many days. It is evident that a repetition of such an event nowadays would bring considerable damage to our application satellites and could interrupt global satellite services in a major way.

Knowledge of satellite applications is important, but it is equally important to understand the whole system starting from the underlying basics of the technology and how satellites are built and operated nationally, regionally, and globally. We need to know the potential of these satellites today and tomorrow as well as understand the threats that can influence their performance.

This handbook is exactly aimed to fulfill this purpose and provides an excellent and outstanding overview of satellite applications, at the same time emphasizing the regulatory, business, and policy aspects. The authors are among the best experts worldwide and it is a pleasure to note that many of them are regular lecturers at the International Space University, which at the same time guarantees the interdisciplinary character of this unique standard work.

ISU is for these reasons proud to fully endorse this important handbook!

President, ISU
Strasbourg, France

Prof. Walter Peeters

Acknowledgments

The multiyear creation of the *Handbook of Satellite Applications* represents a mammoth effort by many dozens of authors. Some of the world's most outstanding experts in their field graciously supported this effort as volunteers. We thus wish, first and foremost, to thank these authors who have so generously contributed toward this attempt to creating a truly definitive reference work across the closely related fields of satellite communications, remote sensing, space navigation, and meteorological sensing from space.

We also wish to thank Dr. Michael Simpson, former President of the International Space University (ISU). It was he who first encouraged the creation of this reference work and had the vision to support a handbook that would comprehensively cover the entire range of major satellite applications. Also, it is important to recognize the current president of the International Space University (ISU) Walter Peeters, dean Angelina Bukley, plus the current and past chairs of the ISU Academic Council, who were also involved with this project, namely, Drs. Edward Chester and Stefano Fiorilli.

Finally we would like to thank the many people at Springer Publishing who nurtured and supported this project during its 2-year gestation. We wish to thank Maury Solomon, who first conceived that such a project would be an important undertaking for the scientific literature in the field. Particular thanks also go to Barbara Wolf, Marion Kraemer, and Saskia Ellis who carefully oversaw the final editing and kept the production schedule more or less on track. Then there was Vasuki Ravichandran, who, along with Ms. Ellis, very scrupulously oversaw the production of this extensive reference work on a day-after-day basis. We thank them both for their constant eye to detail and their tireless efforts.

Joseph N. Pelton
Sergio Camacho-lara
Scott Madry

Contents

Volume 1

Section 1 Satellite Communications 1

1 Satellite Applications Handbook: The Complete Guide to Satellite Communications, Remote Sensing, Navigation, and Meteorology 3
Joseph N. Pelton, Scott Madry, and Sergio Camacho Lara

2 Satellite Communications Overview 19
Joseph N. Pelton

3 History of Satellite Communications 27
Joseph N. Pelton

4 Space Telecommunications Services and Applications 67
Joseph N. Pelton

5 Satellite Orbits for Communications Satellites 93
Joseph N. Pelton

6 Fixed Satellite Communications: Market Dynamics and Trends 115
Peter Marshall and Joseph N. Pelton

7 Satellite Communications Video Markets: Dynamics and Trends 135
Peter Marshall

8 Mobile Satellite Communications Markets: Dynamics and Trends 163
Ramesh Gupta and Dan Swearingen

9 An Examination of the Governmental Use of Military and Commercial Satellite Communications 187
Andrew Stanniland and Denis Curtin

10 Economics and Financing of Communications Satellites 221
Henry R. Hertzfeld

11 Satellite Communications and Space Telecommunications Frequencies 239
Michel Bousquet

12 Regulatory Process for Communications Satellite Frequency Allocations 271
Ram S. Jakhu

13 Satellite Radio Communications Fundamentals and Link Budgets 293
Daniel R. Glover

14 Satellite Communications Modulation and Multiplexing 325
Paul T. Thompson

15 Satellite Transmission, Reception and On-Board Processing Signaling and Switching 359
Bruno Perrot

16 Satellite Communications Antenna Concepts and Engineering 373
Takashi Iida

17 Satellite Antenna Systems Design and Implementation Around the World 397
Takashi Iida and Joseph N. Pelton

18 Satellite Earth Station Antenna Systems and System Design 427
Jeremy E. Allnutt

19 Technical Challenges of Integration of Space and Terrestrial Systems 461
John L. Walker and Chris Hoeber

20 Satellite Communications Regulatory, Legal and Trade Issues 507
Gérardine M. Goh Escolar

21 Trends and Future of Satellite Communications 533
Joseph N. Pelton

Section 2 Satellite Precision Navigation and Timing Section 559

22 Introduction to Satellite Navigation Systems 561
Joseph N. Pelton and Sergio Camacho-Lara

23 Global Navigation Satellite Systems: Orbital Parameters, Time and Space Reference Systems and Signal Structures 573
Rogerio Enríquez-Caldera

24 International Committee on GNSS 603
Sergio Camacho-Lara

25 Current and Future GNSS and Their Augmentation Systems 617
Sergio Camacho-Lara

Volume 2

Section 3 Space Remote Sensing 655

26 Introduction and History of Space Remote Sensing 657
Scott Madry

27 Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing 667
M. J. Rycroft

28 Astronaut Photography: Handheld Camera Imagery from Low Earth Orbit 683
William L. Stefanov, Cynthia A. Evans, Susan K. Runco, M. Justin Wilkinson, and Kimberly Willis

29 Electro-optical and Hyper-spectral Remote Sensing 729
Scott Madry and Joseph N. Pelton

30 Operational Applications of Radar Images 739
Vern Singhroy

31 Lidar Remote Sensing 757
Juan Carlos Fernandez Diaz, William E. Carter, Ramesh L. Shrestha, and Craig L. Glennie

32 Digital Image Acquisition: Preprocessing and Data Reduction 809
Siamak Khorram, Stacy A. C. Nelson, Halil Cakir, and Cynthia F. van der Wiele

33 Digital Image Processing: Post-processing and Data Integration 839
Siamak Khorram, Stacy Nelson, Halil Cakir, and Cynthia Van Der Wiele

34 Remote Sensing Data Applications 865
Haruhisa Shimoda

35 Geographic Information Systems and Geomatics 935
Jesus A. Gonzalez

Section 4 Space Systems for Meteorology 955

36 Introduction to Space Systems for Meteorology 957
Joseph N. Pelton, Scott Madry, and Sergio Camacho-Lara

37 United States Meteorological Satellite Program 967
Sergio Camacho-Lara, Scott Madry, and Joseph N. Pelton

38 EUMETSAT Geostationary Meteorological Satellite Programs 991
Declan Murphy

39 International Meteorological Satellite Systems 1021
Sergio Camacho-Lara, Scott Madry, and Joseph N. Pelton

Section 5 Spacecraft Bus and Ground Systems 1043

40 Overview of the Spacecraft Bus 1045
Tarik Kaya and Joseph N. Pelton

41 Telemetry, Tracking, and Command (TT&C) 1067
Arthur Norman Guest

42 Lifetime Testing, Redundancy, Reliability and Mean Time to Failure 1079
Joseph N. Pelton

43 Ground Systems for Satellite Application Systems for Navigation, Remote Sensing, and Meteorology 1095
Scott Madry, Joseph N. Pelton, and Sergio Camacho-Lara

44 Common Elements versus Unique Requirements in Various Types of Satellite Application Systems 1111
Joseph N. Pelton and Scott Madry

Section 6 Launch Systems and Launch-Related Issues 1129

45 Launch Vehicles and Launch Sites 1131
Joseph N. Pelton

46 Orbital Debris and Sustainability of Space Operations 1145
Heiner Klinkrad

47 Space Weather and Hazards to Application Satellites 1175
Michael J. Rycroft

Section 7 Appendices **1195**

A1 The World's Launch Sites 1197
Arthur N. Guest and Joseph N. Pelton

A2 Major Launch Systems Available Globally 1207
Arthur N. Guest and Joseph N. Pelton

A3a Global Communications Satellite Systems 1221
Joseph N. Pelton

A3b US Domestic Communications Satellite Systems 1225
Joseph N. Pelton

The Authors



Jeremy E. Allnutt Department of Electrical & Computer Engineering, George Mason University, Fairfax, VA, USA

Jeremy Allnutt earned his B.Sc. and Ph.D. in electrical engineering from the University of Salford, UK, in 1966 and 1970, respectively. From 1970 to 1977, he was at the Appleton Laboratory in Slough, England, where he ran propagation experiments with the US satellite ATS-6 and the European satellites SIRIO and OTS. In 1977, he moved to BNR, now Nortel, in Ottawa, Canada, and worked on satellite and rural communications projects before joining the International Telecommunications Satellite Organization (INTELSAT) in Washington, DC, in 1979. Jeremy Allnutt spent 15 years at INTELSAT in various departments. During this period, he ran experimental programs in Europe, Asia, Africa, North and South America, Australia, and New Zealand, finishing as chief, Communications Research Section. Jeremy Allnutt spent 1 year as professor of telecommunications systems at the University of York, England, and then joined the Northern Virginia Center of Virginia Tech in 1986, where he later ran the master's program in ECE, as well as being on the team that designed and set up the masters in information technology program. In August of 2000, he moved to George Mason University with dual appointments: director of the new masters in telecommunications

program and professor in the ECE department. In August 2009, he became director of the new MS in computer forensics program at George Mason University. Jeremy Allnutt has published over 100 papers in conferences and journals and written three books, most in his special field: radiowave propagation. He is a fellow of the UK IET (formerly the IEE) and a fellow of the US IEEE.



Michel Bousquet Institut Supérieur de l'Aéronautique et de l'Espace (ISAE), Toulouse Cedex 4, France

Prof. Michel Bousquet manages the academic and research programs in satellite communications and navigation at ISAE, the French Aerospace Engineering Institute of Higher Education. He chairs the Scientific Board of TeSA (www.tesa.fr), a cooperative research lab on aerospace communications and navigation. With research interest covering many aspects of satellite systems, he participates in many national and international RR&DD programs (French DoD, CNES, ESA, COST, FPs, SatNex NoE). Prof. Bousquet has authored many papers and books including the widely used *Satellite Communication Systems* and sits on several boards of international conferences and journals to promote RR&DD results.



Halil Cakir Air Quality Analysis Group/AQAD/OAQPS, US Environmental Protection Agency, Research Triangle Park, NC, USA

Dr. Halil I. Cakir attended one of Turkey's top universities, the University of Istanbul where he earned a bachelor's degree in forestry. Following his undergraduate program, he was awarded a complete scholarship to study abroad in the USA. He received his master's degree in forestry from Clemson University, and a doctoral degree from the College of Natural Resources at North Carolina State University (NCSU). Upon graduation, he remained on staff as a postdoctoral research associate, then as a research assistant professor. At NCSU

Dr. Cakir's academic focus has been in the geospatial sciences and their application to natural resource issues. As a research associate at NCSU, he worked on and then led increasingly complex and multidisciplinary research projects for the Environmental Protection Agency, Water Resources Research Institute, Department of Defense, and various state and local governments. His research has advanced the geospatial sciences and has two provisional patents for two new image processing techniques and one new change detection technique. These techniques allow users to sharpen the spatial resolution of multispectral imagery to equal that of a more spatially precise geographically coincident panchromatic black and white image. The techniques retain the integrity of the multispectral characteristics of the imagery so that it can be used in most natural resource and environmental applications.

Dr. Cakir maintains an expertise in applied research as well. He has authored peer-reviewed publications, book chapters, and technical reports. He's also served as a reviewer for landmark professional publications like PE&RS.

Dr. Cakir is currently employed in the Office of Air and Radiation at the US Environmental Protection Agency.



Sergio Camacho-Lara Centro Regional de Enseñanza de Ciencia y Tecnología del Espacio para América Latina y el Caribe (CRECTEALC), Tonantzintla, Puebla, México

Dr. Sergio Camacho-Lara is the secretary general of the Regional Centre for Space Science and Technology Education for Latin America and the Caribbean. He was director of the United Nations Office for Outer Space Affairs and chief of the Space Applications Section and the Committee Services and Research Section in the same office. He worked on the organization of the Third United Nations Conference on the Exploration and Peaceful Uses of Outer Space (UNISPACE III) and on implementing its recommendations, including the establishment of the International Committee on GNSS. Prior to joining the United Nations, he carried out research on the interaction of electromagnetic radiation with matter at the *Instituto de Geofísica, Universidad Nacional Autónoma de México*. He obtained his Ph.D. from the University of Michigan.



William E. Carter NSF National Center for Airborne Laser Mapping (NCALM)/ Department of Civil and Environmental Engineering, University of Houston, Houston, TX, USA

William E. (Bill) Carter is a research professor at the University of Houston and a co-PI for the National Science Foundation (NSF) National Center for Airborne Laser Mapping (NCALM). From 1996 to 2010, he was an adjunct professor at the University of Florida (UF), where he taught courses in geodesy and conducted research on advanced geodetic techniques. Prior to joining UF, Bill was chief of the Geosciences Laboratory, NOAA, and led research programs in Very Long Baseline Interferometry (VLBI), absolute gravimetry, and GPS. Bill has coauthored two books and more than a hundred technical papers.



Denis Curtin Eagle Group Partners, Rockville, MD, USA

Dr. Curtin retired from XTAR LLC in July 2010 and is now a communications satellite consultant. Before his retirement he was the COO of XTAR, the joint venture between Loral Space & Communications, Ltd. and HISDESAT Servicios Estrategicos, S.A., providing commercial X-band satellite services to the US and allied governments since October 2003. He was the executive VP, Engineering and Operations, from July 2001 to October 2003. Prior to that from January to July 2001, he led the Loral team that negotiated the joint venture.

Previously, Dr. Curtin served as senior VP, Engineering and Operations, for Loral ORION and as EVP, Loral Cyberstar Broadband Systems. He was responsible for the technical design of the ORION satellites and was instrumental in the formation of the ORION partnership. Before joining ORION, Dr. Curtin held a series of progressively senior engineering and management positions at COMSAT Laboratories, COMSAT General, and COMSAT, culminating in senior director, Satellite Engineering, responsible for all COMSAT's satellite engineering.

Dr. Curtin has an M.S. in physics and a Ph.D. in mechanical engineering. He is a fellow of the American Institute of Astronautics and Aeronautics (AIAA) and a senior member of the Institute of Electrical and Electronics Engineers (IEEE). In 2006, he was named the recipient of the AIAA Aerospace Communications Award, presented for outstanding contributions in the field of aerospace communications. In March 2009, he was inducted into the Society of Satellite Professionals International (SSPI) Hall of Fame. In 2010, he was appointed to the Board of Directors of XTAR LLC. He is the coauthor of the article on "Communications Satellites" in the 10th edition of the McGraw Hill *Encyclopedia of Science and Technology* and has also published extensively on satellite technology and holds a patent on an infrared transparent solar cell.



Rogerio Enríquez-Caldera Centre for Space Science and Technology Education for Latin America and the Caribbean, México Campus (CRECTEALC), National Institute of Astrophysics, Optics and Electronics (INAOE), Coordinación de Electrónica, Tonantzintla, Puebla, Mexico

Rogerio Enriquez-Caldera is an engineering researcher at the National Institute of Astrophysics, Optics and Electronics of Mexico since 2000. He received his Ph.D. from the University of New Brunswick, Canada. He has ample experience in the area of astronomical instrumentation including control engineering, high sensitive detectors, and optical and radio telescopes. His expertise is related to areas of electrical engineering, including GPS receiver's software and low noise amplifiers and correlators for very high dynamics platforms in the presence of high levels of noise as well as information systems. He also works for the Mexico Campus of the Regional Centre for Space Science and Technology Education for Latin America and the Caribbean teaching in the fields of GNSS, State Estimation for Nonlinear Dynamics Navigation and Tracking Systems, and developing an Aero-Space Flight Simulator for formation flying satellites.



Cynthia A. Evans Code KT, Astromaterials Research and Exploration Science Directorate, NASA Johnson Space Center, Houston, TX, USA

Cynthia A. Evans received her Ph.D. in earth science from Scripps Institution of Oceanography in 1983. She joined the Space Shuttle Earth Observations Office at NASA Johnson Space Center (JSC) in 1989 and has participated in or led several Earth Observations experiments from the Shuttle, Shuttle-Mir, and International Space Station Programs. She also managed JSC's Image Science and Analysis Lab. Currently with the Astromaterials Acquisition and Curation office, her research interests include the use of remotely sensed data for investigation of coastal changes, including those related to human activities, and documenting regional responses to dynamic events and global climate change. In addition to training astronauts in Earth science, she participates in education and public outreach activities using astronaut photography.



Juan Carlos Fernandez Diaz NSF National Center for Airborne Laser Mapping (NCALM)/Department of Civil and Environmental Engineering, University of Houston, Houston, TX, USA

Juan Carlos was born in Tegucigalpa, Honduras, in 1976. His formal education includes a B.S. in electrical engineering and an M.B.A. obtained from universities in Honduras, M.S. and Ph.D. degrees in geosensing systems engineering obtained from the University of Florida. Other interests include aviation, Earth/space science and exploration, and the application of space science and technology to bring progress to developing countries.



Craig L. Glennie NSF National Center for Airborne Laser Mapping (NCALM)/ Department of Civil and Environmental Engineering, University of Houston, Houston, TX, USA

Craig is an assistant professor at the University of Houston and co-PI of the NSF-funded National Center for Airborne Laser Mapping (NCALM). Dr. Glennie was formerly the vice president of engineering for Terrapoint, a LIDAR remote sensing company with offices in Canada and the USA. He has been active in the design, development, and operation of kinematic remote sensing systems for 13 years. Craig holds a B.Sc. and a Ph.D. in geomatics engineering from the University of Calgary and is a registered professional engineer in Alberta, Canada.



Daniel R. Glover International Space University, Bay Village, OH, USA

Daniel R. Glover spent a 27-year career working as an electrical engineer at NASA's Glenn Research Center, Jet Propulsion Laboratory, and Johnson Space Center. He has worked on various projects including launch vehicles, shuttle experiments, satellite communications protocol research, image data compression, planetary spacecraft conceptual design, space suit design, software management, systems engineering, and strategic planning. He earned his MSEE and Ph.D. degrees from the University of Toledo and an MBA degree from Cleveland State University. Dr. Glover is a member of the faculty of the International Space University. He served as an engineering duty officer in the US Naval Reserve. He has also worked as an independent consultant.



Gérardine M. Goh Escolar International Court of Justice, United Nations, The Hague, The Netherlands

Dr. Gérardine Goh Escolar is associate legal officer at the International Court of Justice. She is also research scholar at the International Institute of Air and Space Law, Leiden University, the Netherlands. Previously, she was legal counsel and project manager at the German Aerospace Center in Bonn, Germany. In this role, she served as counsel on international transactions, assistant editor on a legal commentary, and legal advisor to the German delegation to the UN Committee on the Peaceful Uses of Outer Space. Prior to that position, she was counsel at a satellite company near Berlin, Germany. A native of Singapore, Gérardine has been awarded the Bachelor of Laws (Honors) from the National University of Singapore, the Master of Laws in Public International Law from University College London, and holds the Doctor of Laws degree from Leiden University. She was appointed assistant secretary to the Board of Directors of the International Institute of Space Law, and was a contributing member to the International Academy of Astronautics' study on space debris remediation. She is a fellow of the International Association for the Advancement of Space Safety and an associate of the Committee on Space Research of the International Council for Science. She is also a member of the International Law Association, the European Centre for Space Law and Women in Aerospace Europe. Gérardine continues to be heavily involved in academia, and has held the positions of assistant professor at Cologne University, research director at The Hague University of Applied Sciences, and lecturer at Webster University. The author of more than 45 research publications, Gérardine's research interests include dispute settlement, international law, international humanitarian law, law and technology, human rights, and international environmental law.



Jesus A. Gonzalez National Institute of Astrophysics, Optics, and Electronics (INAOE)/Regional Center for Space Science and Technology Education for Latin America and the Caribbean (CRECTEALC) Campus Mexico, Tonantzintla, Puebla, Mexico

Jesus A. Gonzalez obtained his Ph.D. in computer science and engineering from The University of Texas at Arlington, Texas, in 2001. He is currently a researcher and professor in the Computer Science Department at the National Institute of Astrophysics, Optics, and Electronics, México. He also holds the academic coordinator position at the Regional Centre for Space Science and Technology Education for Latin-America and the Caribbean, Campus Mexico, affiliated to the United Nations.



Arthur N. Guest International Space University, San Francisco, USA

Arthur Guest is a space system engineering consultant who specializes in the principles of system architecting and is located in San Francisco, USA. Arthur graduated from the Massachusetts Institute of Technology with a master's in aeronautics and astronautics and is a graduate of the Masters of Space Studies Program at the International Space University. He is currently serving as the space system engineering department chair for the International Space University's 2011 Space Studies Program in Graz, Austria.



Ramesh Gupta LightSquared, Reston, VA, USA

Dr. Ramesh K. Gupta is vice president of Satellite Engineering and Operations group at LightSquared where he has supported the development of the next

generation hybrid satellite, its launch, in-orbit testing, and operations. He has more than 25 years experience in the satellite and wireless communications industries. He has held senior management positions as vice president of advanced business and technology at AMCOM Communications and managing director at COMSAT Laboratories and Lockheed Martin Global Telecommunications. His work has included the integration of large satellite/wireless systems, business planning, and strategic management in a high-technology business environment.

He holds a Ph.D. and an M.S. degree in electrical engineering from the University of Alberta, Canada; an M.B.A. degree from the Wharton Business School at the University of Pennsylvania; and a B.S. degree (with Honors) in electronics and communications engineering from India. He has published extensively, and holds four US patents. He has received many honors and awards including Alberta Government Telephone's Centennial Fellowship for graduate research in Telecommunications and COMSAT Laboratories' Research Award. He was corecipient of the Best Paper Award at the 9th International Digital Satellites Communications Conference in Copenhagen, Denmark, and the 29th AIAA International Communications Satellite Systems Conference, 2011, in Nara, Japan. He has served as an adjunct associate professor of strategic management and technology planning at the University of Maryland, College Park, MD. Dr. Gupta has published more than 75 papers on satellite and wireless RF technology and systems in AIAA and IEEE conferences and technical journals. Dr. Gupta is a fellow of the IEEE.



Henry R. Hertzfeld Space Policy Institute, Elliott School of International Affairs, The George Washington University, Washington, DC, USA

Dr. Henry R. Hertzfeld is a research professor of space policy and international affairs at the Space Policy Institute, Center for International Science and Technology Policy, Elliott School of International Affairs, George Washington University.

He is an expert in the economic, legal, and policy issues of space and advanced technological development. Dr. Hertzfeld has served as a senior economist and policy analyst at both NASA and the National Science Foundation, and is a consultant to both US and international agencies and organizations. He is the coeditor of *Space Economics* (AIAA 1992), as well as many articles on the economic and legal issues concerning space and technology. Dr. Hertzfeld has a B.A. from the University of Pennsylvania, an M.A. from Washington University, and a Ph.D. in economics from Temple University. He also holds a J.D. degree from the George Washington University and is a member of the Bar in Pennsylvania and the District of Columbia. He can be contacted at: hhertzfeld@law.gwu.edu.



Chris Hoeber Space Systems/Loral, Palo Alto, CA, USA

Christopher F. Hoeber is senior vice president, Program Management & Systems Engineering at Space Systems/Loral (SS/L). With more than 35 years of industry experience, he leads the systems engineering and program management groups and oversees all of the company's research and development programs. Mr. Hoeber has a broad base of experience in systems test and engineering and program and functional management and was the leader of the systems engineering team that developed SS/L's industry leading geostationary satellite platform, the modular 1300. Before his current position, Mr. Hoeber was vice president of business development for SS/L and before that he was chief engineer of the company.



Takashi Iida Tokyo Metropolitan University, Hachioji, Tokyo, Japan

Takashi Iida received B.E., M.E., and Dr.Eng. degrees from the University of Tokyo, Tokyo, Japan, in 1966, 1968, and 1971, all in electronic engineering. He joined Radio Research Laboratories (RRL), Ministry of Posts and Telecommunications (MPT), in 1971. He was with National Space Development Agency and involved in the CS-3 satellite development, 1987–1989. He was director of Space Communications Division, Communications Research Laboratory (CRL) (former RRL), MPT, 1989–1991. He was visiting professor of University of Colorado at Boulder, 1991–1992. He was director general of CRL, 1999–2001, and President of CRL, Incorporated Administrative Agency, 2001–2004. He was executive director of JAXA (Japan Aerospace Exploration Agency), 2005–2007. He was invited advisor/distinguished researcher of National Institute of Information and Communications Technology (NICT, former CRL), Incorporated Administrative Agency, 2007–2009. He was research professor, Tokyo Metropolitan University, 2009–2010. Dr. Iida is now visiting professor of Tokyo Metropolitan University and Tokyo University of Technology. He is editor and author of the books entitled *Satellite Communications – System and Its Design Technology*, Ohmsha/IOS Press, 2000, and *Satellite Communications in the 21st Century: Trends and Technologies*, AIAA, 2003. He received the AIAA Aerospace Communications Award, 2002, and Hall of Fame Award from SSPI, 2003. Dr. Iida is IEEE life fellow and AIAA fellow.



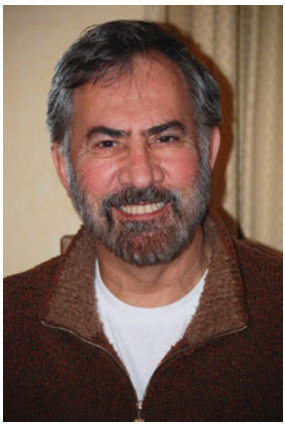
Ram S. Jakhu Institute of Air and Space Law, McGill University, Montreal, QC, Canada

Prof. Ram S. Jakhu has over 25 years of experience in space-related fields. He is holding a position of associate professor at the Institute of Air and Space Law, Faculty of Law, of McGill University in Montreal, Canada, where he teaches several courses covering numerous subjects including public international law, international and national space law and policy, international trade, export controls, space applications, space commercialization, telecommunications, etc. From January 1995 to December 1998, Dr. Jakhu served full-time the International Space University, Strasbourg, France, holding various titles, including a professor and the first director of the Master of Space Studies program. He has authored more than 60 articles in several reputed journals and edited two books: *Space Safety Regulations and Standards* and *National Regulation of Space Activities*. He has presented numerous papers and expert legal opinions at various conferences around the world and participated in several space-related studies. Prof. Jakhu is a fellow as well as the chairman of the Legal and Regulatory Committee of the International Association for the Advancement of Space Safety. He is a member of the Board of Directors of the International Institute of Space law of the International Astronautical Federation (Paris). In 2007, he received a "Distinguished Service Award" from the International Institute of Space Law for his significant contribution to the development of space law. He holds a Doctor of Civil Law (Dean's Honours List) degree in space law from McGill University; a Master of Law (LL.M.) degree in the field of Air and space law from McGill University. In addition, he has earned LL.M. (in Public and Private International Law), LL.B. (in Laws of India), and Bachelor of Arts (in Economics and Political Science) degrees from Punjab University, Chandigarh, India.



Tarik Kaya Mechanical and Aerospace Engineering Department, Carleton University, Ottawa, ON, Canada

Tarik Kaya has been working as a professor at the Mechanical and Aerospace Engineering Department of Carleton University since 2002. He received his Ph.D. from ENSAE, Toulouse, France in 1993. Before joining Carleton University, he worked as a research associate at NASA Goddard Space Flight Center. Dr. Kaya's current research interests includes mathematical modeling of two-phase heat transfer systems (heat pipes and loop heat pipes) for spacecraft thermal control and miniaturization of the heat pipe technology for electronics packaging.



Siamak Khorram Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA

Dr. Khorram received M.S. degrees in engineering and ecology from the University of California (UC) at Davis. He received a Ph.D. in 1975 under a joint program from the University of California at Berkeley and Davis with emphasis in remote sensing and image processing.

From 1974 to 1980, he served as the principal scientist at the Space Sciences Laboratory at the University of California in Berkeley. He joined the faculty in North Carolina State University (NCSU) in 1980. At NCSU, he has served as the Principal Investigator for over 60 major research projects. His research projects have focused on environmental remote sensing, image processing, and geospatial information technology.

In 1982, he established the Computer Graphics Center at NCSU as a university-wide officially recognized center involved in research and training in spatial information technology and special purpose computing. In 1997, he changed the name of the Computer Graphics Center to the Center for Earth Observation (CEO) with the same mission. In 1986 and 1987, he served as a NASA-ASEE fellow at NASA Ames Research Center and as a summer faculty at Stanford University, California.

Since 1988, he has served as a faculty member at the International Space University (ISU). Dr. Khorram has worked with well over 250 educators and world-renowned experts from over 30 countries and has participated in educating over 2,000 multidisciplinary graduate students from 70 countries worldwide.

In 1995 and 1996, he served as the first dean and vice president for academic programs at ISU in Strasbourg, France. In this capacity, he played a major role in establishing academic relationships between ISU and major space organizations such as European, French, Japanese, Russian, German, Austrian, and Indian space agencies. Subsequent to his position as dean, Dr. Khorram served as the principal advisor to the president in 1996 and as the chair of the Academic Council and chair

of the ISU's 23 Affiliates Campuses Network worldwide. He currently serves as a member of the University's Board of Trustees.

He has served as the major professor for over 30 Ph.D. and M.Sc. students. He is the author of over 200 publications in peer-reviewed journals, conference proceedings, and major technical reports. He is a member of several professional and scientific societies.

He has delivered keynote speeches on "Information Technology" in the IEEE International Symposia on Computers and Communications in Morocco in 2008 and on "Geospatial Information Science and Technology" in Italy in 2010.



Heiner Klinkrad European Space Agency ESA/ESOC, Darmstadt, Germany

Heiner Klinkrad graduated from the Braunschweig University of Technology (TUBS) in aeronautical engineering in 1980, and he received his Ph.D. from the same university in 1984. In 1980, he joined the European Space Agency, where today he is head of the Space Debris Office at the European Space Operations Centre (ESOC) in Darmstadt, Germany. In his current position he is ESA's focal point and senior advisor for space debris matters, and he represents ESA, for instance, in the multinational Inter-Agency Space Debris Coordination Committee (IADC).

Heiner is a member of the International Academy of Astronautics (IAA), a fellow of AIAA, and he has served as a member or chairperson of working groups and panels of AIAA, COSPAR, ECSS, IAA, IADC, IAG, ISO, and UNCOPUOS. He is a professor at the Braunschweig University of Technology since 2009, and he published a textbook *Space Debris – Models and Risk Analysis* in 2006.



Scott Madry International Space University, Chapel Hill, North Carolina, USA

Scott Madry is a specialist in the practical applications of satellite data, including remote sensing, GPS, and geographic information systems. His research interests are in regional cultural and environmental analysis, monitoring, and modeling, and he has conducted research in North America, Europe, and Africa. He is on the faculty of the International Space University and is the program manager of ISU's Southern Hemisphere Summer Space Program. He is a research associate professor of archaeology at the University of North Carolina at Chapel Hill and is founder and president of Informatics International, Inc., an international consulting company located In Chapel Hill, NC, USA.



Peter Marshall Royal Television Society, England, UK

After working for the BBC as a journalist and editor, Peter Marshall was a pioneer in developing the use of international satellites for worldwide TV news coverage and distribution. In 1986, he joined INTELSAT in Washington, DC as the first Director of Broadcast Services. Then as competition in the global satellite industry began to emerge, he moved on to the private sector in 1989 as president of the US-based satellite services company, Keystone Communications, which became a leader in worldwide satellite distribution services. Keystone was acquired by France Telecom in 1996 after which Peter Marshall served as a member of the Board of the France Telecom subsidiary, GlobeCast. He is a past president of the Society of Satellite Professionals International (SSPI) and was elected to the Society's "Satellite Hall of Fame" in 2002 for his pioneering work on the development of global satellite services for broadcasting. He was chairman of the Britain's Royal Television Society (RTS) in 1985 and he continues to serve as a director of the Arthur C. Clarke Foundation.



Declan Murphy Met Éireann (the Irish Meteorological Service), Glasnevin Hill, Dublin 9, Ireland

Declan Murphy graduated from University College, Cork, Ireland, with a master's degree in experimental physics in 1973. His career in Met Éireann (Ireland's National Meteorological Service) included roles as an operational weather forecaster and developer of meteorological software before taking a management position in 1980. He became director of Met Éireann in 1989 and represented Ireland in many international meteorological organizations before his retirement from that position in 2009. He served as chairman of the Council of EUMETSAT from 2004 to 2008. He was lead author of the updated "History of EUMETSAT" published in 2011.



Stacy A. C. Nelson Center for Earth Observation, North Carolina State University, Raleigh, NC, USA

Dr. Stacy Nelson is currently an associate professor within the Department of Forestry and Environmental Resources, and the Center for Earth Observation at North Carolina State University. Dr. Nelson received a B.S. from Jackson State University, M.A. from College of William and Mary's Virginia Institute of Marine Sciences, and Ph.D. from Michigan State University. His research centers around the use of remote sensing and GIS technologies to address questions of land use/cover change on aquatic systems at both regional and local scales. He has worked with several federal and state agencies including the Earth Systems Science Office at the Stennis Space Center in Mississippi, the NASA-Regional Earth Science Applications Center (RESAC), and the MI and NC Department of Environmental Quality. Dr. Nelson currently teaches as part of the spatial analyses curricula related to GIS science and technologies at NC State University and is active in several professional societies.



Joseph N. Pelton Former Dean, International Space University, Arlington, Virginia, USA

Dr. Joseph N. Pelton is an award winning author/editor of over 30 books and over 300 articles in the field of space systems. These include the four book series: *e-Sphere*, *Future Talk, Future View*, and *Global Talk*, for the latter of which he was nominated for a Pulitzer Prize. He served as Chairman of the Board (1992–1995) and vice president of Academic Programs and Dean (1995–1996) of the International Space University of Strasbourg, France. He is currently a member of the ISU faculty and series editor for a number of books on behalf of the university. He is also the director emeritus of the Space and Advanced Communications Research Institute (SACRI) at George Washington University. This Institute, which he headed from 2005 to 2009, conducted state-of-the-art research on advanced satellite system concepts and space systems. From 1988 to 1996, Dr. Pelton served as director of the Interdisciplinary Telecommunications Program at the University of Colorado Boulder, which at that time was the world's largest graduate level telecommunications program. Prior to that, he held a number of positions at Intelsat and Comsat including serving as director of Strategic Policy and director of Project Share for Intelsat.

Dr. Pelton is a fellow of the International Association for the Advancement of Space Safety (IAASS), a member of its Executive Board and chairman of its Academic Committee. He is also the Executive Editor of the IAASS publication series. He is also acting president of the International Space Safety Foundation of the USA as well as the former president of the Global Legal Information Network. Dr. Pelton was the founder of the Arthur C. Clarke Foundation and remains as the vice chairman of its Board of Directors. This Foundation honors Sir Arthur Clarke, who first conceived of the Communications Satellite (as of 1945). Dr. Pelton was elected to full membership in the International Academy of Astronautics in 1998 and was awarded in 2000 the Sir Arthur Clarke Award for lifetime achievement in the field of satellite communications. He was elected to the Hall of Fame of the Society of Satellite Professionals International in 2001, an honor only extended to

some 50 people in the field. In 2004, he was elected an associate fellow of the American Institute of Aeronautics and Astronautics. His degrees in physics and International Relations are from the University of Tulsa, New York University (NYU), and Georgetown University.



Bruno Perrot SES, Betzdorf, Luxembourg

Bruno Perrot, senior manager Fleet Planning in SES, has a broad background, encompassing the engineering, services, and business sides of a worldwide satellites operator, leader in its domain. With 25 years in the field of space telecommunications industry, he has experienced first hand the dynamics of the rapidly evolving and changing global telecommunications marketplace.

The early years in Mr. Perrot's career were spent with Aerospatiale, France, a period during which he became a key person of the communication engineering staff. He joined Alenia Aerospazio, Italy, in 1991, as a payload manager, and then Telespazio in 1999. From 2000, he worked for SES ASTRA and moved to SES Global in 2005. Since 2008, he is in charge of the European Fleet Planning in SES.

Bruno graduated at ENSEA of Paris where he earned an honors degree in Telecommunication Engineering application and holds a management degree at INSEAD of Fontainebleau.

As a member of the Technical Committees on Communication Satellites of the AIAA and on the Ka-band & Broadband Communication Conference, he has chaired and served on numerous industry panels, seminars, and roundtable discussions across the globe and is fluent in English, French, Italian, and Spanish.



Susan K. Runco Code KX, Astromaterials Research and Exploration Science Directorate, NASA Johnson Space Center, Houston, TX, USA

Susan K. Runco completed her M.S. in oceanography and meteorology at the Naval Postgraduate School in 1986. She is currently the principal investigator for the Crew Earth Observation Payload on the International Space Station at NASA Johnson Space Center in Houston, Texas, USA. She has provided Earth science training to astronauts and participated in Crew Earth Observations since 1988. As PI, her interests include developing imaging techniques and imagery collections for broadening utilization of astronaut photography in the areas of Earth science, education, and public awareness of Earth and space.



Michael J. Rycroft Cambridge Atmospheric, Environmental and Space Activities and Research (CAESAR) Consultancy, Cambridge, UK

Prof. Michael Rycroft is visiting professor at Cranfield University, UK, and part-time faculty member at the International Space University, Strasbourg, France. Previously he was Head, Atmospheric Sciences Division, British Antarctic Survey, Cambridge; Lecturer in Physics Department at Southampton University; and post-doctoral NAS/NRC associate at NASA Ames Research Center, California. He obtained his honorary D.Sc. from De Montfort University, Ph.D. from Cambridge, and B.Sc. from Imperial College London. Prof. Rycroft has done research on solar-terrestrial physics, ionospheric and magnetospheric physics, and atmospheric studies. He is editor of the *Cambridge Encyclopedia of Space* (1990), *Journal of Atmospheric and Solar-Terrestrial Physics* (1989–1999), and *Surveys in Geophysics* (2002–present). He is author of more than 200 scientific publications and editor of more than 40 books and special issues of journals.



Haruhisa Shimoda Earth Observation Research Center, Aerospace Exploration Agency, Tsukuba, Japan

Haruhisa Shimoda (Member, IEEE) received the B.S, M.E., and Ph.D. degrees from the University of Tokyo. After he got his Ph.D., he joined Tokai University in 1972 as an assistant professor of Department of Electro-Photo-Optics of the Faculty of Engineering. He became an associate professor in 1974. In the same year, he also joined Tokai University Research & Information Center as a senior researcher. In 1985, he became a professor in the same faculty. In 1994, he also joined National Space Development Agency (now Japan Aerospace Exploration Agency, JAXA) as an invited scientist, and has been working as the program scientist of ADEOS, and later also for GCOM. In 1999, he became the deputy director of Tokai University Research & Information Center, and in 2000, he became the director of Tokai University Space Information Center. He has been engaged in the field of remote sensing from 1974. His main achievements are development of remote sensing image analysis system including both hardware and software, developments of high-accuracy classification algorithms, development of IMG on ADEOS, etc. Also, he was elected as the Technical Commission President of ISPRS Commission 8 (Remote Sensing Applications) during 2008–2012.



Ramesh L. Shrestha NSF National Center for Airborne Laser Mapping (NCALM)/Department of Civil and Environmental Engineering, University of Houston, Houston, TX, USA

Ramesh L. Shrestha, Ph.D. is a Hugh Roy & Lillie Cranz Cullen University professor at the University of Houston (UH) and leads the GSE graduate research and academic programs. He is also PI and the director for the NSF funded National Center for Airborne Laser Mapping (NCALM) which is jointly operated by UH and the University of California-Berkeley. Dr. Shrestha's main research activities are associated with the application of advanced geodetic and remote sensing techniques, particularly airborne laser swath mapping (ALSM, aka LiDAR) and digital mapping.



Vern Singhroy Canada Centre for Remote Sensing, Natural Recources Canada, Ottawa, Canada

Dr. Vern Singhroy is the chief scientist at the Canada Centre for Remote Sensing, a Centre of Excellence in Earth Observation research and application for the Government of Canada. He is also the principal scientist of the Radar Constellation Mission (RCM) to be launched in 2015, by the Canadian Space Agency.

Dr. Singhroy received his Ph.D. in environmental and resource engineering from the State University of New York, Syracuse. He has published over 300 papers in scientific journals, proceedings, and books. He is also the coeditor of the *Encyclopedia of Remote Sensing* and was the editor in chief of the *Canadian Journal of Remote Sensing*. He is Professor of Earth Observation at the International Space University in Strasbourg, France, since 1998 and an adjunct professor in planetary and earth sciences at the University of New Brunswick and McMaster Universities in Canada. Dr. Singhroy received the prestigious Gold Medal Award for his contribution and impact to Canadian and International Remote Sensing from the Canadian Aeronautics and Space Institute.



Andrew Stanniland Paradigm Secure Communications Ltd, Stevenage, Hertfordshire, UK

Andrew Stanniland, Senior Vice President, Business Development, Sales and Marketing.

Andrew has over 15 years experience in the satellite communications industry, with a background in systems and aeronautical engineering, including development of satcom prospects in Australia and South Korea. He has been involved with the UK's Skynet 5 Milsatcoms program since its inception, initially on the communications technical analysis before leading the Service Design aspects that eventually evolved into the company Paradigm.

During the Skynet 5 proposal and contract negotiation phase, Andrew focused on the contract and financial aspects of the program and was responsible for liaising with investors and satellite insurers to ensure that the complex and groundbreaking PFI deal was bankable.

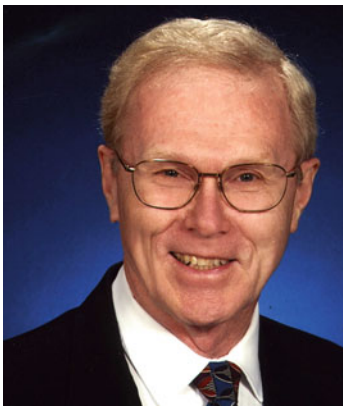
Since the contract signature and formal formation of Paradigm in October 2003, Andrew has been responsible for overseas business development activities, leveraging the capabilities of Paradigm's constellation of Skynet 4 and 5 satellites to sell military satellite communications capacity into NATO, Canada, the Netherlands, France, Germany, Portugal, and, most recently, the US DoD.

He has been the senior VP responsible for the Business Development, Sales and Marketing team, since 2007. He is also an alumnus of the International Space University, having attended the Summer Session Program in 1997 in Houston, TX.



William L. Stefanov Science Applications Research and Development, Code KA/ESCG, Astromaterials Research and Exploration Science Directorate, NASA Johnson Space Center, Houston, TX, USA

William L. Stefanov completed his Ph.D. in geology at Arizona State University in 2000. He is currently chief scientist for the Engineering and Science Contract Group at NASA Johnson Space Center in Houston, TX. In association with the International Space Station (ISS) Program Science Office, he works with instrument science/operation teams to coordinate collection, distribution, and analysis of remotely sensed data from the ISS in response to catastrophic events such as volcanic eruptions, earthquakes, and flooding. His research interests include the use of remotely sensed data for investigation of geohazards, geomorphology, and surface material characterization; mapping of urban riskscapes in the context of global and regional climate change; and assessing the role of humans as geological agents on the landscape. He is an active proponent of geoscience education and public outreach using remotely sensed data.



Dan Swearingen Arlington, VA, USA

Daniel Swearingen is an engineering consultant who worked at Communications Satellite Corporation (COMSAT) for 30 years. After earning degrees at Georgia Tech and Stanford and completing his military service, he worked for 3 years at ITT Telecommunications before joining COMSAT in 1970. After working in the spectrum utilization and special studies departments, he joined the COMSAT's mobile satellite systems group in 1973. There he served as a system design group manager for the first maritime mobile communications satellite system, MARISAT, launched in 1976. The MARISAT standards and protocols he and his group developed became the basis for the first International Maritime Satellite (INMARSAT) system. In 1980–1981, shortly after the INMARSAT was established, he worked with the small startup staff in London to plan the initial space and ground components of the multinational cooperative's system prior to its operational cutover in 1982. In the subsequent years from 1981 to 1996, he served as a member of the Inmarsat technical advisory committee and proposed key system architecture features that were adopted for the new INMARSAT systems. After leaving COMSAT, Mr. Swearingen served for several years as an adjunct lecturer at George Washington University and has been a consultant specializing in communications satellite systems.



Paul T. Thompson University of Surrey, Guildford, Surrey, United Kingdom

Dr. Thompson is a senior researcher in the Centre for Communications and Systems Research (CCSR) at the University of Surrey. He has an extensive background in satellite communications and spent 30 years with British Telecom where he led the Technology Development Division covering a wide range of international communications disciplines. During part of this time he was seconded to the SHAPE Technical Centre in The Hague where he was involved in the development and in-orbit testing of NATO satellites.

Subsequent to his time at BT, Dr. Thompson has worked with ERA Technology, developing a range of radio-communications products and was also the Director of Teledesic, UK.

In addition to research and teaching roles, he currently participates in the standards areas of DVB and the European Telecommunications Standards Institute (ETSI).

Dr. Thompson was the first UK delegate to become chairman of the INTELSAT Board of Governors Technical Committee (BG/T). He is a fellow of the British Interplanetary Society where he played several roles (member of BIS Council 1990–2002, president 1994–1998).

He was a visiting professor of engineering design at the Engineering Science Faculty of the University of Oxford, a role supported by the Academy of Engineering (1993–2001).

He has been a member of the editorial panel of the *International Journal of Satellite Communications* since 1982.

Dr. Thompson is also a senior member of AIAA since 1991.



Cynthia F. van der Wiele Cynthia Van Der Wiele and Associates, LLC, Durham, NC, USA

Cynthia F. Van Der Wiele received the B.S. degree in engineering and the Masters of Landscape Architecture from North Carolina State University; the Masters of Forestry and Masters of Environmental Management with an emphasis in environmental economics and policy from Nicholas School of the Environment, Duke University; and Ph.D. in Community and Environmental Design from North Carolina State University. She has performed landscape change analyses of the Research Triangle Region of North Carolina and the implications for regional planning and conservation initiatives.

She is currently the director of Sustainable Communities Development, Chatham County, NC and a research associate at North Carolina State University. Her research interests include the development of high-accuracy land use/land cover classifications for analysis and formation of evidence-based land use planning policies.



John L. Walker Space Systems/Loral, Union City, CA, USA

John Walker has over 30 years of technical and managerial achievements in the design and development of advanced communication systems operating from ELF to EHF. He has held positions at Lockheed Martin, Lockheed Electronics, and Hughes Aircraft Company responsible for both space and terrestrial developments. He joined Space Systems/Loral (SS/L) in 1996 as the director of RF Electronics responsible for the development, design, manufacturing, and test of space-borne RF electronics equipment. He progressed to the director of advanced development responsible for the end-to-end communication systems activities within SS/L. Dr. Walker most recently led the development and execution of the first two-way Ground Based Beam Forming system from concept through on-orbit integration, verification, and deployment.



M. Justin Wilkinson Code KX/ESCG, Astromaterials Research and Exploration Science Directorate, NASA Johnson Space Center, Houston, TX, USA

M. Justin Wilkinson was born in South Africa and holds a Ph.D. in geomorphology from the University of Chicago. Since 1988, he has held the position of astronaut trainer in geography, geology, and geomorphology with the Crew Earth Observations payload at the NASA Johnson Space Center in Houston, Texas, USA. His research interests include fluvial and desert geomorphology, especially the interface between geomorphology and sedimentology, the role of landscapes in species evolution, and geomorphic analogs for planetary geology. His teaching interests have resulted in the publication of books of astronaut imagery with National Geographic and a bilingual atlas of Costa Rica, for which he was awarded NASA's Public Service Medal. He has a patent in the area of automated identification of fluvial landscapes.



Kimberly Willis Code KX/ESCG, Astromaterials Research and Exploration Science Directorate, NASA Johnson Space Center, Houston, TX, USA

Kimberly J. Willis completed her M.S. in physical science, with a concentration in geology, at the University of Houston–Clear Lake. She began her career at the NASA Johnson Space Center in Houston, TX, over 26 years ago where she first worked in the Lunar Laboratory with samples returned from the Apollo missions. Kim transitioned into Earth Observations where she held progressively more responsible positions. She is currently the manager for Astromaterials Curation, Education, and Crew Earth Observations contractor personnel. In addition to training astronauts in Earth science, she also holds an adjunct faculty position at the University of Houston–Clear Lake in the School of Science and Computer Engineering, where she teaches the fundamentals of earth science.

Section 1

Satellite Communications

Satellite Applications Handbook: The Complete Guide to Satellite Communications, Remote Sensing, Navigation, and Meteorology

1

Joseph N. Pelton, Scott Madry, and Sergio Camacho Lara

Contents

Introduction 4
The Evolution of Commercial Satellite Applications 9
What Does the Term "Satellite Applications" Mean and Why Consider It in a Unified Way? 13
Common Elements of Applications Satellites 13
Organization and Effective Utilization of the Satellite Applications Handbook 15
Notes 17
References 17

Abstract

This chapter introduces what is meant by the term "applications satellite" and addresses why it makes sense to address the four main space applications in a consolidated reference work. This handbook also provides a multidisciplinary approach that includes technical, operational, economic, regulatory, and market perspectives. These are key areas whereby applications satellite share a great deal. This can be seen in terms of spacecraft systems engineering, in terms of launch services, in terms of systems economics, and even in terms of past, present, and future market development. This is not to suggest that there are not important

J.N. Pelton (✉)
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: joepelton@verizon.net

S. Madry
International Space University, Chapel Hill, North Carolina, USA
e-mail: Scottmadry@mindspring.com

S. Camacho Lara
CRECTEALC, Luis Enrique Erro No. 1, Tonantzintla, San Andres, Cholula, Puebla, C.P. 72840, Mexico
e-mail: sergio.camacho@inaoep.mx

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_91, © Springer Science+Business Media New York 2013

technical and operational differences with regard to communications satellites, remote sensing satellites, global navigation satellites and meteorological satellites. Such differences are addressed in separate sections of the handbook.

Yet in many ways there are strong similarities. Technological advances that come from one type of applications satellite can and often are applied to other services as well. The evolution of three-axis body-stabilized spacecraft, the development of improved designs for solar arrays and battery power systems, improved launch capabilities, and the development of user terminal equipment that employs application-specific integrated circuits (ASIC) are just some of the ways the applications satellites involve common technology technologies and on a quite parallel basis.

These applications satellites provide key and ever important services to humankind. Around the world, people's lives, their livelihood, and sometimes their very well-being and survival are now closely ties to applications satellites. Clearly the design and engineering of the spacecraft busses for these various applications satellite services as well as the launch vehicles that boost these satellites into orbit are very closely akin. It is hoped that this integrated reference document can serve as an important reference work that addresses all aspects of application satellites from A to Z. This handbook thus seeks to address all aspects of the field. It thus covers spacecraft and payload design and engineering, satellite operations, the history of the various types of satellites, the markets, and their development – past, present, and future, as well as the economics and regulation of applications satellites, and key future trends.

Key Words

Applications satellite • Committee on the Peaceful Uses of Outer Space (COPUOS) • Earth observation • Global Navigation Satellite Systems • Launch services • Markets for satellite applications • Military satellite communications • Satellite broadcasting • Satellite communications • Satellite meteorology • Satellite navigation and positioning • Satellite remote sensing • Scientific satellites • United Nations

Introduction

Artificial satellites have now been around for more than a half century. The launch of Sputnik in October 1957 ushered in the space age and confirmed Sir Isaac Newton's theoretical explanation of how an artificial satellite could be launched into Earth orbit. Today the world of satellites can be divided into two broad areas – scientific satellites and applications satellites. Scientific satellites explore and help humanity acquire new information about our world, our solar system, our galaxy, and the great cosmos within which we exist. The scientific satellites explore radiation from the Van Allen Belts to cosmic radiation. They engage in geodesy to measure our Earth and tectonic movements. They study the workings of the Sun and the characteristics of the our Solar System, including the planets, their moons, asteroids, comets, and the Oort cloud well beyond the orbit of Pluto. Astronomical

observatory satellites explore the stars and galaxies and give us a view of the Universe near its beginning as well as of exo-planets in other star systems. This handbook, however, is about the applications satellites that provide practical services to people here on Earth. These are the communications satellites, the remote sensing satellites, the space navigation and positioning satellites, and the meteorological satellites that truly serve humankind.

Thousands of applications satellites have now been launched over the past half century. These practical satellites now represent a huge global industry. These satellites are a part of our everyday lives whether we know it or not. Every time you hear a weather report or every time you use a GPS or Glonass device to navigate your car you are relying on an applications satellite. Services such as worldwide news, satellite entertainment channels, coverage of sporting events, communications to ships at sea or aircraft in the skies, and many more frequently depend on satellites. Farmers now rely on satellites to irrigate their crops, add the right amount of fertilizer, or detect a crop disease. Fishing fleets use satellites to know where to fish. Energy and resource companies employ satellite imaging to know where to dig or drill. Efforts to combat global warming, preserve the Ozone layer that is essential to life on Earth, and other activities to sustain the biodiversity of plant and animal life on our planet all depend on applications satellites. Responding to major disasters routinely involves analysis of satellite imagery and mobile satellite communications. This then is a comprehensive reference work about the practical use of satellites to serve humanity and make our lives better.

The multibillion-dollar (US) world of commercial satellite applications and services continues to expand each year. This means that the technology is becoming more sophisticated and reliable and the practical applications ever broader. Commercial satellite applications and satellite technology are both becoming more sophisticated and efficient. This is particularly true in terms of finding more and more applications in different fields and in the expanded use of automation and the application of expert systems and artificial intelligence to allow more autonomous operation of satellites in outer space. The size of satellite applications markets are now measured in the hundreds of billions of dollars (US). Virtually every country and territory in the world relies on applications satellites for multiple space-based services. The diversity of the submarkets within the field we have defined as "satellite applications" continues to expand and becomes more complex. Indeed, in view of this growing dependency on space applications and the expanding number of satellites in near-, middle-, and geostationary Earth orbits, the international community, through the United Nations and other fora, is working to ensure the long-term sustainability of activities in outer space.

Today the field of "satellite applications" includes at least: (1) satellite communications, (2) satellite broadcasting, (3) satellite navigation and positioning, (4) geostationary satellite meteorology, (5) remote sensing and Earth observation, and (6) space-based information systems. And this is just the beginning. The above-cited satellite applications activities generate other major space-related activities and industries which are themselves of significant size. For instance, the field of commercial satellite applications creates a substantial part of a multibillion-dollar

(US) launch vehicle industry around the world. It also creates yet another multibillion-dollar market for earth station antennas, very small aperture antennas, micro-terminals, and hand-held satellite transceivers. Finally there are also important ancillary markets that also feed off of commercial "satellite applications." The supporting industries include:

1. *Space-related insurance and risk management industries* (such coverage requires expenditures equivalent to 10–20% of the value of the satellite and its launcher).
2. *Engineering, design, reliability testing, and regulatory support activities*. Key technical support is required to design new systems and carry out research related to new space and ground systems. (These engineering companies and research organizations prepare detailed technical specifications for satellite systems and work with specialized law firms to prepare requests for regulatory approvals and frequency assignments and allocations at the national, regional, and even global levels.)
3. *Financial institutions, investment banks, and underwriting corporations*. These institutions help to raise the billions of dollars in capital needed to build and launch the satellites and deploy hundreds of millions of earth station antennas, receive only terminals and two-way satellite telephones around the globe.
4. *Marketing and sales organizations*. These companies help with the sales of satellite applications services around the world to literally billions of people who depend on applications satellites for severe weather warnings, for radio and television services, for Internet connection, for navigational and routing information, and for vital information for farming, fishing, mining, or urban planning.

Today the overall field of "satellite applications" thus represents not only the primary sectors that build or operate commercial space satellite systems and launch them but a huge supporting work force as well. These include hundreds of service, engineering, manufacturing, specialized banking, and insurance companies. These supporting service industries represent an important set of commercial enterprises representing billions of dollars (US) in revenues (see Fig. 1.1).

Without all of these diverse satellite applications providing meteorological and weather information, communications, broadcasting, navigational, remote sensing, and supporting space-based information services, the world we live in would be greatly different. Without these systems, for instance, many more lives would be lost to hurricanes, tornadoes, tsunamis, earthquakes, volcanoes, and other violent acts of nature. Without these systems there would be less effective global communications systems. Satellite telecommunications systems support telephone, Internet, and other IP-based information services across virtually all of the world's 200 plus countries and territories. These satellite systems also provide communications to ships and offshore platforms and buoys in the seas as well as to the Polar region and to aircraft in the skies. Applications satellites are an important part of the world's search and rescue (SAR) infrastructure for downed pilots, stranded passengers and crews of shipwrecked vessels, or people lost in wilderness areas or subject to natural disasters. Other space navigational systems and geodetic satellite systems provide key real-time information to all parts of the world whether on land, the seas, or in the air, including tracking of goods in our global transportation network. These space systems, with increasing

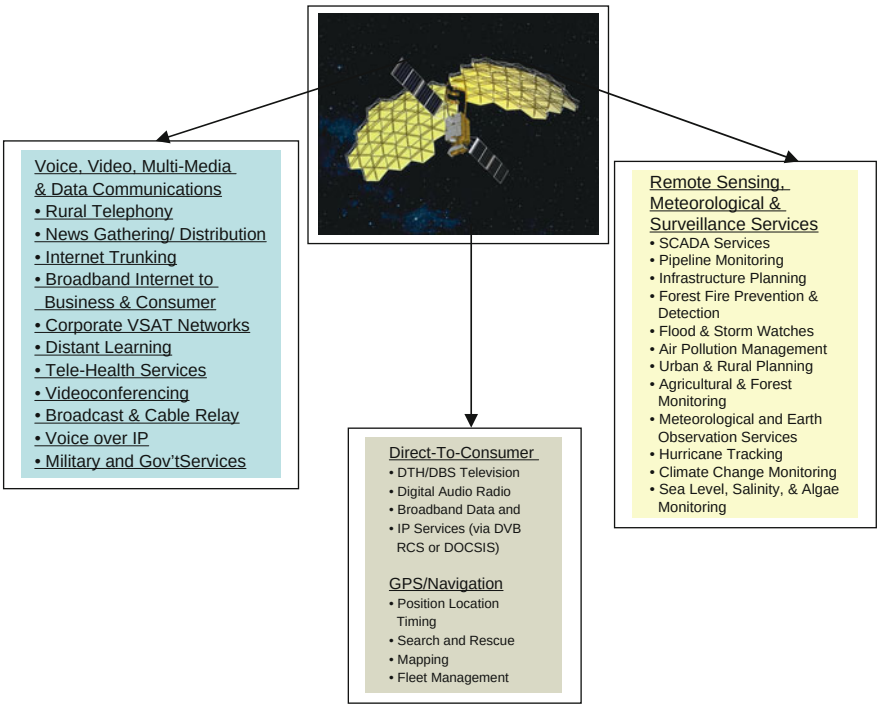


Fig. 1.1 The wide range of satellite applications services provided from space today (Graphic courtesy of the author)

accuracy, can tell us where people or vehicles or buildings or a myriad of things are located for a wide range of applications. Without broadcasting satellites there would be limited television, radio, and other broadcasting services around the world.

Over a billion people receive television, radio, or communications live via satellite to their homes and offices. Satellite services have become so pervasive that they have almost disappeared from the public consciousness as something unique and special. The use of outer space has become almost commonplace in a span of a half century. Much like electric motors or batteries, the vast and extensive use of satellites in our everyday lives has thus often become "invisible." These key machines in the skies help us to predict the weather, receive an entertainment broadcast, connect to the Internet across the globe, know where we are in our cars and how to get to our destinations, protect us from fires, or help us to have access to a wide range of resources from apples to zirconium.

The growth of the satellite applications market will continue to be rapid and diverse – and for many years to come. The first graph below shows the evolutionary nature of the markets from a decade ago. Even then the satellite communications industry in terms of satellite and earth station manufacturing, launch services plus communications services when combined represented total annual revenues of nearly $90 billion. The global positioning system (GPS) to support satellite

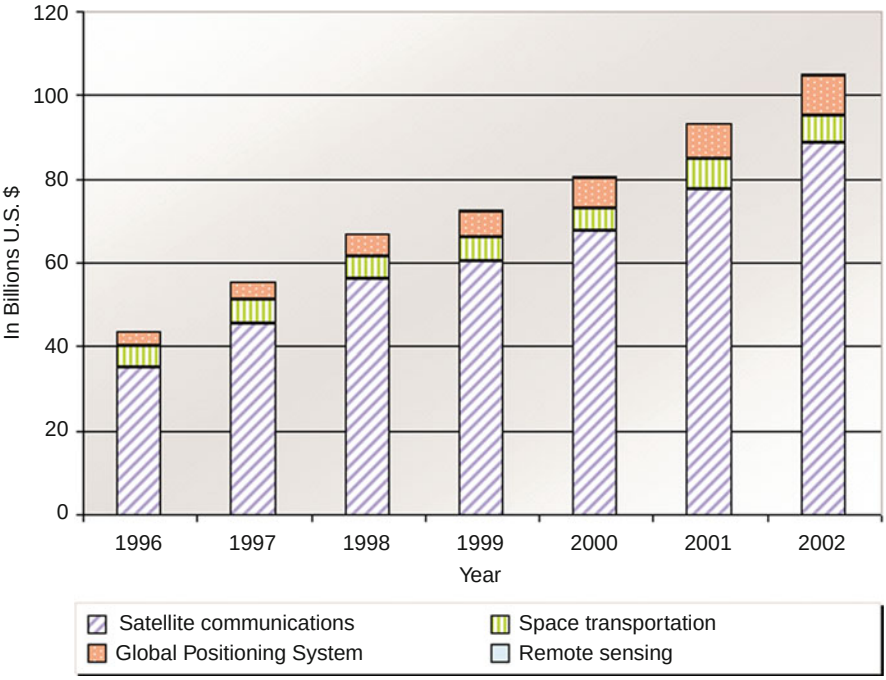


Fig. 1.2 Early stages of growth of the commercial satellite applications market. (U.S. Department of Commerce, *Trends in Space Commerce, Office of Space Commercialization*. (U.S. Department of Commerce, Washington, D.C, 2002))

navigation applications was beginning to emerge as a key market when GPS receivers, and services related to GPS were combined. Commercial space transportation to support these industries was also a significant market, but space navigation and remote sensing was quite small (Fig. 1.2).

The combined revenues for 2009 – just for global commercial satellite applications *services* – totaled around $105 billion (see Fig. 1.3). If one then adds in expenditures for satellite launches, the manufacturing of satellites, manufacturing of earth stations and various types of user terminal equipment, technical consulting support, licensing fees, and insurance and risk management, the total revenues associated with the commercial applications satellite industry – for satellite service revenues plus all other costs and expenditures – the net industry annual revenues climb to over $200 billion per annum.

If one were then to add in the additional costs associated with governmental and defense-related communications systems and the cost of governmentally operated geosynchronous meteorological satellite services, then the annual financial turnover for the satellite industry would exceed $250 billion or over a quarter trillion dollars. It is thus safe to say that overall the combined field of "commercial satellite

REVENUES FOR IN APPLICATIONS SATELLITE SERVICES (2004-2009)[1]
(Note: Figures Do Not Include Revenues for Satellite Manufacturing, Earth Station Manufacturing, Launch Services, Engineering, Licensing, Risk Management, etc.)
(In billions of dollars (U.S.) per year)

Type of Satellite Service	2004	2005	2006	2007	2008	2009
Total Direct to Consumer	$35.8 B	$41.3 B	$48.9 B	$57.9 B	$68.1 B	$75.3 B
DBS Televsion	$35.8 B	$40.2 B	$46.9 B	$55.4 B	$64.9 B	$71.8 B
DBS Radio	$0.3 B	$0.8 B	$1.6 B	$2.1 B	$2.5 B	$2.5 B
Direct Broadband Internet	$0.2 B	$0.3 B	$0.3 B	$0.4 B	$0.8 B	$1.0 B
Fixed Satellite Services	$8.9 B	$9.3 B	$10.7 B	$12.2 B	$13.0 B	$14.4 B
Transponder Lease	$7.0 B	$7.3 B	$8.5 B	$9.5 B	$10.2 B	$11.0 B
Managed Network Services	$1.9 B	$2.0 B	$2.2 B	$2.6 B	$2.8 B	$3.4 B
Mobile Services	$1.8 B	$1.7 B	$2.0 B	$2.1 B	$2.2 B	$2.2 B
Space Navigation Services	$5.8 B	$ 6.8 B	8.0 B	$9.4 B	$10.8 B	12.0 B
Remote Sensing	$ 0.4 B	$ 0.6 B	$0.4 B	$0.4 B	$0.8 B	$1.0 B
TOTAL	$52.7 B	$69.7 B	$70.0 B	$82.0 B	$94.8 B	$105.0 B

[1]Satellite Industry Association "Executive Summary, 2009 State of the Satellite Industry Report ", Prepared by the Futron Corporation, 2010 Washington, D. C.
www.sia.org/news_events/pressrelease/2010StateofSatelliteIndustryReport2010(Final).pdf

Fig. 1.3 Annual revenues for applications satellite services

applications" represents quite a large global industry. Further this is an industry that has shown consistent growth for several decades and has continued to grow even in times of global recession.

The Evolution of Commercial Satellite Applications

With the launch of the Sputnik satellite in October 1957 people began to think of outer space not as something in science fiction novels, but as a real and increasingly important activity. Everett Edward Hale as early as 1867, when he wrote *The Brick Moon*, speculated on the use of artificial satellites for communications, navigation, and remote sensing. But as of the late 1950s, scientists and engineers began to conceive of practical ways to utilize artificial satellites for needed services. The first application was satellite telecommunications. In the late 1950s and early 1960s, international communications capacity for overseas links was both very limited in scope and the per-minute rate of a telephone call was quite high. (Submarine cables for voice communications could only connect 36–72 voice circuits at a time and could not handle even

low-quality black and white television transmissions on a live basis. One might have to pay $20–50 a minute for an overseas telephone connection.)

In the various sections that follow, the history of satellite applications is provided in detail, but the following provides a brief overview. The first practical application of satellites was in the field of telecommunications. A series of experimental satellites were launched in the early 1960s to test the feasibility of communications satellites for commercial purposes. These satellites, known variously as SCORE, Courier 1B, Echo, Telstar, Relay, and Syncom, proved vital to the design of the operational systems that were to follow. These early experimental satellites helped space system designers to discard the idea of using passive satellites for telecommunications. Echo was a metallic coated balloon launched for meteorological experiments, but bounced electronic signals off its reflective surface without amplification. These experiments confirmed that this type of "passive satellite" represented much too low of a capacity for commercial needs. These experiments and many others – particularly the Syncom satellites – showed that deploying satellites into geosynchronous orbit and providing telecommunications services from orbiting spacecraft was technically feasible (Pelton 1974).

This special Geo orbit (sometimes call the Clarke orbit in honor of Sir Arthur Clarke who first suggested this orbit for communications satellites) allowed virtually complete global coverage with only three satellites and eliminated the need for Earth Station antennas to track rapidly across the sky. This is because the Geo orbit allows the satellite to "seem to hover constantly" above the same location over the equator. (Note: A more formal definition of the geostationary orbit is an Earth orbit having zero inclination and zero eccentricity, whose orbital period is equal to the Earth's sidereal period. The altitude of this unique circular orbit is very close to 35,786 km.)

These various experiments led to the deployment of operational telecommunications in 1965. The three initial satellite telecommunications systems, all launched in 1965, were the Intelsat system that deployed the "Early Bird," the low-orbit Initial Defense Communications Satellite System deployed by the US defense department, and the Molniya satellite system for the USSR. There were three Molniya satellites deployed into highly elliptical orbits that were suited to northern latitude coverage over Russia and to the satellite countries known as Soviet Socialist Republics.

The rest is history. The Intelsat satellite system grew into a truly global network that ultimately served nearly 200 countries and territories around the world. A number of national satellite systems were launched to meet domestic communications needs (particularly to meet television and radio broadcasting needs and service to rural and remote areas). In time regional satellite systems evolved and yet other systems were deployed to meet maritime, aeronautical, and land mobile communications. Military-, security-, and defense-related satellite systems were also launched to meet the specialized needs of military agencies. Today there are many hundreds of communications satellites in orbit and well over a thousand have been launched since the late 1950s. Some of these telecommunication satellites are in geosynchronous orbit, others are in medium earth orbit, and yet others are deployed as constellations in low earth orbit. Some of these are multipurpose and support various types of telecommunications services for telephone, radio, and

television broadcasting or distribution or data networking and Internet-related services. Other satellites are designed and optimized for mobile communications for land, sea, or aircraft communications.

Close on the heels of the telecommunications satellites came other types of applications satellites. Military reconnaissance satellite systems were a very high early priority for both the Soviet Union and the USA in the cold war. Fully half of the first 20 Soviet Cosmos series space launches were for military Zenith imaging systems, and the US Corona satellite system was developed in secret starting in 1959. The Corona program was started under the camouflage of public statements that these satellites were scientific payloads. The Corona program was not publicly acknowledged for many years and not until well after being out of service. These "spy" satellites set the stage for remote sensing and weather satellites.

First came the weather or meteorological satellites, which were initially developed in order to provide weather and cloud cover information for the military imaging systems. The US President's Science Advisory Committee reported in 1958 that "The satellite that will turn its attention downward holds great promise for meteorology and the eventual improvement of weather forecasting." But the potential benefit of weather satellites was evident and the first civil weather satellite launched was the TIROS (Television and Infra-red Observation Satellite), which started as a defense department program and was transferred to the new NASA in April of 1959. Its first images in 1960 provided a synoptic view of weather patterns, sea ice, and other features that were immediately analyzed on the ground to great effect, and were the first in an unbroken series of weather satellites that operate to this day. TIROS was in a low Earth (435 miles or approximately 700 km) orbit, but the potential for a permanent geostationary orbital view was clear. The first geo weather satellite was the US GEOS-1, launched on October 16, 1975, demonstrating the benefit of the geostationary orbit for weather observation. Over the past 30 years, additional weather satellites have been launched by Europe, Japan, India, Russia, and China. These now provide a constant global view of our world and have revolutionized our understanding of global weather patterns and our ability to accurately forecast the weather.

This was followed by remote satellite sensing systems and specialized Earth observation satellites, with the launch of the Earth Resources Technology Satellite in July 1972 (later renamed Landsat 1). The Landsat series led the way in the development of dedicated Earth resources satellites that were specifically designed for a wide range of applications such as agriculture, forestry, and water resources. These systems have continued to develop and have improved their capabilities, with spatial resolution improving from 80 m with Landsat 1 to under 0.5 m with the current ultrahigh resolution systems.

The most recent class of satellite applications to evolve are those associated with satellite navigation, also referred to as precision timing and navigation systems (PNT) or Global Navigation Satellite Systems (GNSS). These systems were first devised to assist with military- and defense-related purposes such as targeting and mapping, such as the early US TRANSIT and Soviet Tsikada Doppler-based systems, which were first fielded in 1959 (Transit) and 1974 (Tsikada). Today, however, there are a wide range of commercial uses for space navigation satellite systems and these satellites actually represent a multibillion-dollar industry

worldwide. This market has grown rapidly and continues to develop new uses. Next-generation system development started with the US Navstar GPS system, first developed in 1973 and fully deployed in 1994. Europe, Russia, India, Japan, and China are all developing and launching their own advanced systems, and the future of this class of satellites is bright.

Later in this handbook the more detailed histories of these various types of applications satellites are provided for those that would like to know how these various types of satellite systems evolved over time.

More than a thousand applications satellites are now in low, medium, or geo-synchronous orbits and these are being used to provide one or more types of commercial satellite services. Indeed there have been a number of instances where a satellite built for one type of application such as telecommunications had another "package" added to the satellite to provide meteorological imaging such as was the case with an Indian "Insat" satellite. Sometimes an operational satellite will have an experimental package attached to test out a new technology. One example of this was the Orion international communications satellite that had an experimental inter-satellite link (ISL) package added to it for performance testing. Today most satellites are designed for a particular application because the frequency bands (or radio frequency spectrum) allocated for space applications through the International Telecommunication Union are (ITU) are typically different for different types of applications.

Later in this handbook specifics of these allocations are provided. Nevertheless satellites can have a primary payload for one application and then have one or more secondary payload(s) for other applications or to experiment with a new technology or new frequency band. An example of this is the US NOAA polar orbiting POES satellites which also carry the Cospas_SARSAT search and rescue and ARGOS telemetry systems. For 50 years now there have been applications satellites in the Earth orbit. Thousands of these various types of satellites have been launched and some of them have come back from their orbit and burned up in the atmosphere or crashed backed into Earth. Others have been pushed out into space above a geosynchronous orbit where they will stay for millions of years. There are, however, many thousands of defunct and derelict satellites or parts of satellites or launchers still in orbit known as orbital debris. Currently on the order of 20,000 bits of "orbital debris," about the size of a tennis ball or larger, are known to be in orbit and millions of microscopic elements are present – especially in low earth and polar orbits. By December of 2010, 4,765 launches and 251 on-orbit breakups had led to an orbital debris tracking condition where at least 16,200 objects had been entered into the US Space Surveillance Network (SSN) catalog.

The problem of orbital debris crashing into a satellite, space station, or other active space object is thus a growing concern. The crash between a defunct Russian Kosmos satellite and an active Iridium communications satellite in 2009 has served to highlight these concerns. The United Nations Committee on the Peaceful Uses of Outer Space (COPUOS) that has been addressing this issue for some time has now agreed as of June 2007 to voluntary procedures to reduce the threat of orbital debris creation in future years.[1]

What Does the Term "Satellite Applications" Mean and Why Consider It in a Unified Way?

Why a Handbook of Satellite Applications? One can find handbooks and reference sources in many areas that include the various "fields" of satellite applications. There are reference handbooks on satellite telecommunications, satellite broadcasting, satellite-based remote sensing, or Earth observation. There are also some reference materials on satellite-based meteorology and space navigation. The key to "satellite applications" is to recognize that while these space-based services all tend to have a different range of users – and require different specialists to use the information, the underlying technology with regard to designing, manufacturing, launching, insuring, financing, and getting regulatory approval for "applications satellites" are in many ways quite similar. As noted above an applications satellite designed for one type of service or application can also have a secondary or even tertiary package (i.e., payload) to perform operations in an entirely different field. The platform used for telecommunications, broadcasting, remote sensing, Earth observation, meteorological purposes, or space navigation start out to be remarkably similar in their design, manufacture, testing, launch requirements, and, in many cases, even their deployment.

Common Elements of Applications Satellites

An applications satellite's mission is defined by its payload, which carries out its specialized function, but the platform on which the payload resides is quite often similar in terms of structure, power systems, tracking, telemetry, command and monitoring systems, stabilization, positioning, pointing and orientation systems, thermal control systems, and so on. Many manufacturers of applications satellites have evolved toward the design of various size platforms that meet various customer needs. At the smaller end of the spectrum the Surrey Space Center's microsatellite "bus" or platform has been used for communications, IT related, remote sensing, and other scientific purposes. For progressive larger satellites with more ambitious objectives to support larger payloads, commercial aerospace manufacturers have progressive larger platforms that support larger and more sophisticated missions.

Over the last 40 years there have been more telecommunications and broadcasting satellites designed, manufactured, and launched than other types of applications satellites. These "communications satellites" have been deployed for commercial, governmental, and military purposes. Just because of their sheer volume, the platforms developed for communications satellites have generally tended to characterize the range of platforms available for other purposes in terms of size, structural integrity, maneuverability, lifetime, power systems, and pointing accuracy. At the very beginning a new platform was designed for each satellite. This custom design process was often driven by the fact that satellites were becoming more capable in size and performance. This constant upgrading of performance and the need for larger satellite antennas required greater pointing accuracy.

In another context, the satellites were also being designed for longer life. Several prime characteristics defined the design of these increasingly complicated and larger platforms. These characteristics were: (1) increased capability for the payload; (2) prime or peak power requirements over the satellites' lifetime; (3) the pointing accuracy and orientation requirements for the satellite platform and the size and shape of the antennas or functional elements that must be supported by this platform; and (4) the lifetime desired for the satellites operation and the need to remove satellites from their operational orbit at end of life – factors that required more fuel and larger fuel tanks. These four characteristics were the main drivers that led to a wide range of platform designs. Over time the manufacturers realized, just like manufacturers of automobiles, that one did not have to design a new "bus" or "platform" every time there was a new order for a communications or other type of applications satellite. Thus manufacturers began to standardize classes of platforms from nano- or microsatellites up through giant 10,000–12,000-kg platforms that are built for the largest type of direct broadcast or mobile satellites that can be launched by currently available launch systems such as the Atlas 5 or the Ariane 5.

Although the platforms might be quite different in volume and mass, varying from about 200–12,000 kg, they usually contain many of the same features. These are: batteries (for emergencies and when the satellite might be in eclipse); solar cell arrays (as a prime source of electrical power); a strong but lightweight structure to hold the satellite and its components together; a thermal control system to keep the components and its payload from becoming too cold or too hot; an electrical system for controlling components; a tracking, telemetry, command, and monitoring system so that people on the ground can actively know how the satellite platform and its payload is operating and send commands to maintain effective operations; a source of fuel and a thruster maneuvering system to aid in keeping a proper orbit; and a finely tuned pointing and orientation system, particularly to help position and point the satellite antennas or onboard sensing system for best performance. Finally there is a star, sun, and/or Earth sensing system to allow people on the ground to know exactly how the satellite is pointing on an X, Y, and Z axis or there is something like an RF alignment system to allow very precise pointing.

This platform or bus will also contain a payload (or in some cases multiple payloads) to perform a particular function such as communications, broadcasting, meteorological sensing, Earth sensing, Earth observation, space navigation, or perhaps a scientific experimental mission. Regardless of the payload and its mission, these elements will largely be common to the "bus" that delivers payload to where it needs to go and to support the payload's operation 24 h a day, 7 days a week, until the mission is complete. When the mission is complete the payload is then employed to help with the final disposition of the satellite, such as bringing the satellite from low earth orbit back into the atmosphere where it will burn up or crash harmlessly into an ocean. If the satellite should happen to be in geosynchronous orbit (i.e., a distance that is equivalent to one tenth of the way to the Moon), then the usual maneuver to remove this type of spacecraft from orbit is to raise it above Geo orbit where it will stay for many thousands of years.

Although there are many common elements to an applications satellite "bus" there still remain quite a large diversity of design elements that are described in the later chapters of this handbook. Some satellites need to be very accurately oriented to perform their mission and others much less so. Thus they range from very simple and low-pointing orientation systems to much more sophisticated ones. The simplest orienting system that is still in use is a gravity gradient system. With this type of platform design there are long booms that can be deployed to extend out into space away from the satellite. Once the booms are extended perhaps 5 m or more away from the satellite in different directions, the pull of gravity from the Earth can more or less orient the satellite toward the Earth. Other designs include satellites that spin around at speeds like 50–60 rpm, while their payloads inside spin in the opposite direction to achieve constant pointing toward the Earth with a stable orientation. These "spinners" were quite common in the early days of satellite communications. Today the most common "bus" is a three-axis-oriented platform that has one to three momentum or inertial wheels inside of the core of the satellite that spin at very high speeds, such as 4,000–5,000 rpm; this serves much like a spinning top to provide very accurate pointing orientation toward the Earth or wherever the platform needs to be oriented.

Just as there are options with the pointing and orienting system there are also options with regard to the thermal control system. Different types of reflective surfaces can be used to control solar heating. There are devices called heat pipes that can transfer heat from the interior of the satellite to the exterior. Despite the diversity of design options, most commercial manufacturers of satellites have a series of four or so basic platforms from which to build desired application satellites again, just as automobile manufactures have four or so chasses from which they build new automobiles. These various elements of the satellite platforms are discussed in great detail in the special section devoted to this subject.

In general, however, smaller and lower cost satellites will have shorter lifetimes with smaller capacities if they are communications satellites. If they are sensing satellites they will have lower resolution or lesser sensing capabilities. In short, smaller applications satellites will have lesser capabilities than the larger spacecraft. Further, there are often economies of scale that are achieved in the design of larger and longer-lived satellites and they also tend to be most cost efficient to launch on a "per kilogram to orbit" basis.

Organization and Effective Utilization of the Satellite Applications Handbook

This handbook is organized to be useful to a wide range of potential users from design engineers, faculty members and teachers, to reference librarians and students. It is organized into major sections. These sections are each self-contained and provide the history, demand for the service, and technology – past, present, and future. Thus there are sections on: (a) *Space Telecommunications* (this section includes the main categories of Fixed Satellite Services, Mobile Satellite Services,

Broadcast Satellite Services, plus store and forward data services, etc.); (b) *Satellite Precision Timing and Navigation* (this type of satellite application has now become the second largest commercial satellite service in terms of market size); (c) *Space Remote Sensing* (this section not only covers remote sensing and Earth Observation, but it also addresses the *Global Information System* (*GIS*) and related software); and (d) *Space Systems for Geosynchronous Meteorology*. This is the remaining key civilian practical use of outer space. It is different in that the provision of this service is largely by governmental agencies rather than commercial companies. The practical value of this service is more difficult to quantify but each and every year this type of satellite applications serves to save many, many lives and greatly minimizes property damage.

Conclusion

This handbook thus addresses the above-described four areas of commercial satellite applications and seeks to do so in considerable depth. It does not, however, specifically address classified military and defense-related satellite applications.

What is presented is specific and detailed information about all forms of telecommunications satellites, remote sensing and Earth Observations, satellite navigation, and satellite meteorological satellites. Coverage is provided for the commercial use by defense organizations of applications satellites in a so-called dual use mode. This term applies to civilian or commercial satellites that are also used to meet certain largely "non-tactical" military applications. In this regard it is important to note that military usage of satellites is in many ways quite parallel to civilian space applications, and often presages the development of commercial systems. This is to indicate, for instance, that the basic engineering and design characteristics of telecommunications and remote sensing satellites are often quite similar, although military systems may add special features such as radiation hardening, antijamming capabilities, and encryption capabilities.

The aforementioned four satellite applications are today the prime commercial and practical civilian uses of outer space. To be comprehensive the handbook also presents current and detailed information regarding global launching capability around the world and also addresses the design, manufacture, test, and deployment of the application satellite spacecraft platforms or "buses" that are launched to support these various types of commercial satellite services. As noted above, the platforms for these various applications satellites are quite similar even though the payloads may be quite different.

Finally the last part of the handbook addresses the key economic, regulatory, social business, and trade issues that are associated with applications satellites. Again, although the payloads that are contained on various types of applications satellites are different, the economics, trade, and regulatory aspects in these various commercial systems are similar. This is to say that applications satellites of various types need to use radio frequencies (RF) to send information to and from earth and that accordingly there is a need for RF allocations by the International Telecommunication Union for these various operations. There are

national processes for the approval of specific frequency assignments to particular spacecraft in a way to prevent undue frequency interference. There are also a host of technical, economic, regulatory, trade, and even social and religious issues that arise from the use of applications satellites since by their very nature these satellites are international in their operation. Thus commercial applications satellites and their operation, for instance, come under some degree of regulatory control by the World Trade Organization (WTO) with regard to how these services are distributed or sold and the nature of international and national regulation and control. Orbital debris is an increasing threat to the safe operation of applications satellites. This final section thus addresses all of these types of issues and especially regulatory, trade, business, and social issues.

There are several other elements of the reference handbook that should be particularly noted in terms of convenience of use. First of all, the various parts of this handbook are divided into major sections and chapters with highly descriptive titles. Secondly, each chapter contains a list of key words so that if one is interested in "orbital debris," "photo voltaic solar cells," "lithium batteries," "frequency allocations," "precision timing," or the "United Nations Committee on the Peaceful Uses of Outer Space (COPUOS)" these terms should be easily identified. The organization of the handbook is typically structured to put information about any one subject in a concentrated location. This means that power systems are discussed together rather than in four different sections for each type of applications satellite. Finally the appendices are a key source of information about actual applications satellite systems and other types of technical information.

Notes

1. United Nations Committee on the Peaceful uses of Outer Space Voluntary Guidelines on Orbital Debris, http://www.orbitaldebris.jsc.nasa.gov/mitigate/mitigation.html

References

J.N. Pelton, *Global Satellite Communications Policy: Intelsat, Politics and Functionalism* (Lomond Systems, Mt. Airy, MD, 1974), pp. 47–50

Satellite Communications Overview

2

Joseph N. Pelton

Contents

Introduction 19
Overview of Commercial Satellite Services..... 20
Notes 25
References..... 25

Keywords

Broadcast satellite service (BSS) • Fixed satellite service (FSS) • Geosynchronous earth orbit (GEO) • International Telecommunication Union (ITU) • Intersatellite link (ISL) • Low earth orbit (LEO) • Mobile satellite service (MSS) • Satellite constellations • Store and forward satellite service

Introduction

The serious consideration of the provision of satellite communications from space dates from 1945 when the first technical descriptions were written with regard to launching a spacecraft into geosynchronous orbit and the design of space stations as extraterrestrial radio relays was specifically outlined. In the historical section that follows, however, it becomes clear that the idea or concept had been around many years, indeed centuries before. The 1945 article, however, described the possible delivery of telecommunications services from space and presented detailed calculations as to how this might efficiently be done from a special orbit known as the geosynchronous (or sometimes the geostationary) orbit (Clarke 1945). The use of

J.N. Pelton
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: joepelton@verizon.net

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_2, © Springer Science+Business Media New York 2013

radio waves for long-distance communications up until the 1960s was limited to microwave relay between towers or the use of shortwave or high-frequency (HF) transmissions that were, in effect, bounced off of the ionosphere. This latter technique was quite limited in transmission throughput and unreliable because the ionosphere was subject to distortions largely due to solar radiation and the so-called solar wind and solar storms. Launch technology that could place satellites in orbit came into being in the late 1950s.

This new launch capability that came to the USSR and the USA in the late 1950s and the expanded capability to launch a satellite into geosynchronous orbit that came in the early 1960s allowed the rapid evolution of satellite communications technology. Within a decade a wide variety of telecommunications services from satellites in different types of orbits became possible. The International Telecommunication Union (ITU), the specialized agency of the United Nations that oversees the use of radio frequencies (RF) for practical and scientific purposes assumed responsibility for satellite radio frequencies. This began with a globally attended Extraordinary Administrative Radio Conference (EARC) in 1959. The ITU thus provided for the first time a formal process by which radio frequency (RF) spectrum could be allocated to support satellite communication (Pelton 1974).

Over time, the ITU defined a number of satellite communications services that might be offered via different satellites in different types of orbits. The ITU international processes also defined a system and a process whereby there could be technical coordination of such satellites to limit interference between and among them. The number and type of satellite communications services have grown and expanded over the years as is discussed in the following sections (ITU 2008).

There are today many types of technical designs for satellite communications and these technologies are optimized to support a variety of services around the world. A wide range of commercial satellites now operate at the national, regional, and global level. These satellite systems support various types of data, telephone, television, radio, and various networking services around the world.

Overview of Commercial Satellite Services

The services defined by the International Telecommunication Union (ITU) include the following.

Fixed Satellite Services (*FSS*): FSS spacecraft support telecommunications services between antennas that are at fixed points. These fixed antennas can be used for reception only, for two-way communication (like a cable in the sky), or for communications within a network that can start with only a few nodes or can grow to a very large network indeed with thousands of interconnected nodes. The first of the commercial communications satellite services were these FSS systems. The Intelsat FSS system was the first to begin to provide international commercial telecommunications services in 1965. Also deployed in 1965 was the FSS system called Molniya, which provided telecommunications services for the Soviet Union,

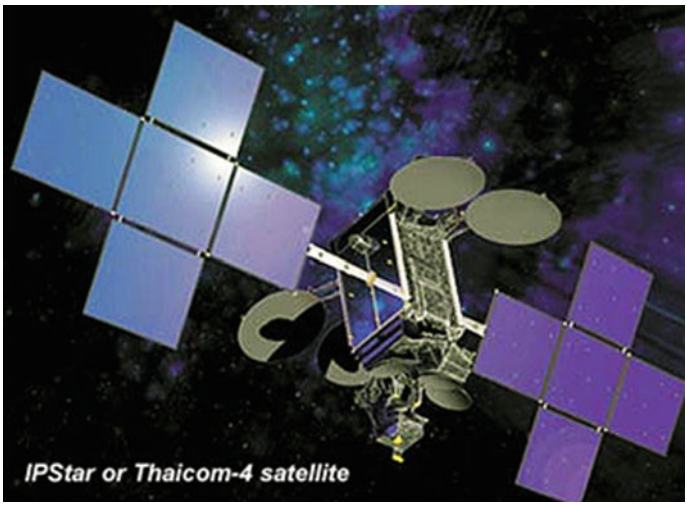


Fig. 2.1 The IP Star Satellite also known as Thaicom-4 (Graphic courtesy of IPStar)

other Soviet Socialist Republics, and Cuba. An Initial Defense Satellite Communications Satellite system was also deployed to provide FSS services to support US defense-related telecommunications services. These initial FSS systems have now multiplied to support satellite telecommunications for over 200 countries and territories around the world (Pelton 2006, p. 30). Figure 2.1 shows a current generation broadband FSS satellite, the IP Star that operates in the Asian region of the world.

In addition to over 200 commercial communications satellites that supply fixed satellite services, there are now scores of military communications satellites that provide fixed satellite services in support of defense-related missions. Although the largest fleet of defense-related communications satellites are owned and operated by the US military, there are a number of strategic communications satellite systems owned by over a dozen countries around the world. In addition commercial satellite systems leased capacity to military systems for so-called dual-use purposes to supplement the capabilities of defense satellite communications systems. Figure 2.2 shows the WGSS military satellite designed to provide communications services.

Broadcast Satellite Services (*BSS*): BSS satellites use very high powered beams to deliver radio or television services directly to end users. In order for this service (also known informally as the direct broadcast satellite (DBS) service) to be economical and efficient, the receiver terminals must be small in size, low in cost, and easy to install and operate. Different RF bands are used for radio or direct audio broadcast services (DABS) in contrast to direct broadcast satellite (DBS) television services. The broadcast satellite service began later than the initial FSS offerings, but this industry has grown rapidly and is now by far the largest revenue generator in the satellite world by a wide margin (Pelton 2006, p. 31).

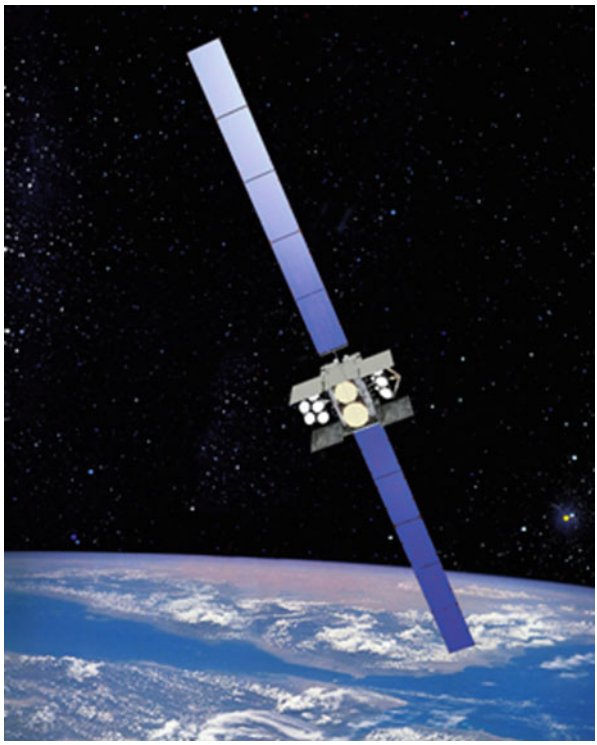


Fig. 2.2 The wideband global Satcom satellite (Graphic courtesy of the U.S. Military)

The Nimiq BSS satellite (Fig. 2.3), operated by the Canadian Telesat organization, provides direct broadcast services to Canada and the USA.

Mobile Satellite Services (*MSS*): The MSS services provide telecommunications to end-user antennas that move rather than remain stationary. The MSS services today include telecommunications connectivity for maritime, aeronautical, and land-based users. The first MSS satellites were designed for maritime service. Next these satellites were used to support both maritime and aeronautical services. The last type of mobile communications satellites that has evolved are those designed for land-based mobile services. This is the most demanding of the MSS services technically, but in terms of market, this is also the most demanding. A variety of different designs in different orbits have evolved with the initial land mobile systems known as Iridium, Globalstar, and ICO experiencing severe financial and market difficulties with their initial service offerings. These organizations have been reorganized and the Iridium and Globalstar systems are deploying their second-generation systems while ICO is developing a new mobile satellite service for the US market (Fig. 2.4).

Today there are a variety of MSS satellites deployed in a variety of different orbits. Some of the latest systems are those designed to work in conjunction with

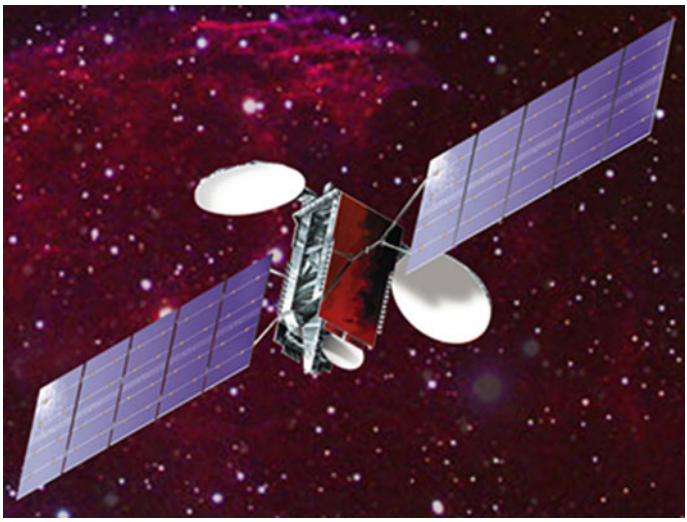


Fig. 2.3 A Canadian Nimiq direct broadcast satellite (Graphic courtesy of Telesat)



Fig. 2.4 A constellation of 66 iridium satellites provides global mobile services (Graphic courtesy of Iridium)

terrestrial cellular telephone services within urban areas. These hybrid systems that integrate mobile communication satellites with terrestrial cellular systems are called MSS with "Ancillary Terrestrial Component (ATC)" in the USA. The equivalent service is called MSS with Complementary Ground Component (CGC) in Europe. These hybrid mobile systems combine urban terrestrial cellular systems in a seamless manner to allow very high powered MSS satellites to cover the rest of a country or region. Unlike the initial constellations like Iridium

and Globalstar, these new systems with a terrestrial component are deployed in Geosynchronous orbit and are targeted to service to a single country like the USA or a single region like Europe (Pelton 2006, p. 31).

These satellite services, FSS, BSS, and MSS, are the so-called big three of the commercial satellite services and represent a very significant part of the total worldwide market for the satellite industry. Nevertheless there are other types of telecommunication satellite systems that can be, and indeed are, deployed. One additional system is the so-called store and forward type data relay satellite that can support messaging services to remote areas. The more satellites deployed in low earth orbit to support in this type of system, the more rapidly a message can be relayed from one part of the world to another. If there are enough satellites of this type, like in the Orbcomm system, you can have almost instant messaging. In some cases the receiver can be configured to not only receive short messages but also to receive space navigation signals to support vehicular or ship navigation. One can also design a transceiver to send short data messages as well as to receive them, as has been done with several store and forward satellite systems.

Some commercial satellite systems employ what are called cross-links (CLs) or intersatellite links (ISLs), or in ITU parlance Intersatellite Service (ISS) in order to operate. These can be used in low earth orbit or medium earth orbit to interconnect satellite constellations. ISLs were a part of the design of the low earth orbit Iridium satellite network and they have been used in some military communications satellites. ISLs could also be used to interconnect geosynchronous satellites on an interregional basis in order to avoid double hops when communicating halfway around the world. Today, satellite service connections providing global linkages are more likely to combine with a fiber-optic submarine cable to achieve rapid connectivity since intersatellite links for regional interconnectivity remain relatively rare. The more common use of ISLs is to interconnect satellites within a large-scale low earth orbit constellation where the satellites are typically hundreds of kilometers apart from one another rather than many thousands of kilometers apart such as the case when geosynchronous satellites are serving different regions of the world (Pelton 2006, p. 31).

Finally there are satellites for military or defense-related communications. These satellites for military purposes are allocated different frequency spectrum than commercial satellites. These satellites resemble commercial satellites in many of their technical features, but they often have special features. Special capabilities can include radiation hardening, additional redundancy, and special encryption capabilities. Military communications are not operated on a commercial basis for the most part. There is a special chapter in this handbook that does describe such commercial defense-related communications satellite systems such as X-TAR as well as the "dual use" of commercial satellites for civilian requirements as well as defense-related applications.[1]

Other applications satellites are designed for different purposes other than telecommunications. Yet these too must be able to relay information to various users on the ground. Thus there are many types of satellites systems, other than commercial satellite systems, which are designed to support scientific communications, exact timing, remote sensing, earth observation, search and rescue, geodetic

measurements, or various types of environmental services such as to monitor tsunamis, volcanoes, earthquakes, etc. These types of satellites are not addressed in this part of the Handbook, but are covered in later sections.

Conclusion

The chapters that follow in this section seek to provide a comprehensive and interdisciplinary overview of satellite communications services and applications, markets, economics, technology, operations and continuity of service, regulation, and future trends. Specific information on commercial satellite systems is provided in the appendices at the end of the handbook. ► Overview of the Spacecraft Bus address the common technical elements found in essentially all applications satellites. Thus ► Overview of the Spacecraft Bus addresses spacecraft power systems; thermal balancing and heat dissipation systems; orientation, pointing, and positioning systems; structural design elements; diagnostic systems; tracking, telemetry, and command systems; manufacturing and integration; and quality and reliability testing processes. ► Major Launch Systems Available Globally also address how the various applications satellites are launched by different rocket systems from various launch sites around the world.

Notes

1. J. Oslund, Dual use challenge and response: military and commercial uses of space communications, in *Communications Satellites: Global Change Agents*, ed. by J.N. Pelton, R.J. Oslund, P. Marshall (Lawrence Erlbaum, Mahwah, 2004), pp. 175–194. Also see Satellite security and performance in an era of dual use. Int. J. Space Commun., www.spacejournal.ohio.edu/Issue6/pdf/pelton.pdf

References

A.C. Clarke, Extraterrestrial radio relays. Wireless World (Oct 1945)

ITU, Radio regulations as published in 2008 (ITU, 2008), www.itu.int/publ/R-REG-RR/en

J.N. Pelton, *Intelsat: Global Communications Satellite Policy* (Lomond Systems, Mt. Airy, 1974), pp. 40–47

J.N. Pelton, *The Basics of Satellite Telecommunications International* (Engineering Consortium, Chicago, 2006), p. 30

History of Satellite Communications

3

Joseph N. Pelton

Contents

Introduction 28
Early History of Satellite Communications 29
The Modern History of Satellite Communications 30
Separate Systems for Maritime and Mobile Satellite Services 42
Evolution of Regional and Domestic Satellite Systems 46
Satellite Systems to Support Defense- and Military-Related Services 51
Direct Broadcast Satellite Systems 55
Satellite Radio Broadcasting 58
Economic and Political Evolution of Global Telecommunications 60
Cross-References 65
Notes 66
References 66

Abstract

The history of satellite communications is a rich one that began centuries ago with the efforts to interpret the meaning of the "wandering planets" among the stars and to understand the structure of the cosmos. Early scientists such as Sir Isaac Newton and writers of speculative fiction both contemplated the idea that humans might one day launch artificial satellites into orbit for practical purposes. This chapter provides a brief overview of that rich international history up through the early days of global satellite operations. This history continues to provide a narrative concerning the different types of satellites that have evolved to offer various kinds of services and the development of competitive satellite networks that are at the core of the communications satellite industry today. A brief history of military satellite systems and the

J.N. Pelton
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: joepelton@verizon.net

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_35, © Springer Science+Business Media New York 2013

"dual use" of commercial satellite systems to support defense-related communications needs is also addressed.

Keywords

Arabsat • CEPT (Council European Post and Telecommunications) • Clarke, Arthur C. • Cold War • Communications satellite corporation (Comsat) • Courier satellite • Department of defense (DOD) • Divestiture of AT&T • Domestic satellite leases • Early bird satellite • Edward Everett • Eutelsat • FR-3 satellite • Galileo • Global information grid (GIG) • Global maritime distress safety system (GMDSS) • Hale, Edward Everett • Initial defense satellite communications systems (IDSCS) • Inmarsat • Intelsat • International maritime organization (IMO) • International telecommunication union (ITU) • Kefauver, Estes • Kennedy, John Fitzgerald • Kerr, Robert S. • MARECS satellite • Marisat • Ministry of post and telecommunications • Molniya • Moon landing • Newton, Isaac • Pickering, William • Relay satellite • Robert Goddard, Greene • SCORE satellite • Shockley, William • Submarine cable systems • Syncom satellite • Telstar • Tsoilkovsky, Konstantin • TV-Sat • United Nations (UN) • van Allen, James • von Braun, Werner • Wells, H.G. • Wideband global satellite (WGS)

Introduction

The idea of satellite communications is a powerful one that has spawned a billion dollar industry on which the people of Planet Earth now depend every day. The idea that humans could actually launch a satellite into Earth orbit, however, was dependent on certain key knowledge about the Solar System that was lacking for many millennia. The concept of an artificial satellite revolving around our home planet was first and foremost dependent on the understanding that Earth itself is a planetary body that revolves around the Sun and that Earth and Moon are subject to universal laws of gravity. It further requires the understanding that the Moon, as a satellite, revolves around the Earth.

In short, before the orbital mechanics of the Solar System were understood and the concept of gravity clearly comprehended, the idea that one might launch an artificial moon or satellite into Earth orbit made no sense. But once one did grasp the basic physics involved, the idea that an artificial satellite could serve as a very high "artificial relay tower" for communications was a quite logical concept to follow. Clearly an artificial satellite circling the Earth might indeed be designed to receive radio waves or some form of signal transmitted up from the Earth out to space and return them to a desired distant location. How then did this historical thought process occur and who were the key players? This chapter not only outlines the history of satellite communications, but also indicates how the current structure of today's complex satellite markets is now evolving.

Early History of Satellite Communications

The starting point in this thought process began with the understanding that the Earth revolves around the Sun and that the Moon revolves around the Earth. This correct conception was confirmed by Galileo Galilei (1565–1642) in the sixteenth and seventeenth century when he was able to look through a telescope to observe Jupiter and note that four satellites were revolving around Jupiter. The limited magnification of his telescope prevented him observing that there were in fact many more artificial moons revolving around this giant planet and that there were other moons circling other planets. The discovery of the moons circling Jupiter provided sufficient physical data to draw a reasonable conclusion about the basic physics of the Solar System's orbital mechanics. Galileo's discoveries aided the thought process to posit that the Earth was also one of the "wanderers" or "planets" that revolved around the Sun. It also helped to confirm that the Earth had its own orbiting satellite, which we call the Moon.

Actually it was Galileo who first coined the term "satellite" that we use today. He applied to these distant moons the Latin word *satelles*. Galileo thought this word might appropriately be used to describe the "moons" of Jupiter. The Latin word was at the time used to describe an attendant or servant who was bound to obey the commands of his master. To Galileo the distant moons flying around Jupiter were bound to obey the commands of this mighty distant planet. Today we indeed have a large number of application satellites which do the bidding of their human designers. Many applications and scientific satellites now launched into Earth orbit carry out communications, navigation, remote sensing, or meteorology as well as various types of scientific discoveries. Galileo, however, did not understand at that time the concept of gravity and thus did not understand what force was used to command the moons of Jupiter to circle in their orbits, nor why the Moon should circle Earth. Indeed, because Galileo's observations ran counter to the dicta of the Catholic Church it was quite a while until the workings of the Solar System became widely comprehended and understood in a correct scientific sense (Pelton and Madry 2010).

The next key historical step essential to the understanding of how an artificial satellite might be launched into Earth orbit and then provide services to people back on the ground came with the seventeenth-century discovery of gravity. Isaac Newton's discovery had many implications that impacted everything from astrophysics to zoology. He figured out how the mechanics of gravity worked within the Solar System. He did this through his own observations as well as by studying the writings of Galileo and Copernicus. His writings described how a very powerful cannon might shoot an object with enough velocity in order to allow the "launched object" to travel greater and greater distances. He then concluded that if the object could be shot with sufficient velocity it would overcome the pull of Earth's gravity and would attain orbital speed and thus start circling the Earth. There is even a wonderful illustration from his writings in *Philosophiæ Naturalis Principia Mathematica* that show how this might be accomplished (Pelton 1981).

It is interesting that the next step in the thought process that led to the actual launch of applications satellites came not from the annals of science but from the

imaginative literature of the nineteenth century. Writers such as Achille Eyraud (*Voyage to Venus*, 1863), Jules Venus (*From Earth to Moon*, 1865), Edward Everett Hale (*The Brick Moon*, 1869), and H.G. Wells (*The First Men in the Moon*, 1897) inspired popular and scientific thought about the possibility of space travel and the construction of rocket ships that could launch people and things into orbit or even beyond. It is the writings of Edward Everett Hale that today seems to be the most remarkable in its anticipation of today's application satellites. His book in 1869 anticipated the ability to launch a satellite into so-called polar orbit. In his book he described how an "artificial moon" could be deployed as a practical device for communications, Earth observation, or navigation.

The Modern History of Satellite Communications

By the twentieth century technology was evolving very rapidly. Konstantin Tsoilkovsky (1857–1935), in Russia at the very outset of the new century, gave careful and deliberate thought to the design of rockets that could carry people into outer space. Robert Goddard (1882–1945) began experiments in the USA to build viable rocket launchers only to be laughed at in a New York Times editorial as the "Moon Man." Goddard persevered and in 1926 proved that viable liquid-fueled launchers were indeed possible. During World War II, the German government assembled a team of scientists to develop rockets as weapons systems based, in part, on Goddard's earlier work. These led to the development in Germany of buzz-bombs, V-1 and then the V-2 rockets with ever-increasing range and accuracy. After the war a part of the German rocket team was brought to the USA to work on this technology and the other part went to the Soviet Union to develop rocket systems there. From these two efforts came the launcher systems that became so prominent a feature of the so-called Cold War.

In 1945, a young man named Arthur C. Clarke wrote an article that brought into clear focus exactly what a communications satellite system might do, how it might be launched, and even presented in detail the reasons why such a space-based communications network should be place into geosynchronous orbit. Arthur C. Clarke, who spent World War II in the British Radar Establishment, first developed his ideas and sent a detailed letter to colleagues in June of 1945. Then in October 1945, he published his ideas and calculations in the journal *Wireless World*. At the time this landmark article did not attract a great deal of attention. It was not the cover story of that edition and he only received only a modest 15 pounds sterling compensation for his efforts. His 1945 article at the time was thus a largely unheralded event, even though his brief eight-page article contained the basic concepts on which a multibillion industry would be born and global television news reporting "live via satellite" would become commonplace only a few decades later (Clarke 1945).

Arthur C. Clarke, who died at age 90 in 2008, explained to colleagues before he died that he did not seek to patent the idea of a geosynchronous communications satellite. This was simply because he anticipated that the space stations he wrote about would be realized many years into the future. He believed that his "space

stations" would require a crew to replace the radio tubes that would frequently burn out. In short, Sir Arthur Clarke, who was knighted by the British Government for his many predictions and farsighted writings, did not anticipate the transistor and the integrated circuit. These devices were to make possible not only the reliable solid state technology that would enable reliable satellite technology but also would facilitate the development of high-speed electronic computers that could calculate the celestial mechanics associated with their accurate deployment into space.

In fact the invention of the transistor came only a few years later at Bell Labs, in December 1947.[1] This fundamental breakthrough by William Bradford Shockley, John Bardeen, and Walter Houses Brattain led to many innovations that ranged from the transistor radio to the modern electronic computer. The "transistor" and the integrated circuitry that followed have transformed the world in almost every conceivable way over the past half century, from the World Wide Web to the cell phone. Certainly the transistor transformed the concept of a communications satellite and the practical utilization of outer space from a far off dream to only a difficult technical challenge.

On October 7, 1957, the Space Age began with the launch by the Soviet Union of the world's first artificial satellite, Sputnik 1. In light of the "Cold War" that then existed between the USA and the Union of Soviet Socialist Republics (USSR) this launch, even though carried out in the context of the International Geophysical Year (IGY) for global scientific research, was broadly interpreted in a political context. Thus there was an immediate perception that the USA was subject to a so-called missile gap. This led to immediate efforts by the USA to launch and orbit a satellite of its own. Another almost immediate response to the launch of Sputnik was for the US Congress to pass a new law in 1958 to create a new space organization known as the National Aeronautical and Space Administration (NASA).

In the months that followed, the Soviet Union (i.e., the shortened name of the USSR) continued to launch increasingly sophisticated and larger satellites while the USA experienced a series of embarrassing launch failures. On February 1, 1958, however, the USA did manage to launch the Explorer 1 satellite into orbit. This satellite and the launch team, headed by Dr. William Pickering of the Jet Propulsion Lab, Dr. James Van Allen of the University of Iowa, and Werner Von Braun of NASA, confirmed the existence of the powerful belts of radiation that surround the Earth. The second Soviet satellite, Sputnik 2, had also sensed the presence of orbital radiation.

From the period from 1957 to the early 1960s a number of satellites were launched by the Soviet Union and the USA – the only two countries with orbital launch capability at that time. The Soviet Union also demonstrated an early capability to launch heavier satellites, and to orbit animals and then even people into orbit. On April 12, 1961, Vostok 1 was launched with Yuri Gagarin aboard to become the first person in space. The USA, with lesser launch capability, initially focused on miniaturization so that it could launch more capable satellites with a smaller mass. The US presidential election of 1960 hinged in part on the issue of the "missile gap" and President John F. Kennedy focused one of his first major speeches to Congress on the issue of outer space.

Kennedy's speech to a Joint Session of Congress on May 25, 1961, is most memorable for his challenge to the USA to send people to the Moon and

successfully return them by the end of the 1960s decade. This speech, known formally as the "Special Message to Congress on Urgent National Needs," was the one which launched the NASA Moon mission known as Project Apollo. In that same speech Kennedy also called for other space achievements. He called for funding for the Rover nuclear launch system and the rapid development of satellite communications technology and systems. He urged Congressional funding of $50 million (equivalent to perhaps $500 million in 2010) for "accelerating the use of space satellites for world-wide communications" (Special Message to the Congress on Urgent National Needs 1961). Clearly the Moon mission was what dominated the press coverage the next day, but President Kennedy also put great personal stress on the future potential of a global communications satellites network. In September 1961, some 4 months later, President Kennedy went to the General Assembly of the United Nations and called for the establishment of a single global satellite system that would: "... benefit all countries, promote world peace, and allow non-discriminating access for countries of the world."[2] As a result of US urgings, the United Nations adopted resolution 1721 that included Section P, which stated "communications by means of satellite should be available to the millions of the world as soon as possible on a global and nondiscriminatory basis" (United Nations General Assembly Resolution of Satellite Communications 1961).

The ongoing political processes led to the creation of the Communications Satellite Corporation (Comsat) in 1962 when the US Congress enacted the Communications Satellite Act of 1962. This led to the subsequent signing in Washington, DC, in August 1964 of the Initial International Agreement to create the International Telecommunications Satellite Consortium known as Intelsat (and its companion Operating Agreement). The creation of Intelsat was largely spearheaded by US initiatives and especially through representatives of the US State Department and of Comsat. This new Intelsat entity, which was initially organized as an international consortium, started with mainly Western countries as members (USA, Australia, Canada, Japan, and most of the Western European nations) and grew to include well over 100 member countries around the world (Pelton and Alper 1986). Seven years later after the 1964 launch of Intelsat, the Soviet Union, in response to its growing international membership, launched another entity known as the Intersputnik International Organization of Space Communications (or simply Intersputnik) with a membership of eight socialist countries, namely, the Soviet Union plus Bulgaria, Cuba, Czechoslovakia, East Germany, Hungary, Mongolia, Poland, and Romania.[3]

While the national and global political processes were moving along, the related satellite technology was developing at an even swifter pace.

In December 1958, the US Signal Corps launched what might be characterized as the world's first "broadcasting satellite." This satellite, known as SCORE, simply repeated a brief message from President Eisenhower: "Peace on Earth, Goodwill Toward Men." It was launched just before Christmas on December 18, 1958, and its batteries were exhausted just before the end of the year. On August 12, 1960, the Echo I, a giant aluminized balloon was launched to carry out meteorological experiments, but AT&T Bell Labs experimenters also tested the idea that such



Fig. 3.1 Courier 1B satellite – world's first active repeater satellite (Photo courtesy of the US army signal corps)

a satellite could serve as a passive reflector of radio signals as way to relay signals back to Earth, somewhat like bouncing shortwave radio transmissions off the ionsphere. These experiments were in a way successful by demonstrating that the signal throughput for a "passive communications satellite" would be too modest to serve as a commercially viable communications service.

It was not until October 1960 that the first active communications satellite, Courier 1B, was launched (see Fig. 3.1). This experimental spacecraft only supported the transmission of 16 teletype channels. Yet this satellite actively demonstrated that the relay of a signal to a satellite and then its retransmission of teletype messages back to Earth could be technically achieved. Its active transponders were powered by solar cells. From this landmark demonstration, quite rapid progress toward more and more capable communications satellites continued apace. Although today's space systems, a half century later, literally possess a billion times more capacity, the basic technical concept is in many ways the same.

By 1962, there was a surge in the technical sophistication of the design of active communications satellites. On July 10, 1962, the Telstar satellite, as designed by Dr. John R. Pierce and his team at Bell Labs, was launched into low Earth orbit (Fig. 3.2).

For the first time in human history the Telstar satellite demonstrated how a live and real-time television signal could be relayed across an ocean. This was followed by the launch of the Relay satellite on December 14, 1962. This satellite, as built by RCA in accord with NASA Goddard Spaceflight Center design, was similar to the Telstar and conducted similar transmission experiments. This NASA design specified an

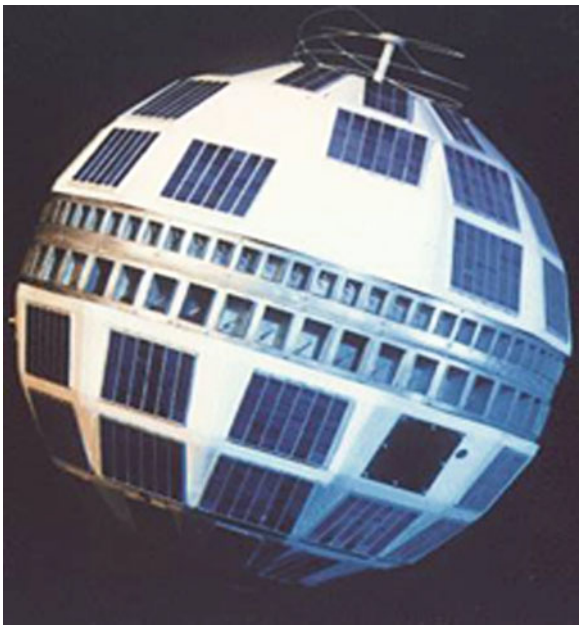


Fig. 3.2 The AT&T designed Telstar satellite that first transmitted live television (Photo courtesy of Bell Labs)

augmented power system that provided this satellite with a longer in-orbit life. Thus Relay 1 remained in service through 1965. This satellite, in addition to conducting television transmission tests, was also designed to measure the impact of the Van Allen Belt radiation on the satellite communications subsystem (Fig. 3.3) (U.S. Congressional Hearings 1962).

The technical feasibility of satellite communications to support teletype, voice, and even television had been demonstrated by the end of 1962. The remaining key technical question was whether a communications satellite could be successfully launched into geosynchronous orbit (sometimes call the Clarke orbit in honor of Arthur C. Clarke) and operated reliably from this great distance – almost a tenth of the way to the Moon.

This question was answered in 1963 when the Hughes Aircraft Company designed and built the so-called Syncom satellites (for geoSYNchronous COMmunications satellites). The three satellites of this design were launched by NASA on Delta launch vehicles. The first launch was a failure, but the second, Syncom 2, was successfully launched on December 14, 1963, exactly 1 year after the launch of Relay. The Syncom 2 and subsequent Syncom 3, as engineered by Dr. Harold Rosen and his team at Hughes, demonstrated that reliable communications to geosynchronous orbit with a return link to Earth was indeed technically and operationally viable. For the 1964 Olympics in Japan, television signals were transmitted from Japan to the USA via Syncom 3 and the signal was transmitted from the USA to Europe via the Relay 1 satellite. The idea of global television relay of major sporting and world events "live via satellite" across the oceans thus date back to the early 1960s.

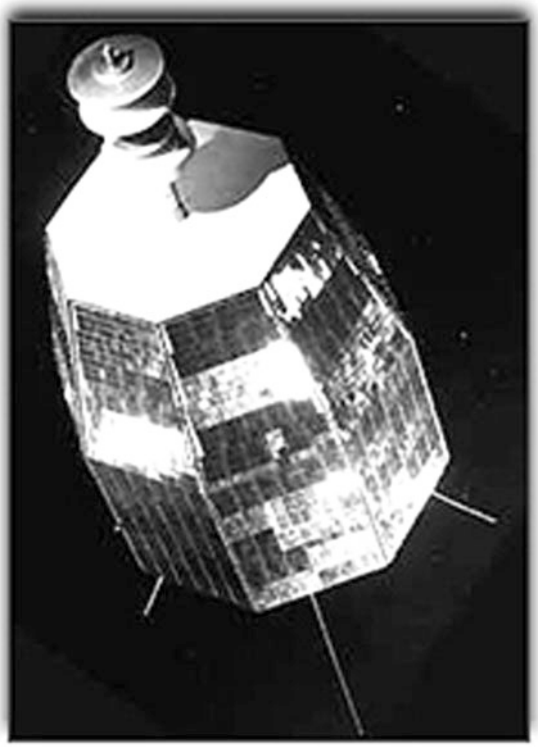


Fig. 3.3 The Relay 1 satellite designed by NASA and built by RCA (Photo courtesy of NASA)

Although satellite telecommunications technology was moving swiftly ahead, the political and economic processes to establish a mechanism to provide satellite services to the world were subject to a number of key challenges (Fig. 3.4). The enactment of the Communications Satellite Act of 1962 in the USA constituted a protracted and very difficult political process. This conflict arose because the telecommunications industries wanted satellite communications services to be completely commercialized and Sen. Robert S. Kerr of Oklahoma, who headed the powerful public works committee, led the fight in this direction. Sen. Estes Kefauver, who had been the Democratic candidate for Vice President on the ticket with Adlai Stevenson, argued that public expenditures through NASA had brought this new technology to the level of industrial feasibility and he led the fight for a public agency for satellite communications. President Kennedy, on the other hand, was eager for a bold new space initiative and badly wanted to put the USA in a leadership role with regard to establishing global satellite communications. In addition, he needed to heed the advice of the powerful Senator Kefauver and his colleagues. In short, he wanted a compromise solution. The threatened filibuster in the Senate required skillful action. He relied on John A. Johnson, then general counsel at NASA, to draft a compromise bill that created the Communications Satellite Corporation (COMSAT) as a private corporation, but with half of the

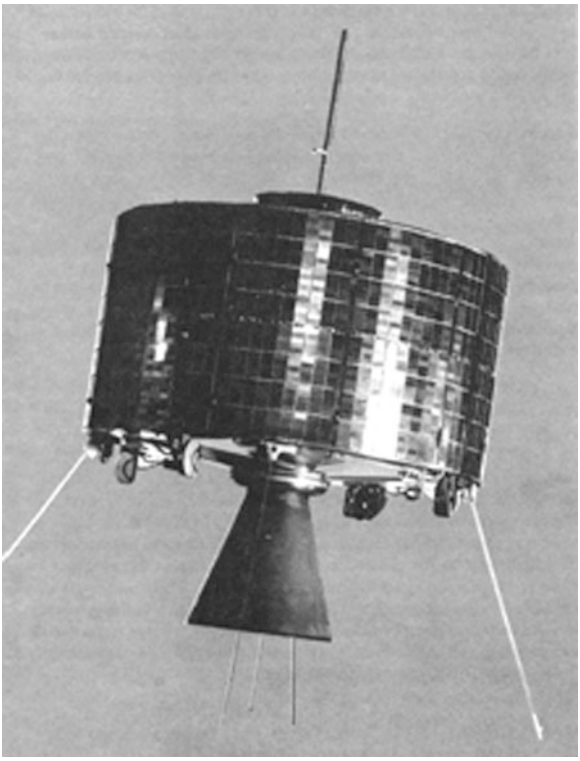


Fig. 3.4 Syncom experimental satellite that first demonstrated feasibility of operation from a geosynchronous orbit (Photo courtesy of NASA)

shares going to major telecommunications companies such as AT&T, IT&T, RCA, Western Union, and Western Union International and with the other half of the shares to be sold to the public on the New York Stock Exchange. Under this compromise bill, COMSAT was subject to instruction by the US Government on matters of national policy by the State Department, the Federal Communications Commission (FCC), and the White House Office of Telecommunications Policy (OTP). In addition technical advice was also to be provided by NASA. This compromise bill managed to pass and break the deadlock between the Kefauver and Kerr factions and an ongoing filibuster avoided.

The next challenge was the international negotiations to create a framework for international satellite communications services. The original thought within the US State Department was that COMSAT would undertake to establish a series of bilateral agreements with countries that wished to establish satellite links. When the US delegations arrived in Europe to discuss international arrangements for satellite communications they were confronted with a unified European position via the Committee on European Post and Telecommunications (CEPT). These European telecommunications officials insisted that a new international agency would need to be formed for this purpose.

Two years of tough international negotiations ensued. The final outcome was the signing of the Interim Intelsat Agreements, as described earlier. There were two

agreements. One document was an Intergovernmental Agreement signed by nation states on behalf of their governments. The other document was called the Special Agreement and this was signed by "participating" telecommunications organizations as variously constituted within the countries that signed the Intergovernmental Agreement. The purpose of having this second agreement was to allow private or semi-private companies such as Telespazio of Italy, KDD of Japan, the Overseas Telecommunications Corporation (Australia), the Canadian Overseas Telecommunications Corporation (COTC), or Comsat of the USA to participate directly in the organization as a partial owner as well as governmental agencies such as post and telecommunications agencies.

These documents were deemed to be "interim" in nature because European countries maintained that the USA possessed an unfair advantage due to their technical lead in launch vehicle technology. They successfully maintained that after 5 years of experience, a new set of permanent arrangements should be negotiated to reflect newly gained operational capabilities and strengthened new space technologies that were evolving around the world. These countries (especially European nations) believed, and correctly so, that under the permanent arrangements there would be the opportunity for a more thoroughgoing internationalization of the Intelsat management. In particular, COMSAT was designated as the Manager of the Intelsat system in the Interim Intelsat Arrangements largely due to US official insistence. The supporters of the "interim arrangements" believed that after experience had been gained, the "U.S. dominant technical and operational role" could and would decrease as space capabilities spread around the world.

The two Interim Intelsat Agreements were signed by 15 countries in Washington, DC, on August 20, 1964. Some countries had the ability for these signatures to take immediate effect and others had to obtain ratification by their national legislatures. In the months and years that followed more and more countries joined this initial satellite communications consortium. An official report on experience gained was completed in 1969 and this led to 2 years of negotiations that concluded with the so-called Final Agreements in 1971. In was not until 1973 that enough signatures were gained for these new agreements to enter into force. By this time membership had swelled to well over 80 countries.

The Communications Satellite Corporation, COMSAT, in its role as Manager for the Intelsat Consortium, sought to bring the Intelsat system into operation as soon as possible once it was established in 1962. As a result of the successful deployment of the Syncom 2 and 3 satellites into geosynchronous orbit, COMSAT signed a contract with the Hughes Aircraft Company to build a somewhat larger version of Syncom. This satellite with a larger bank of solar cells, a "squinted beam" antenna that provided increased pointing ability back toward the Earth – and thus higher gain – was the result. This satellite once deployed, was able to provide the equivalent of 240 voice circuits (or complete two-way voice channels) or alternatively one low-quality black-and-white television channel.

This satellite that was officially known as the Intelsat I (F-1) was actually more popularly known in the world press as "Early Bird" (the "F" stood for flight model and indicated it was successfully launched into orbit). This satellite, which was

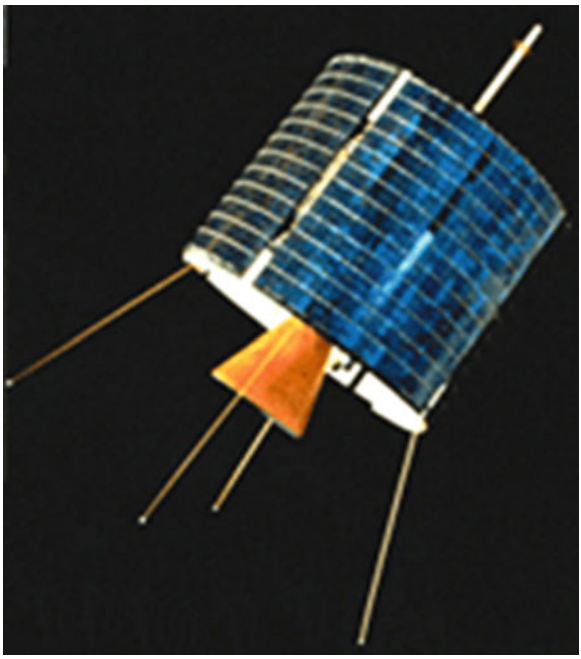


Fig. 3.5 The early bird satellite, world's first commercial communications satellite (Graphic courtesy of the Comsat Legacy Project)

launched in April 1965 just 8 months after the formation of the Intelsat Consortium, surprised the world by achieving practical commercial satellite communications in a remarkably short period of time.

Exciting satellite video experiments were conducted. For example, Dr. Michael Debakey conducted open heart surgery in Switzerland and the procedures were watched live via satellite by heart surgeons in Houston, Texas, who were able to ask questions in real time. Coverage of the LeMans auto race in France were beamed to the USA and Heads of State were able to exchange greetings (Fig. 3.5).

Early Bird, when it was launched in 1965, was in many ways an experimental satellite. But the Intelsat II series was able to provide multi-destination service and video, audio and data service to ships at sea in support of the US Gemini space program. Next came the Intelsat III series with more than five times the capacity of the early Intelsat satellites. Each of these satellites, with much higher gain antennas, could provide 1,200 two-way telephone circuits plus two color television channels.

This Intelsat III series was the first to complete a fully global network. It was in June 1969 that a network of these satellites were deployed and fully configured so that they could send voice and television channels not only across the Atlantic and the Pacific Oceans, but even across the Indian Ocean. It was this global Intelsat III network that allowed a worldwide audience of over 500 million people to see the Moon Landing of Apollo 11 on the Lunar surface and the first space walk (Fig. 3.6).

In the following years, from 1969 to the early 1970s, Permanent Management Agreements were negotiated for the newly named International

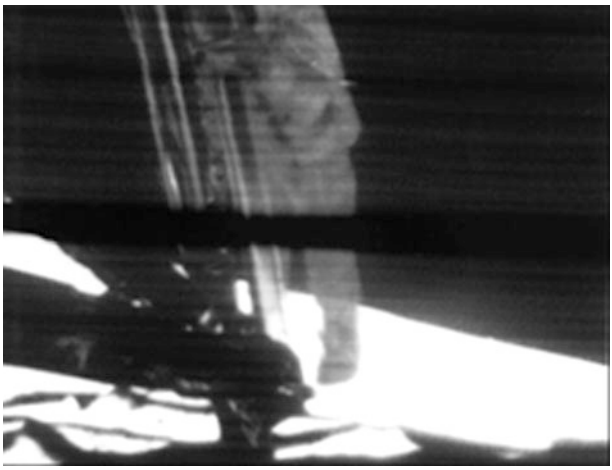


Fig. 3.6 Low-resolution TV picture of Neil Armstrong's first step on the Moon (Image courtesy of NASA)

Telecommunications Satellite Organization (Intelsat). During the period an international management was established that assumed these responsibilities from the USA-based COMSAT over a transitional period. The communications satellites increased in size, power, lifetime, and performance and migrated from analog to digital communication services. The size of Intelsat Earth Stations decreased in size and very small aperture terminals (VSATs) supporting what was called the Intelsat Business Service became commonplace. Intelsat provided not only tens of thousands of international voice circuits, data networks, and eventually hundreds of television channels, but it also increasingly leased spare capacity to support domestic voice, data, and television distribution in scores of countries around the world. In the years that followed the Intelsat system grew in satellite capability and performance, especially as the multiplexing systems migrated from analog to digital.

The number of members and participating countries and territories also expanded and international and domestic traffic surged, despite competition from the greatly expanded channel capacity of fiber-optic submarine cables that were laid across the oceans.

This significant expansion of the satellite communications business and its perception as a viable and attractive business led to efforts to restructure the global system within which these services were provided. There was an increased move, particularly within the countries of the OECD (Organization Economic and Cooperative Development), toward competitive telecommunications service. In the early 1980s, during the Reagan Administration in the USA there were several competitive filings, particularly with systems known as Orion and Panamsat, that proposed that they be authorized to compete directly with Intelsat. It took a number of years for this whole issue to be resolved in terms of the restructuring of Intelsat to become a commercial entity and for ground rules to be agreed as to how competitive systems

might be authorized by governments and allowed to operate within their borders. This "macro change" in the structure of telecommunications toward "liberalization" and "competition" led to the conversion of Intelsat and other publicly structured "public monopolies" to become competitive private telecommunications industries by the later part of the 1980s and the early 1990s and for the new commercial satellite systems to be licensed and to deploy their competitive networks.

These efforts to create competitive systems at the international level were, to a certain extent, stimulated by efforts to create at the regional level communications systems, such as Eutelsat for the European region, and Arabsat, for the areas of the Middle East and Northern Africa. Further the decision to create a separate international organization for maritime and aeronautical satellite communications, called INMARSAT, also served to create momentum toward creating satellite systems separate from Intelsat. This thought process argued that systems designed and optimized for specific markets could be better optimized than a single system configured to meet all possible requirements. This thought process, namely, of more competition for telecommunications services began to arise in the 1980s. The argument arose among economists that instead of just having national monopoly communications networks, competition would help improve services and reduce consumer costs. Up until the 1980s, in most countries telecommunications organizations were regulated by so-called rate base oversight. This meant in practical terms that the more they invested in new "allowable communications infrastructure" the more "return" they could realize. In the 1980s, many countries switched over to the idea of competitive telecommunication systems. The thought was that this competitive system might be more responsive and cost less than simply having a monopoly provider.

These various national and regional decisions of the 1980s led to new levels of competition for telecommunications services. A 1983 "judgment" for Federal Judge Harold Greene that settled a suit against AT&T undertaken by the US Justice Department led to the breakup of the AT&T monopoly in the USA as of January 1984. The development of competitive systems in Europe, Japan, and elsewhere followed in the next few years. This context of moving monopoly telecommunications systems to competitive networks clearly set the context for the authorization of competitive international satellite systems.

This process strongly contributed to the ultimate decision among the Intelsat Assembly of Parties to transform Intelsat from a public international organization with national governments acting as members and investors to entirely new arrangements. After the restructure of Intelsat, it became a privately held corporation as of July 18, 2001. Henceforth Intelsat became just another corporation offering satellite telecommunications services around the world. This led to the "privatized" Intelsat spinning off part of its assets to a new European-based company known as "New Skies." The same sentiments and logic ended with both Inmarsat and Eutelsat also being "privatized" so that these organizations were entirely owned by private equities and no longer owned by national governments. This meant they were no longer international organizations operating under international treaty arrangements, but simply commercial competitors operating in a commercial marketplace along with other competitors with no special rights and privileges.

In the case of both Intelsat and Inmarsat small international organizations were set up to address special concerns about the "public good" and "public services" these organizations as public international organizations had previously performed. These involved services such as the right of access to international communications via satellite for public safety, for other special public needs and also to assist developing countries to achieve equitable access to telecommunication satellite services. In the case of Intelsat, a small part of the former INTELSAT Organization was not privatized on July 18, 2001. This modestly sized residual group remained an international organization, under the acronym ITSO (standing for "International Telecommunications Satellite Organization"). The role of this organization, with 150 members from around the world, and which had previously owned Intelsat when it was "spun off," was officially defined to be as follows:

- Act as the supervisory authority of the new Intelsat Ltd.
- Ensure the performance of Core Principles for the provision of international public telecommunications services, with high reliability and quality.
- Promote international public telecommunications services to meet the needs of the information and communication society.[4]

This action was taken to assuage those members of Intelsat that had been reluctant to "privatize" Intelsat. As a practical matter the ITSO has limited ability to affect the commercial policies of Intelsat Ltd.

The same parallel was followed in the case of privatizing Inmarsat that in fact occurred before the Intelsat restructuring. In this case the Inmarsat derivative body became known as the International Mobile Satellite Organization (IMSO). This intergovernmental body was likewise established to ensure that Inmarsat continues to meet its public service obligations, including obligations relating to the Global Maritime Distress Safety System (GMDSS). IMSO is also designated an observer to attend meetings of the UN Specialized Agency, the International Maritime Organization (IMO). In April 1998, the Inmarsat Convention was amended to create this IMSO in its current form when Inmarsat Ltd. was restructured as a privatized organization. In addition to its public maritime safety role the IMSO seeks to guarantee that services are provided by Inmarsat Ltd. free from any discrimination and in a peaceful way to all persons living or working in locations that are inaccessible to conventional, terrestrial means of communication. IMSO also ensures that the principles of fair competition are observed.[5]

Over time this commercialization or privatization process led to a series of mergers and acquisitions. New Skies was purchased by the group known as SES Global, based in Luxembourg, as part of its global network of satellite assets. Perhaps most ironically of all, the privatized Intelsat eventually ended up purchasing the satellite organization known as Panamsat. This company, i.e., Panamsat, had originally been its biggest international competitor and driver of the competitive process that led to Intelsat being restructured as a private competitive satellite provider.

Apart from the move to create a competitive global industrial structure for the provision of worldwide fixed satellite services starting in the late 1980s and 1990s, the overall history of satellite communications was punctuated by several key events. These will be addressed in appropriately titled sections ahead and these

events relate to: (1) the creation of separate satellite systems for maritime and mobile satellite services; (2) the evolution of regional and domestic satellite systems; (3) the development of satellite systems to support infrastructure for defense- and military-related services; and (4) the development and launch of direct broadcast satellite systems, known in the parlance of the International Telecommunication Union (ITU) as the Broadcast Satellite Service (BSS).

Separate Systems for Maritime and Mobile Satellite Services

The success of satellites for international communications, and especially the new ability to provide broadband service across the oceans, quickly led to interest in using satellite technology for maritime communications (Fig. 3.7).

As noted above, the Intelsat II satellite series was sponsored by the National Aeronautics and Space Administration (NASA) essentially to support communications between launch vehicles ascending from Cape Canaveral and to establish links with tracking ships in the Atlantic Ocean in support of the Gemini Manned Space Program. But this was an inefficient system because of the small antenna size of the ship-mounted reflectors. Intelsat satellites, at least in earlier years, were designed to communicate between and among larger-scale fixed location Earth Stations. The desire by the US Navy to communicate more effectively with its globally deployed fleet led to the planning of a dedicated maritime satellite known as Marisat. This satellite as pictured above was manufactured by the Hughes Aircraft Company (now the Boeing Corporation). It was deployed in 1976 on the basis that half of the Marisat system capacity would dedicated to meeting US Navy fleet communications needs and the other half to commercial maritime communications needs. COMSAT General, a subsidiary of Comsat created to enter into other satellite ventures, served as the operator of the system and marketed the additional maritime capacity to other entities desiring maritime services.

The success of fixed satellite services stimulated worldwide interest in "the next step" in terms of maritime satellite services. The European Space Agency had developed and launched some experimental fixed satellites known as the European Communications Satellites (ECS). It followed this program with the European Communications Satellites (ECS) for Maritime Service, known as (MARECS). These satellites were launched and performed a number of successful tests and demonstrations. Within Intelsat there was active discussion as to whether it should expand its services into the maritime mobile communications satellite services area. Under Article XIV of the definitive Intelsat Arrangements Intelsat was granted by its member states the right to enter into what were characterized as "specialized services" that included maritime, aeronautical, or land mobile services. There was, however, a quite important caveat added to this authorization. This required an active determination by the Intelsat Assembly of Parties (the plenipotentiary body of all member states) that the provision of such specialized services would not involve an economic penalty to member states that were not users of these additional services. In the mid to late 1970s, Intelsat was in the



Fig. 3.7 The Marisat satellite, the world's first dedicated maritime communications satellite (Picture courtesy of Comsat Legacy Project)

process of acquiring its fifth generation of satellites known as the Intelsat V. These satellites, with 12,000 voice circuit capacity and two television channels, were procured from the Ford Aerospace Corporation (now Space Systems/Loral) with an initial purchase of six satellites. After considerable discussion and a vote within the Intelsat Board of Governors and finally a favorable decision by the Intelsat Assembly of Parties it was decided to acquire three additional Intelsat V satellites with a maritime communications package aboard. These Intelsat V-MCS satellites were also launched successfully into orbit.

The launch of the Marisat, MARECS, and ISV-MCS capacity into geosynchronous orbit created a great deal of maritime mobile satellite capacity, but the institutional and organization situation was certainly quite unclear. The Intelsat organization had a strong interest in extending its worldwide sway over maritime and possibly aeronautical and other mobile services. However, the institutional situation was complicated by several factors. One key factor was that the Soviet Union was not a member of Intelsat. The USSR was not a major user of international telecommunications services, but it was certainly a key player when it came to maritime communications. Another key factor was that a number of countries tended to see Intelsat, even after the negotiation of the permanent management arrangements and the creation of an internationally staffed Executive Organ headed by a Secretary General (and later a Director General), to be largely controlled and staffed by the USA.

These "complications" led to preliminary discussions held in the UK about the possibility of creating a new international organization to provide maritime communications satellite services. The Safety Committee of the International Maritime Organization (IMO), a United Nations specialized international organization, also endorsed the idea of creating a separate international organization dedicated to maritime satellite communications and safety. This led to a series of three Conferences in 1975 and the formal signing of an international agreement to establish INMARSAT in 1976 that actually went into force in 1979.

In this new organization European, Soviet Union, and other major shipping interests would have a predominant voting share in contrast to Intelsat where the USA had predominant control. In short, the thought was to create a new organization that would be structured around maritime shipping interests, maritime fleets, and maritime safety and not international telecommunications usage as reflected in the Intelsat Organization.

In its structure and its enabling agreement, however, this new international organization closely resembled the Intelsat organization. Like Intelsat, INMARSAT had an Assembly of its members, a Board and specialized advisory committees of its Board. Its membership, however, was focused on the major maritime powers and most notably differed from Intelsat by including the Soviet Union in its membership and ownership. It was also headquartered in London rather than Washington, DC, and thus it was largely seen as a "European" entity rather than an "American" institution. Unlike Intelsat, that had to gradually build up its space infrastructure that took from 1965 to 1969 to establish a global network, Inmarsat "inherited" a global space network that included the Marisat satellites, the MARECS satellites, and the Intelsat V MCS that together covered most of the world's oceans except for a thin strip of the Southern Pacific off the coast of Chile.

Once the Inmarsat Agreement was in place the issue of mobile satellite communications to support aeronautical services began to arise in the 1980s. Inmarsat not only began to plan its own dedicated satellites to support future maritime needs but also began working toward space segment capability that could not only meet maritime needs but also provide communications to aircraft with appropriately designed antennas that could be easily mounted on airplanes.

In 1994, the INMARSAT Assembly proceeded to amend its charter to create the International Mobile Satellite Organization (IMSO) that would address the needs of maritime and aeronautical satellite communications for safety. In 1998, it also spun off the new "privatized" commercial Inmarsat Ltd. that would own and operate the Inmarsat satellite system and would provide commercial mobile satellite services.

During the early 1990s, the Motorola Corporation initiated a project to provide a global satellite network to provide land mobile satellite services on a global basis. From the outset, officials from Motorola met with officials of both Intelsat and Inmarsat to explore whether either organization would like to engage in a joint venture to deploy such a global land mobile satellite system. Intelsat and Inmarsat both declined, but in the case of Inmarsat it decided to not only privatize and commercialize its maritime and aeronautical satellite services under the name

Inmarsat Inc. but also to create an entirely new commercial organization first known as the International Circular Orbit (ICO) Ltd.

This new ICO corporation was capitalized by an Initial Public Offering (IPO) and had the objective of deploying a global land mobile communications service. The Secretary General of INMARSAT before it was privatized and restructured as a private corporation, Mr. Olof Lundberg, decided to resign as head of INMARSAT and to become the head of this new ICO commercial entity. At the time this new commercial land mobile satellite business seemed to be quite promising and was projected to grow more rapidly than the maritime or aeronautical satellite communications business. The prospects seemed so bright that the billion dollar IPO offering for ICO was oversubscribed. Motorola proceeded on its own and formed a new global satellite consortium of commercial partners known as Iridium to launch a low Earth orbit land mobile satellite constellation. Further the aerospace corporation Space Systems/Loral also formed yet another consortium with telecom partners around the world to launch the Globalstar low Earth orbit satellite consortium for land mobile satellite services. On the order of US $15–$18 billion were put at risk to create these new satellite systems for land mobile satellite services at the time terrestrial-based cell phone systems were expanding and maturing at a rapid pace. Unfortunately all three of the dedicated land mobile satellite systems, Iridium, ICO, and Globalstar, failed financially and commercially and the ICO system, as originally conceived, was never launched even though several satellites for this system were designed and manufactured.

In 1995, 1996, and 1997, the Iridium consortium launched, deployed, and began operating a network of 66 satellites Leo constellation that was also supported by a number of operational space satellites. The projected satellite cell phone traffic in the millions of circuits did not materialize. Marketing and licensing agreement problems and technical performance problems associated with the inability to call reliably from buildings and automobiles plus the large size of the user handheld transceivers led the Iridium system going into bankruptcy in 1998 after less than 2 years of operation. An estimated $7–$8 billion of losses were incurred by Motorola and its many international partners around the world. The Globalstar satellite that deployed some 48 satellites in a low Earth orbit constellation plus spares was deployed just shortly after the Iridium system. It also declared bankruptcy in 1998. Finally the ICO system that had purchased medium Earth orbit satellites from Boeing also declared bankruptcy after the failure of Iridium and Globalstar without ever deploying its network. The staggering losses of $7–$8 billion for the Iridium system, the $6–$7 billion losses for Globalstar, and the over $2 billion losses for ICO had a dramatic impact on the overall satellite communications industry in the late 1990s and early 2000s.

Ironically, the Inmarsat Ltd. commercial venture continued to expand its maritime and aeronautical satellite services successfully and proved to be quite financially viable. Particularly with the deployment of its latest quite powerful and large aperture Inmarsat 4 satellites, Inmarsat Inc. has managed to expand into the land mobile satellite services market in most recent years. Inmarsat has continued to deploy larger and more capable satellites from geosynchronous orbit to support all

of these services. Today, New Iridium (which is the name of the commercial entity that took over the assets of the original bankrupt Iridium consortium) and a reorganized and restructured Globalstar have both recovered from the catastrophic failures of the late 1990s and are providing global services from low Earth constellations. Further, on a regional and global basis there now are two geosynchronous-based networks. These are the Inmarsat system, already discussed, as well as the geosynchronous-based Thuraya system that serves not only the Middle East, but parts of Europe, North Africa, and Asia. These satellite networks both offer broadband land mobile services to a large number of customers. The Inmarsat Ltd. system supports maritime and aeronautical communications as well as land mobile. Iridium has ordered a second generation of satellites that will also allow higher powered and broader band services and Globalstar has also ordered new satellites to upgrade their capabilities as well.

The latest innovation in land mobile satellite services comes from nationally based services for the USA, which the Federal Communications Commission (FCC) has designated as mobile communications satellites with ancillary terrestrial component (ATC). (Note this mobile satellite service is known as Complementary Ground Component (CGC) in Europe.) This concept involves the active marriage of terrestrial cellular service (i.e., land mobile services using terrestrial towers to cover the largest urban areas) which is then linked together with extremely high-powered, multiple-beam satellites. As with all new systems there are key issues to be solved. In the case of Light Squared the key issue that has arisen is how to avoid undue interference between its ground networks and GPS satellite signals. Indeed these new types of communications satellites currently deploy the world's largest communications satellite antenna systems. These antennas effective cover with relatively high power all of the rural parts of the USA. There are two of these systems now deployed and they are known, respectively, as Light Squared (formerly Skyterra and prior to that MSV) and Terrestar. These satellites with their extremely large antennas with a total area almost equivalent to a soccer field can generate very powerful beams to support the service demands of mobile consumers anywhere outside the coverage of the terrestrial cell towers in urban areas. The Light Squared satellite with its huge multi-beam antenna is shown in Fig. 3.8.

Evolution of Regional and Domestic Satellite Systems

At the time the original Intelsat Agreements were formed, the USA was essentially the only source of launch services since the Soviet Union chose not to participate in the consortium and launched its own network known as the Molniya satellite system. There was considerable feeling in Europe and especially in France that the Intelsat arrangements were too much under US control. Specifically, they maintained that other satellite networks should be allowed. When the definitive arrangements were negotiated between 1969 and 1971, one of the more contentious issues was over the possibility of separate regional satellite systems and if this should occur what type of coordination process would appropriately be employed.

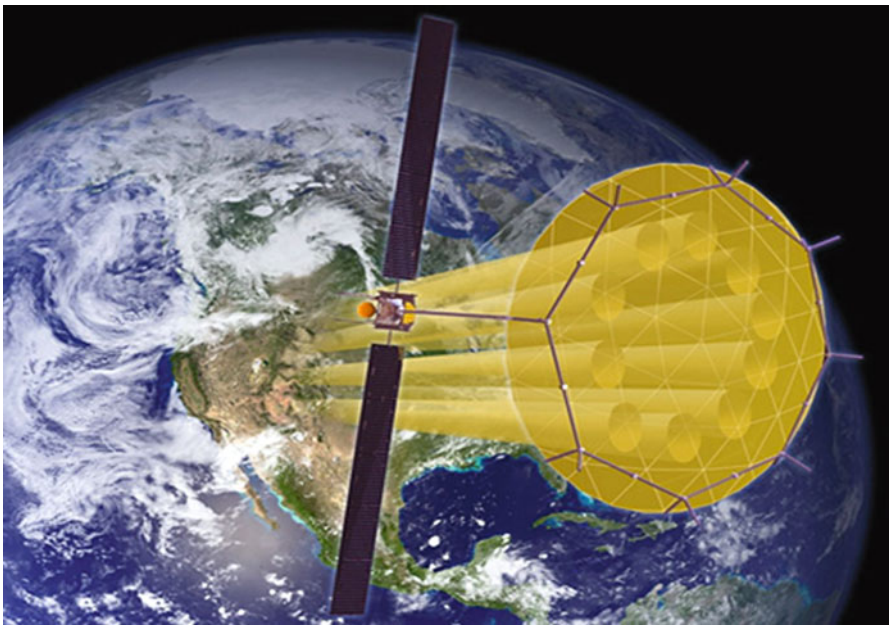


Fig. 3.8 The Light Squared land mobile satellite with ancillary terrestrial component with its huge deployable antenna system (Note: This system was formerly known as SkyTerra and before that MSV) (Photo courtesy of Light Squared)

After months of negotiations, with a block of countries largely composed of the USA and developing countries on one side and European nations largely on the other, a deadlock of opinion occurred. One of the prime barriers to agreement was Article XIV that sought to address what services Intelsat might provide in the future and the coordination processes that would be employed in the event of other satellite systems. One provision that was generally conceded to be appropriate was that there would need to be technical coordination between Intelsat and other communications satellite systems owned or operated by Intelsat members. The provisions of Article XIV(d), however, required that any separate communications satellite system would also be subject to "economic coordination" and that the members participating in another such system would need to demonstrate that there would be no economic harm to Intelsat.

The creation of Inmarsat in the late 1970s set the stage for serious consideration of what other satellite systems might be deployed. This was particularly relevant to the European region, because the European Space Agency (ESA) and the French Space Agency (known as CNES) had seriously begun the development of the Ariane, a European launch vehicle. Earlier efforts to create a four-stage launcher within what was called the European Launcher Development Organization (ELDO), where different countries developed different stages of the launcher, had failed. This new effort

under a unified management and a consolidated technical design and fabrication capability proved to be successful. Thus in 1977, the agreements under which a European Telecommunications Satellite Organization (Eutelsat) could be established were signed. This organization, just like Intelsat, was initially created as an intergovernmental organization (IGO) to develop and operate a satellite-based telecommunications infrastructure for Europe. One of the Deputy Director Generals of Intelsat, Mr. Andrea Caruso, formerly of Telespazio of Italy, left Intelsat to head up this organization and the Eutelsat Agreements were not surprisingly much akin to Intelsat in their nature and many of the European members of Intelsat also were members of Eutelsat. (Also just as the case with Intelsat and Inmarsat, this organization was later "privatized" and is now a private enterprise with private equity ownership.)

The question of Article XIV coordination, under the Intelsat Agreements, immediately arose with regard to Eutelsat. Documents were presented to Intelsat by the European organization describing the technical characteristics of the proposed Eutelsat satellites and indicating how and why these satellites would not pose harmful technical interference to Intelsat satellites. The more challenging issue was that of Article XIV(d) regarding economic coordination. These documents showed the traffic currently carried on Intelsat and indicated that there was very little traffic between and among European countries on the global system. This economic coordination document indicated that only very minor potential streams of traffic such as between Norway and Turkey would be involved and that these potential streams would be much less than 1% of the traffic carried by Intelsat at that time, and the bulk of intra-European traffic for which Eutelsat was designed would very likely never be economically viable streams for the Intelsat system.

This proved to be a very contentious issue for Intelsat and its Board of Governors. It was perhaps very much the case because Intelsat officials knew that this was not only a test case, but that the decision would set precedents for other regional systems that could and indeed would follow. Finally with carefully worded language the Board of Governors and then the Intelsat Assembly of Parties agreed that if a number of restrictions were observed the Eutelsat satellite system would not constitute technical nor economic harm to Intelsat.

Eutelsat was thus clearly on its way to deploy its regional system without any significant legal or technical constrictions to its operations. It proceeded to launch its first satellite in 1983 on an Ariane launch vehicle. This was some 18 years after the launch of the first Intelsat satellite that was manufactured entirely in the USA and launched on an American launcher. Eutelsat thus demonstrated that European industry was now able to launch its own regional satellite, manufactured entirely in Europe. This Eutelsat spacecraft was launched from the Ariane equatorial-sited launch facility in French Guyana in South America. As was to be expected, a number of different proposals for separate satellite systems ensued that followed the European precedent. The next of these regional systems was called Arabsat. This regional system was designed to cover the Arab world, within the Middle East, and also covered the Arab states of Northern Africa.

The most dramatic shift to the world of satellite communications came in 1983 from filings for new satellite systems proposed to the FCC in the USA. The RCA Corporation plus two new start-up firms filed applications to create commercial satellite systems to provide international services directly in competition with Intelsat. When the first of these applications were officially filed, starting with Orion, and then PanAmSat, RCA, and others, the Intelsat Board of Governors was meeting in Sydney, Australia. At this meeting there was general dismay that such a direct attack on the single global system had come so unexpectedly. Part of the dismays originated from the fact that Intelsat was largely created through the efforts started by the John F. Kennedy Presidency in the USA.

The leadership for the Orion system came from top Congressional aides who were in close contact with the Reagan White House Office of Telecommunications Policy (OTP). The Reagan White House and the leader of OTP Tom "Clay" Whitehead were advocates of the breakup of the American Telephone and Telegraph (ATT) monopoly and its divesture to create competition that they believed competition, whether domestically or internationally, would fuel innovation and drive down the cost of service. In short the Reagan administration favored a pro-competitive policy for services both within the USA and abroad and provided a favorable attitude toward international satellite telecommunications competition.

On the other hand, the Panamsat initiative came from television broadcasters from Mexico and South America who had found the cost of television broadcasting under Intelsat tariffs and especially under the ultimate price charged by Intelsat Signatories around the world to be exceptionally high. Other broadcasters such as CNN in America had also particularly encouraged the creation of competitive satellite systems, again because they found the television broadcasting and distribution tariffs of Intelsat and its signatories to be quite expensive.

Since USA was the largest member of Intelsat and since the US Government had played a key role in the formation of Intelsat, government officials knew they had to proceed cautiously. Thus the process whereby the applications for competitive systems was considered within the US Government and discussed within the Intelsat Intersystem Coordination processes took a number of years. Lobbyist organizations and politically savvy law firms entered the fray as the applications for competitive satellite systems took center stage within Intelsat during the period 1983 through the end of the decade.

The following report from the conservative think tank, the Cato Institute, that describes events at this time and provides its views about the evolution of the thought processes that moved forward toward "competition" and "liberalization" summarizes these events and the political context that evolved from the late 1970s through the course of the 1980s.

> In the 1970s, however, U.S. telecommunications policy began to take a path that brought cold-war-era concerns about world leadership and a single global system into conflict with domestic trends favoring competition and diversity. An increasingly pro-competitive U.S. government [sic: i.e. the Reagan Administration from 1981–1988] deregulated satellite communications for domestic traffic. Later, the United States allowed its domestic satellites

to carry trans-border traffic on an ancillary basis. The latent policy conflict came to a head in 1983 when the U.S. government received applications from RCA, Orion, PanAmSat, and others to launch and operate private satellite systems that would carry international traffic in direct competition with Intelsat (Mueller 1991).

The regional systems such as Arabsat, Asia Sat, along with Eutelsat that proceeded them, resulted in successful technical and economic coordination with Intelsat through the normal Board of Governors and Assembly of Parties processes. The issue of competitive international satellite systems raised a wide range of more fundamental issues. This ultimately led to proposals to restructure and "privatize" Intelsat, Inmarsat, and Eutelsat and move from state ownership to private ownership. Inmarsat was the first to make this transition but within 2 years Eutelsat and Intelsat also soon followed. The restructure of Intelsat proved the most complicated with not only Intelsat moving from an intergovernmental organization (IGO) but also led to the spin-off of a number of Intelsat satellites to create a new European-based operator called "New Skies." This action was designed to create a more competitive market structure more quickly. The claim by competitive systems was that the official status of Intelsat as an IGO gave it an unfair competitive advantage in the market place.

Satellite systems for international or regional satellite communications services were not the only types of systems to evolve over time. Intelsat had created tariffs and commercial arrangements for a number of countries to lease spare capacity on the Intelsat system for domestic purposes. Some of this "domestic traffic" – such as traffic between the US Mainland and Hawaii and Alaska, or between Denmark and Greenland, or between France and its Overseas Departments such as French Polynesia and Martinique, of course – seemed very much like international traffic. But beginning in the mid-1970s, Intelsat began to lease spare capacity to countries such as Algeria, the Sudan, Nigeria, Saudi Arabia, Malaysia, the Philippines, China, India, Brazil, Argentina, Colombia, etc. Over time up to nearly 100 countries or territories leased capacity from Intelsat for domestic telecommunications and television services. This was "spare capacity" on Intelsat satellites that had been launched to restore service outages that might occur, but was not currently needed to provide international satellite services. These lease arrangements allowed countries to establish national long-distance telecommunications networks as well as national television broadcast networks. Not all of this leased capacity was by developing or industrializing countries. Some developed countries such as Australia, Germany, and the UK leased capacity to provide national television distribution or to otherwise supplement their terrestrial communications networks.

It was only a matter of time before the countries with the largest market needs proceeded to deploy their own dedicated satellite communications systems. The Soviet Union had indeed deployed their Molniya satellite network back in 1965. Canada was the first of the Western nations to deploy a domestic satellite network, but then a host of countries, both developed and newly industrializing, followed suit. France, Germany, the USA, Australia, Japan, Indonesia, Argentina, Brazil,

China, India, Malaysia, Nigeria, Pakistan, Spain, Sweden, Taiwan, and Turkey among others have at least one if not several dedicated national satellite systems for telecommunications services. Many of these national systems are for business services to interconnect large-scale VSAT (Very Small Aperture Terminal) networks, others are for radio and television services and some are for both.

The combination of international, regional, and national satellite systems today results in a quite large number of satellites in orbit. Virtually all of these types of satellites supporting fixed and broadcasting television and radio satellite services are in geosynchronous orbit and nearly 300 of such "Geo satellites" for communications services are in orbit today. In addition there are hundreds more of communications satellites in low Earth and median Earth orbit constellations for a wide range of services – particularly for land mobile satellite services. The operation of these satellites requires careful technical coordination of these satellites to lessen the problem of inter-satellite interference and to maintain their orbital positions with some precision, as assigned through a process established through the United Nations specialized agency, the International Telecommunication Union. Today over 100 countries around the world either lease satellite capacity to meet domestic telecommunications needs or have established one or more separate satellite systems to meet their telecommunications needs. In some cases countries have separate systems for domestic needs but are still leasing capacity to meet additional needs that they might have, such as France which uses international FSS satellites to reach several of their overseas departments in the Caribbean and Pacific regions.

Table 3.1 provides a summary of countries that are either leasing capacity from Intelsat or other international or regional systems or now have separate networks to meet domestic needs. In total, the number of countries that today rely on satellite networks for long-distance overseas communications or for domestic links constitutes over half of the countries and territories in the world.

The regional and domestic satellite systems now deployed are designed to provide a wide range of services that include telephone, data, VSAT networking, virtual private networks to support corporate tele-working, television distribution of programs to support cable television systems, and direct satellite broadcasting for both radio and television services.

Satellite Systems to Support Defense- and Military-Related Services

The launch of the Intelsat system beginning in 1965 was not the only system deployed in that year. The Molniya satellite system, launched by the Soviet Union, was the world's first domestic satellite system. The US Department of Defense also launched the Initial Defense Satellite Communications System (IDSCS). This was a series of low Earth orbit (Leo) satellites in a random orbit constellation that allowed for more or less global communications, although there

Table 3.1 Domestic satellite systems and domestic satellite leases around the world (Countries with one or more separate systems are designated with SS)

1. Afghanistan	36. Guinea	71. Paraguay
2. Algeria	37. Guyana	72. Peru
3. Angola	38. Hong Kong (China)	73. Philippines (SS)
4. Argentina (SS)	39. India (SS)	74. Poland
5. Australia (SS)	40. Indonesia (SS)	75. Portugal
6. Austria	41. Iran (SS)	76. Qatar
7. Barbados	42. Iraq	77. Romania
8. Belgium	43. Israel (SS)	78. Russia (SS)
9. Bolivia	44. Italy (SS)	79. Saudi Arabia
10. Bosnia-Herzegovina	45. Ivory Coast	80. Senegal
11. Brazil (SS)	46. Japan (SS)	81. Solomon islands
12. Bulgaria	47. Kenya	82. South Africa, Rep. of
13. Cameroon	48. Korea, Rep. of (SS)	83. Spain (SS)
14. Canada (SS)	49. Kuwait	84. Sri Lanka
15. Central Africa Rep.	50. Libya	85. Sudan
16. Chad	51. Madagascar	86. Surinam
17. Chile	52. Malaysia (SS)	87. Sweden (SS)
18. China (SS)	53. Mali	88. Switzerland
19. Colombia (SS)	54. Martinique (France)	89. Taiwan (SS)
20. Congo, Democratic Rep.	55. Mauritius	90. Tanzania
21. Congo, Rep. of	56. Mexico (SS)	91. Thailand (SS)
22. Costa Rica	57. Mongolia	92. Trinidad and Tobago
23. Croatia	58. Mozambique	93. Turkey (SS)
24. Cuba	59. Myanmar	94. Tuvalu
25. Czech Republic	60. Namibia	95. Uganda
26. Denmark	61. Nepal	96. Ukraine
27. Egypt	62. Netherlands	97. UK (SS)
28. Equatorial Guinea	63. New Zealand	98. USA (SS)
29. France (SS) + depts.	64. Nicaragua	99. Vatican State
30. Gabon	65. Niger	100. Venezuela
31. Georgia	66. Nigeria (SS)	101. Vietnam (SS)
32. Germany	67. Norway	102. Zambia
33. Ghana	68. Oman	103. Zimbabwe
34. Greece	69. Pakistan (SS)	
35. Greenland (Denmark)	70. Papua New Guinea	

J. N. Pelton, *Future Trends in Satellite Communications: Markets and Services* (International Engineering Consortium, Chicago, 2005), pp. 138–141 and from Intelsat

were some periodic service interruptions due to gaps in the satellite coverage. This initial limited capacity system led to a wide range of military communications satellites being launched, not only by the US military but by a number of other defense forces around the world in the years and decades that followed. Most of these systems are classified and most of them are deployed in geosynchronous orbit but

there are also some defense-related communications satellites in low and medium Earth orbit and even in super-synchronous orbit. In addition to the US defense–related systems there are a wide range of other satellites for defense purposes that are now deployed and in service on behalf of Russia, the UK, Spain, France, and China. In addition the Japanese Government has now authorized such systems to be deployed to support Japanese defense as well. The most complete defense-related satellite networks are those deployed by the USA. These satellites are collectively known as the Military Satellite Communications Program (MILSATCOM), and their various functions can be summarized as follows.

The MILSATCOM architecture has three major elements: (a) There is one type of satellite system for mobile tactical support services that operates in the ultrahigh frequency (UHF) band and has limited throughput capacity. (b) Another type is for long-haul protected communications. This type is represented by the three different generations of the Defense Satellite Communications System (DSCS), sometimes known as Discus. (c) Thirdly there is the type of satellite for Wideband Defense Communications Services, known as the Military Strategic, Tactical, and Relay (or MILSTAR) satellites (Fig. 3.9).

The latest version of this type of architecture is the Wideband Global Satellite (WGS) network. This was once known as the Wideband Gapfiller Satellite. These satellites today represent the most capable and most rapid throughput system in the US military communications satellite network. There were also plans until 2009 for a so-called Tranformational Satellite System (T-SAT) that would provide world-wide connectivity to the Global Information Grid (GIG). The GIG is the name for the entire global network of all forms of ground, air, and space communications systems for the US military. The T-Sat system is now on hold. Figure 3.10 provides an integrated view of all of these US defense–related space communications systems, starting with the "narrowband" systems and ranging to the "protected" and then the "wideband systems." Except for the Advanced Polar Satellite and the planned low Earth orbit constellation Mobile User Operator System (MUOS) all these satellites are in geosynchronous orbit.

Despite all of this considerable global communications satellite capability, there still are gaps in the information and communications networks of the US satellite defense–related systems. The reason for these gaps is, in part, because military conflicts occur in different and sometimes unexpected parts of the world. The solution that has been found by the US and other defense forces around the world has been to rely on commercial satellite systems. This so-called dual use of commercial satellite facilities has been to adapt commercial systems to military, defense, or emergency rescue use as special needs and demands arise.

For many decades now, going back to the 1970s, commercial communications to support defense-related services have been leased from satellite operators around the world. These "dual use" services support routine communications or even television or radio broadcasts to overseas personnel. In many of the applications special capabilities such as jamming and "rad hard" protection (i.e., capability to survive severe radiation) are not required. For these and other reasons commercial services can often meet demand and do so at lesser expense. Specific examples of such

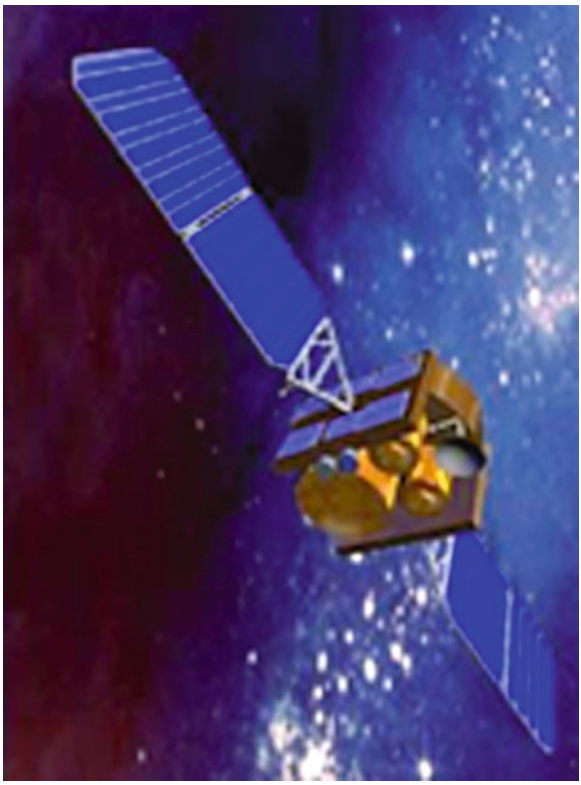


Fig. 3.9 The Phase III, or third-generation, DSCS satellite (Graphic courtesy of the US department of defense)

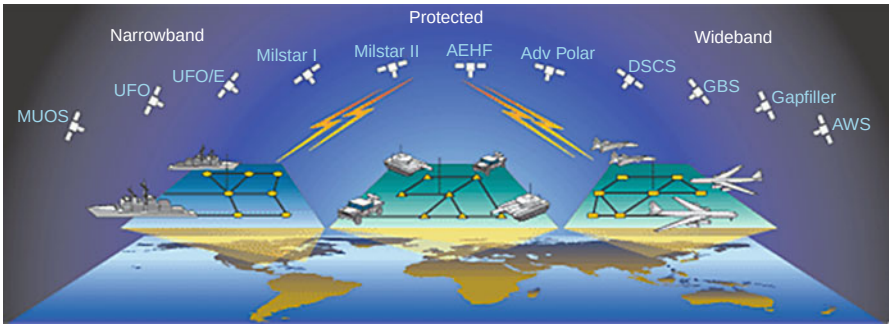


Fig. 3.10 US military communications satellite systems (Credit to US department of defense)

"dual use" satellites by the US Department of Defense (DoD) are the distribution of entertainment to overseas troops or to support e-mail and video messages to families. Such services often require a good deal of bandwidth but not special security.

Other "dual use" applications such as reliance on commercial mobile satellite services in Afghanistan and Iraq, however, often represent more strategic defense

communications. The Enhanced Mobile Satellite Services (EMSS) project provides the DoD with a secure, global, handheld communications capability. DoD now employs New Iridium, Globalstar, and the Inmarsat 4 commercial mobile satellite systems, among other systems, to provide key "in-the-field" mobile communications service. These commercial mobile satellite systems thus provide connectivity to the Defense Information Systems Network (DISN).

More detailed information about the dual use of commercial satellite systems around the world and the specifics of how these arrangements work is provided later in ▶ Chap. 2, "Satellite Communications Overview." This discussion also provides detailed information about how other countries have found new and innovative ways to meet their military satellite communications needs in conjunction with commercial suppliers.

Direct Broadcast Satellite Systems

The Intelsat organization evolved a system to sell television services over extended periods of time rather than just on a short-term basis. Although Intelsat from the outset sold telephone and data circuits on a full-time basis, it began by selling television services on a minute-by-minute basis with a 10-min minimum. This was because television commanded a great deal of the satellite capacity and indeed the entire satellite for Intelsat I and II. During the Intelsat IV and Intelsat V era, however, Intelsat evolved a new charging principal of selling spare capacity for domestic systems.

This type of sales began with Algeria in the mid-1970s, and in the 1980s Intelsat decided to sell full-time transponders to support television service. This full-time television service lease began with Australian broadcaster Kerry Packer and quickly expanded to other types of full-time television leases around the world. This led to the idea of forming a company that could lease a full-time transponder and then sell spot capacity to broadcasters. The first company to do this was named Brightstar, a joint venture of Western Union in the USA and Visnews in the UK (Visnews was later acquired by the Reuters news agency). In time, other companies such as Wold International, Bonneville, AP_TV, and Globecast (the merged entity that combined the holdings of Wold and Bonneville) expanded this type of satellite television business greatly. These services provided the "distribution" of television news, sports, and entertainment among terrestrial television systems worldwide. In short, television programs and news material was delivered to cable television networks, to terrestrial microwave distribution systems, or over-the-air terrestrial broadcast systems for distribution to consumers. A number of scientists, engineers, and businessmen around the world and especially in North America and Europe began to ask why not send the satellite signal directly to the home and bypass the terrestrial networks. As domestic satellite systems were deployed in Canada, the USA, Europe, and Asia, many consumers, particularly those in rural and remote areas without access to cable television or over-the-air broadcast began buying backyard television receive only (TVRO) dishes to receive satellite TV. In many cases they also bought "descramblers" and

began to watch so-called premium channels such as HBO and Cinemax. After a few years there were literally millions of these backyard dishes and it became obvious that so-called direct broadcast satellite television services would be a viable business.

Indeed today the direct broadcast satellite service represents the largest commercial satellite market as measured in the size of its global revenues. Because of this history, whereby the initial service evolved from Fixed Satellite Service (FSS) satellites providing service to backyard dishes and then new types of more powerful direct broadcast satellites designed for direct-to-the-home services (DTH) which evolved later, there is a confusion as to where one service ends and another begins. Today DTH services covers both types of services, namely, backyard dishes that can obtain programming from so-called FSS satellites that operate in several downlink frequency bands and Broadcast Satellite Services (BSS) spacecraft that operate in another downlink band.

BSS is the formal terminology used by the International Telecommunication Union (ITU) for this direct-to-the-consumer television service. This BSS offering is considered by the ITU to be a separate type of offering. Thus for BSS offerings higher transmit powers are authorized, different frequencies are allocated for the downlinks, and the user terminals are typically much smaller and compact and accordingly cost less money than the backyard dishes. BSS terminals are typically 30 cm to 1 m in size while backyard dishes can be 3 m to even 7 m in size.

The "true" BSS direct broadcast service is one in which many television channels are uplinked to a broadcast satellite in geosynchronous orbit and then downlinked at very high satellite broadcast power via a highly concentrated beam to a country or region so that the signal can be directly received by quite small home- or office-mounted satellite antennas. Under the ITU allocation of frequency bands for this type of service there is a different spectra used in different parts of the world. These high-power downlink transmission bands are as follows: The spectrum 11.7–12.2 GHz is allocated in what is known as ITU Region 1 (this region includes Europe, the Middle East, Africa, and parts of Russia). The downlink spectrum of 12.2–12.7 GHz is allocated for ITU Region 2 (this region includes all of the Americas). Finally the spectrum band of 10.7–12.75 GHz is allocated for downlinking BSS services in Region 3 which includes Asia and Australasia.[6]

Many countries, especially developing countries and those without satellite communications capacity, thought that the initial process by which frequencies were allocated for FSS services in Special and Extraordinary sessions of the World Administrative Radio Conferences that were held in 1959 and 1963 were biased in favor of the most advanced countries. Thus when sessions were held to allocate frequencies for this new Broadcast Satellite Service in the 1970s, a number of countries supported the idea that allocations of frequencies for this service should also include allotments for countries that did not yet have satellite capabilities against their future needs.

Thus in the 1977 international BSS Plan each country (at least in Regions 1 and 3) was allocated specific frequencies at specific orbital locations for domestic service. In this ITU negotiation process a number of BSS channels were assigned for specific countries regardless of their current ability to launch and deploy BSS satellites. This

was not the result for Region 2, however, because the USA in particular contended that this was too arbitrary of a process and that some allocations would likely never be used. Thus in Region 2 allotments ended up being made on the basis of actual need with the opportunity for new entries being accommodated as new systems have arisen.

Not surprisingly, the much more static plans for Regions 1 and 3 have needed to be amended as a result of changes that resulted from the shift from analog to digital technology, the shift of national systems to regional coverage, and many other changes (including accommodating new countries that have emerged in Eastern Europe, Africa, and Asia). Satellite systems that provide this type of service today include the BSkyB system in the UK and Europe, Europesat and Eutelsat Hotbird Satellites in various parts of Europe, plus the latest Astra satellites by SES. In the USA these systems include Dish, DirecTV, and SES Americom services; Anik and other DBS systems in Canada; Sky Perfect JSAT and NHK in Japan; and Insat, Koreasat, Asiasat, and Chinasat satellites in Asia. (See the section on broadcast satellite markets for more complete details. Also see the Section on ITU allocations for more complete details of how this process actually transpires.)

In all there are over 20 broadcast satellite systems around the world transmitting many thousands of television channels to individual subscribers. Virtually all of these are now digital television channels and an ever-increasing number of these are High-Definition Television Channels.

Today there is increasing clarity about the Broadcast Satellite Service (BSS) or Direct Broadcast Satellite systems in terms of what frequencies they use around the world and the digital transmission standards they use. These broadcast systems virtually all use efficient digital compression standards in order to send more television channels through available transponder capacity. These systems use the standards developed by the so-called Motion Picture Expert Group or MPEG standards. These are known as the MPEG 2, 4, or 6 digital compression standards. When this type of broadcast television service via satellite first emerged there was considerable debate about which analog standard would be used and whether a global high-definition television standard for use by BSS could be adopted. The conversion to the more efficient digital transmission systems and the development of the MPEG 2, 4, and 6 standards have served to bring standardization to the BSS world.

In the late 1970s and early 1980s, the European Space Agency was interested in launching an experimental direct broadcast satellite that was over time given various names, i.e., L-Sat, then H-Sat, and eventually Olympus. This project was complicated by a procurement process in which the French and German governments decided, after the contractor was selected without a participant from their country, that they would not sign on to fund their "voluntary allocations" associated with this project. Instead they decided to proceed with their own joint project known as TV-Sat in Germany and the FR-3sat in France.

As these various "official projects" proceeded to develop direct broadcast satellites in accord with ITU allocations, the Luxembourg-based company known as SES decided that it would use a high-powered Fixed Satellite Service (FSS) satellite to distribute a quasi-direct broadcast service directly to consumer homes and multi-dwelling units. SES proceeded to provide this FSS-based quasi-direct broadcast

television service via a Ku-band satellite that had complete coverage for Europe. They provided what became known as direct-to-the-home (DTH) service. With this service consumers could have small antennas installed at their home and receive a wide range of television programs – far wider than that offered by national terrestrial broadcasters.

In fact, it turned out that consumers cared little about ITU allocations or service definitions. In short SES, via its Astra satellites, stole a march on the official BSS alternatives. In the UK, the company known as Sky Television PLC also designed and launched a BSS system on an Astra platform. When there were some financial difficulties, Sky Television merged with another new project, namely, British Satellite Broadcasting (BSB) backed by Rupert Murdoch and this TV service became known as BSkyB. Today direct-to-the home (DTH) television satellite services is the largest identifiable satellite market and has grown consistently around the world in both developed and developing economies. Nevertheless actual BSS or DBS systems continue to grow and have become predominant in the USA and Japan. In Europe, the competitive dual between BSS systems and DTH systems continue on in a significant way.

Satellite Radio Broadcasting

Satellite Radio Broadcasting Services represents an important additional aspect of the satellite broadcasting industry today. Commercial satellites from the earliest days were able to use broader band channels to send high-quality audio, music, and radio shows from one location to another. As the transition was made from analog to digital satellite transmission, the idea that satellite radio systems might be deployed came into much clearer focus. The motivations for this type of service came from many different perspectives. One motivation was from "official" national radio broadcasting systems that might be characterized as sending favorable "propaganda" on behalf of one country or another. During the Cold War years very large amounts of money was spent on establishing high-power terrestrial broadcast systems to send music, entertainment, and news to locations across "Cold War boundaries."

It was not surprising that many satellite planners thought that a global or regional satellite system that was able to broadcast to small compact shortwave radio receivers could be a technically more efficient and cost-effective method to send this information to millions of listeners. From another perspective, broadcasters, educators, and news people in the developing world recognized that there were more radios than television sets in some of the poorest countries. They believed that a satellite radio broadcasting services might be an effective way to reach a new and broader audience in these parts of the world. In the most economically advanced areas, entrepreneurs envisioned that a satellite radio broadcasting system might be a way to reach a broad new audience in their automobiles and even home listeners and office workers who wanted high-quality news, entertainment, and sports on a commercial-free basis by paying just a small monthly subscription fee. For these various reasons the ITU allocated frequencies for a satellite broadcast service that is

variously known as Digital Audio Broadcast Service (DABS), Direct Access Radio Service (DARS), and Broadcast Satellite Service-Radio (BSS-R).

It took over a decade of efforts to get such a new allocation for the service through the ITU processes. Part of the difficulty was that one needed to be able to have reliable service to a very small and low-cost receiver to make this offering viable. Allocations in the higher frequency bands would have problems with "rain fade" during times of very rapid rain rates. Also, signals in these regions, because of very, very small wavelengths, would need to be in direct line of sight connections to the satellite. This would make reception in an automobile or inside a building quite difficult if not impossible. The lower bands used for mobile service, i.e., UHF (300MHz to 3000MHz) do not require direct line of sight unlike the SHF band (3GHz to 30 GHz) and especially the EHF band (30GHz to 300 GHz) which require direct and uninterrupted access. After a very lengthy process, the result was to allocate the 2.3 GHz frequencies in North America for downlink broadcasting of this service. (These frequencies are sometimes known as being in the S Band.) In the rest of the world the lower frequencies for downlinking in the 1.4 GHz band were allocated to the Digital Audio Broadcast services (this is sometimes known as the "L" band). These UHF frequencies are well suited for sending signals to vehicles and other locations without having a direct line of sight to the receiving antennas and are not subject to any significant atmospheric disruptions even with high rain rates, snow, or fog.

Once a frequency allocation for satellite radio broadcasting was finally agreed a number of companies actively pursued this business. At the lead was a company called Worldspace which was headed by a charismatic Ethiopian visionary, Noah Samara, who had actually been a leading advocate of this new radio service and who had led the fight for new satellite frequency allocations within the ITU processes. Worldspace proceeded with the immediate design, manufacture, and launch of Worldstar satellites in partnership with the French firm of Matra Marconi. The first of these launches put Worldstar 1 into geo orbit to provide this new radio broadcasting services with coverage for all of Africa, the Middle East, and parts of Europe.

The business model for Worldspace was to offer a lease of one or more individual radio channels to broadcasters who would then provide their own programming for these satellite broadcasting downlinks. These channels could be used not only for radio news, sports, and entertainment, but at nighttime (or even daytime) might optionally be utilized to provide educational programming. Indeed these digital radio channels could even be used to download short video educational programming over longer transmission periods. (This actually required a period of some hours to send a "slower and narrower signal" using the smaller "digital pipe".) The cost of the radio sets and the attendant national import tariffs (that doubled the cost of the satellite radio receivers) ranged from about US $100 to US $200. This was unfortunately a high enough cost to create an unsuccessful business model. As a result Worldspace encountered major financial difficulties almost from the very beginning.

In the USA, two systems known as Sirius (using highly elliptical orbits) and XM Radio (using geosynchronous satellites) were licensed by the FCC and both

companies managed to successfully deploy radio broadcast satellite systems. The very high cost of building and deploying very large aperture satellites for this type of service plus high overhead and programming costs led to financial difficulties for these satellite radio services as well. This type of service was marketed largely to automobile owners on a subscription basis. Consumers were offered, on a 24/7 basis, a very wide range of radio programming and emergency communications and antitheft services. XM was offered via General Motors automobiles and Sirius was offered via Chrysler and Ford automobiles. When the financial difficulties mounted these two systems merged in 2009 with XM Radio essentially acquiring Sirius.

In Europe, there were also plans for a radio broadcasting satellite service, but the financial problems with the other systems delayed the deployment of a European system.

Economic and Political Evolution of Global Telecommunications

The history of satellite communications is most often told in terms of the evolution of the technology that has developed rather continuously for the last half century. This has allowed the creation of larger, more powerful, and more proficient spacecraft with longer lifetimes and greater opportunity for automated operation. The evolution of the technology has allowed the ground devices to become simpler, easier to use, and less costly. It is quite remarkable that the initial communication satellite Earth stations were 30 m giants that require a 24 h a day crew of 40 people or so and today, almost a century later, there are handheld satellite transceivers that any individual can purchase and carry around much like a cell phone. The development of the technology and the allocation of new frequency bands has also allowed the creation of a wider and wider array of satellite communications services that include fixed satellite services; large-scale networking among multinode corporate business satellite networks; broadband digital services based on the Internet Protocol (IP) standards; various types of television, audio, and radio broadcasting and media distribution; aeronautical, maritime, and land mobile satellite services; various types of search and rescue services; as well as various types of satellite communication links to support other satellite applications for remote sensing, precision timing, satellite navigation, meteorological and geodetic services, as well as scientific satellite missions.

The history of satellite communications also has an important economic, political, and regulatory dimension that is important to recount as well.

The International Telecommunication Union (ITU) made its first important effort to bring international order to the allocation of radio frequencies for the purpose of satellite telecommunications when it convened the Extraordinary Administrative Radio Conference in 1959 to address the issue of allocation of radio frequencies for the purpose of satellite communications. The ITU had for many years previously held its periodic World Administrative Radio Conferences (WARCs) to allocate frequencies for terrestrial radio-wave applications and applications such as radio

astronomy and it continues to do so although the name is now simply World Radio Conferences. The ITU is actually the largest specialized agency of the United Nations with over 200 member countries and its role is to internationally coordinate all matters related to telecommunications and all forms of broadcasting such as technical standards, radio frequency allocations, interference mitigation, and telecommunications development for developing countries.

With the launch of Sputnik in 1957, the issue of allocation of frequencies for satellite usage was clearly an important new matter to address. At that meeting initial radio frequencies were agreed and another WARC in 1963 established a more mature framework. The International Frequency Registration Board (IFRB) was assigned the responsibility for recording each national administration's allotment of frequencies for satellite communications and they also recorded the orbital locations assigned to geosynchronous satellite networks and the orbital characteristics of low or medium Earth constellations. The ITU also established the procedure for registration of national, regional, or global satellite networks. In the case of regional or global networks, a member country is designated to provide the information to the ITU. For instance, by way of example, the USA was designated to register intersystem coordination information for Intelsat and the UK was designated to register information on behalf of Inmarsat. The ITU also established procedures for the circulation to all members of official notices concerning new satellite system registration and procedures. This process also defined how technical coordination would be conducted in the event there were concerns with regard to technical interference between or among the various satellite systems in close proximity to the new satellite network.

At the start of the satellite age most countries assigned the responsibility for post, telephone, and telegraph (PTT) services to a government ministry that had the monopoly right to provide these services. When the Intelsat agreements were negotiated these international agreements had to be structured into two parts – one agreement for the governments and another for the operators. At that time most operators were Ministries of the PTT although there were a few commercial companies such as Comsat (USA), Telespazio (Italy), and KDD (Japan) as well as what were called "crown companies" such as the COTC (Canada) and OTC(A) (Australia). The "Operating Agreement" allowed for the commercial operators within Intelsat to assume an official role.

The dramatic increase in capacity that communications satellites represented over the transoceanic submarine cables in the late 1960s led to a significant decrease in the cost of overseas calls. The initial annual cost of an Intelsat two-way telephone circuit was set at $64,000 in 1965 when service first started, but within 7 years this rate had dropped to $8,000 and continued to drop as satellite capacity increased and satellite lifetime was also extended. The introduction of digital service (i.e., Time Division Multiple Access (TDMA) multiplexing) to replace analog service (i.e., Frequency Division Multiple Access (FDMA) multiplexing) drove costs and pricing even lower. At first there was a thought that the high capacity, low cost, and multiple destination satellites might replace submarine cables altogether as satellite costs and pricing plunged. But in the 1980s

and particularly the 1990s, coaxial submarine cables grew in performance and throughput quite rapidly as well. Then fiber-optic submarine cables together with digital multiplexing and what was called wave division multiple access (WDMA) served to give a cost-efficiency edge back to the cable side of the telecommunication industry – at least for all of the heaviest routes of traffic.

In the 1950s through the mid-1960s, international overseas telephone line connections worldwide were measured in hundreds of circuits and an international call could be $15 a minute or more. By the end of the 1990s, the overseas connections were measured in the hundreds of thousands and the cost of an international call had dropped to levels equivalent to national long-distance calls. Today subscribers who use Voice over Internet Protocol (VoIP) services can call all over the world at virtually no additional cost other than a monthly connection fee. The annual cost of a submarine cable or an international satellite telephone circuit is so low (now well under US $50 per annum) that it is a minor part of the cost structure. Today the major costs associated with an international telephone or data connection, whether it is via cable or satellite, relate to marketing and billing and not to transmission costs. If the automobile cost curves had followed the satellite and submarine cable industry efficiency gains on a proportionate basis over the last 50 years, one could today purchase a Rolls Royce for under $100 and one could drive over 100 km on 1 L of gasoline.

Much has been written about the economic competition between submarine cable systems and satellite communications networks, but in many ways there has been a codevelopment of both systems. In many ways these systems have been complementary as often as competitive. First of all the two systems over time have tended to be mutually available for emergency restoration of service. There have been so-called Mutual Aid Working Groups (MAWGs) that have coordinated the ability to switch from one facility to another in case of loss of service and to respond to various emergencies that might occur. Cable systems (both in national terrestrial systems or international submarine systems) are quite vulnerable to earthquakes, volcanoes, or other natural disasters while satellites are not. Japan after the great Kobe earthquake that disrupted most communications and transportation on the southern part of Honshu decided it must undertake a fundamental change. It thus undertook to create a satellite system operating to a new network of Earth stations provided just to provide emergency backup in the case of natural disasters. These emergency Earth stations are largely installed at post office buildings all over the country of Japan.

Submarine cables with very high efficiency fiber-optic transmission and dense wave division multiplexing (DWDM) are extremely efficient for very heavy telecommunications traffic between the USA and Europe. When it comes to television distribution or broadcasting over very large areas that are thinly populated, fiber-optic networks on the other hand are not well suited to such applications. Satellites, of course, also excel over cable systems for large-scale networking and mobile applications. Table 3.2 and 3.3 both indicate the relative strengths and weaknesses and the relative performance levels of satellite networks vis a vis fiber-optic cable systems. Table 3.3 in particular provides a comparison of

Table 3.2 Comparing cable and satellite networks strengths and weaknesses (Chart courtesy of the author)

Relative performance strengths of communications satellites and cable systems			
Communications satellites		Coax and fiber-optic cable systems	
Strength	Weakness	Strength	Weakness
Mobile services	Point-to-point trunks	Point-to-point trunks	Mobile services
Large-scale business networks	Thick routes for concentrated cities	Dense urban networks	Dynamic large-scale business networks
Multicasting	Fixed distribution	Fixed distribution	Multicasting
Broadcasting for TV and radio	Fixed distribution in small geo. areas	Fixed, urban TV, or radio distribution	Broadcasting over very large areas
Connecting rural and remote areas	Intra-urban services	Intra-urban services	Connecting rural and remote areas
Dynamic networks with changing nodes	Stable and fixed node networks	Stable and fixed node networks	Dynamic networks with changing nodes
System reliability and rapid restoration	Systems with low tolerance for delay	System reliability and rapid switch over to backup systems	Vulnerable to natural disasters, construction digging, etc.

the cost-efficiencies and performance of different types of satellite systems when compared to cable.

Cable and communications satellite systems strengths and weaknesses more often than not complement each other. The advent of the World Wide Web and a global Internet has only strengthened this ability for the two technologies to reinforce each other. The other way of viewing the relationship between communications satellites and broadband cable systems is to examine technical performance. The following table shows the technical performance of a typical satellite and a typical cable transmission system. These performance levels will vary from year to year and from system to system around the world, but the relative scale has tended to be similar for the last decade or so. Fiber systems have considerably faster throughput and higher quality of service as measured in bit error rates, but their reliability (or system availability) can be less than satellites, particularly in transoceanic submarine cable installations. Satellites can be reconfigured more rapidly and are at their strongest for broadcasting, networking, or mobile services.

The advent of the Internet changed the world of global telecommunications more than any single factor in the past two decades. More and more telecommunications systems are digital and use the IP protocol to support every service whether it is for telephone, data, television, high-definition television, or mobile telecommunications. The world of satellite communications has had a more difficult time adjusting to this IP-based digital service because of the satellite transmission delay associated with geosynchronous satellites. New techniques to compensate for this transmission delay and the creation of IP over Satellite (IPoS) standards have allowed satellite systems to adapt to the global use of IP protocols.

Table 3.3 A technical comparison of satellite communications and broadband cable systems (Chart copyrighted by author and provided as a courtesy by the author)

Comparing satellite and fiber characteristics				
Capability	Fiber-optic cable systems	Single GEO satellite in a global system	Single MEO satellite in a global system	Single LEO satellite in a constellation
Transmission speed	10 Gbps–3.2 Tbps	1 Gbps–10 Gbps	0.5 Gbps–5 Gbps	0.01 Gbps–2 Gbps
Quality of service	10^{-11} 10^{-12}	10^{-7} 10^{-11}	10^{-7} 10^{-11}	10^{-7} 10^{-11}
Transmission latency	25–50 ms	250 ms	100 ms	25 ms
System availability	93–99.5%	99.98% (C-Ku band) 99% (Ka band)	99.9% (C-Ku band) 99% (Ka band)	99.5% (C-Ku band) 99% (Ka band))
Broadcasting capabilities	Low	High	Low	Low
Multicasting capabilities	Low	High	High	Medium
Trunking capabilities	Very high	High	Medium	Low
Mobile services	None	Medium	High	High
Cost-efficiency	Very high ($5–$10 per year per transoceanic voice channel)	High ($20–$50 per year per transoceanic voice channel)	Low (more than $100 per year per transoceanic voice channel)	Low (more than $150 per year per transoceanic voice channel

J.N. Pelton, *The New Satellite Industry: Revenue Opportunities and Strategies for Success* (International Engineering Consortium, Chicago, 2002)
The rate of 3.2 Tbps assumes 40 monomode fibers with each being able to transmit 80 Gbps

Conclusion

Satellite communications have now been in commercial service since 1965 and over 500 telecommunications satellites in a variety of low, medium, and GEO orbits provide every conceivable service for commercial telephone, data, networking, audio, and video service. These satellites heavily support governmental and military communications as well. This history has been punctuated by several key drivers that have led to the rapid evolution of communications services. These drivers include the following.

- *The rapid evolution of the satellite technology*. These developments have allowed satellites to become more capable, higher in capacity, more powerful, longer in lifetime, and able to support more and more services as the user devices on the ground, on the sea, or in the air have become simpler, smaller, and lower in cost.
- *The evolution of new satellite services*. Today satellite communications support fixed and mobile telephone and data services plus television and

radio distribution and broadcast services. These are known under their official ITU definitions as the Fixed Satellite Services (FSS), the Broadcast Satellite Services (BSS), and the Mobile Satellite Services (MSS) for aeronautical, maritime, and land mobile. There are also other communications satellites services such as short messaging services (also known as machine to machine (M2M) relay) and hybrid communications services directly linked to space navigation services but these represent a very small percentage of the overall market revenues for the industry.

- *The development in the satellite world of a competitive services industry.* Although commercial services started with one global enterprise known as Intelsat, the demand for different types of services, a global change in telecommunications regulation that brought about competitive systems, and market innovation have all led to many competitive systems in the satellite world.
- *The development of special needs to use satellites to support military- and defense-related purposes.* These satellites tend to have special features such as antijamming and radiation hardening and operate in different frequency bands.
- *The parallel and competitive development of satellite communications and broadband cable systems on the ground has reshaped both industries.* Satellite communications initially had greater capacity than telephone submarine cables and could offer services at lower rates. Then fiber-optic technology reversed the trend. The different strengths and weaknesses of satellites and fiber-optic technology have led to a parallel deployment of these systems that are sometimes quite complementary and sometimes competitive.

The decision to use satellite communications systems today hinges on many factors such as the type of service needed, whether the service demand is for heavy (and concentrated) or thin streams of traffic; whether the service involves two-way traffic, multicasting, large-scale networks, or a broadcast service; and whether the service is to a fixed location or a mobile service. Another factor is whether the service is to the same location or whether it is a dynamic network with nodes being added or subtracted on a continuous basis.

This is an overview history of the satellite communications industry since it started some 50 years ago. This history provides many of the key events and indicates the general technical, operational, regulatory, and market trends. For specific details of historical events one can also consult various sources, as noted in the endnotes below (Pelton et al. 2004; Pelton and Alper 1986; Whalen 2002; Logsdon et al. 1998).

Cross-References

- ▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ▶ Fixed Satellite Communications: Market Dynamics and Trends
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends

► Satellite Communications Video Markets: Dynamics and Trends
► Satellite Orbits for Communications Satellites
► Space Telecommunications Services and Applications

Notes

1. History of the Transistor, http://www.inventors.about.com/library/weekly/aa061698.htm
2. Op cit. J. Logsdon et al. 1998, p. 42
3. Intersputnik International Organization for Space Communications, http://www.intersputnik.com/
4. The International Telecommunication Satellite Organization, http://67.228.58.85/dyn4000/itso/tpl1_itso.cfm?location
5. The Creation of the International Maritime Satellite Organization (IMSO) in its current form, www.imo.org/conventions/contents.asp?doc_id=674&topic_id=257
6. ITU Radio Regulations, Article 1, Definition of Radio Service, Section 1.25, http://www.ictregulationtoolkit.org/en"practiceNote.aspx?id=2824

References

A.C. Clarke, The space station. Wireless World, London, Oct 1945
J.M. Logsdon, R. Launius, D.H. Onkst, S.J. Garber, Exploring the Unknown Volume III: Using Space. (NASA, Washington, DC,1998)
J. Pelton, *Global Talk* (Sihjthoff and Noordhoff, Alphen aan den Rijn, 1981), pp. 15–16
J.N. Pelton, J. Alper, *The Intelsat Global Satellite System* (AIAA, Washington, DC, 1986)
J.N. Pelton, J. Alper (eds.), *The Intelsat Global Satellite Organization* (A.I.A.A., New York, 1986)
J.N. Pelton, S. Madry, How satellites service the world, Chapter 6 in *The Farthest Shore: A 21st Century Guide to Space* ed. by J.N. Pelton, S. Madry (Apogee Press, Burlington, 2010)
J.N. Pelton, J. Oslund, P. Marshall (eds.), *Satellite Communications: Global Change Agents* (Lawrence Erlbaum Associates, Mahwah, 2004)
M. Mueller, *Intelsat and the Separate System Policy: Toward Competitive International Telecommunications, CATO Institute Policy Analysis 150*, 21 Mar 1991. (CATO Institute, Washington, DC, 1991), http://www.cato.org/pubs/pa-150.html
Special Message to the Congress on Urgent National Needs "John F. Kennedy address to a Joint Session of Congress, 25 May 1961, John F. Kennedy Library and Museum, http://www.jfklibrary.org/Historical+Resources/Archives/Reference+Desk/Speeches/
United Nations General Assembly Resolution of Satellite Communications, Resolution 1721, Section P, Sept 1961
U.S. Congressional Hearings, House of Representatives, Communications Satellites Hearings, Part II, Interstate and Foreign Commerce Committee Hearings, 1962, pp. 683–685
D. Whalen, *The Origins of Satellite Communications, 1945–1965*. (Smithsonian Press, Washington, DC, 2002)

Space Telecommunications Services and Applications

4

Joseph N. Pelton

Contents

Introduction 68
Satellite Communications Services as Defined by the ITU 71
Fixed Satellite Services 71
Broadcast Satellite Services 72
Mobile Satellite Services 75
Other Types of Commercial Satellite Services 77
Overview of FSS Services: Telephony, Information Technology (IT) Services, and Enterprise Networks via Communications Satellite 80
Overview of Video and Audio Broadcast Satellite Services 83
Defense-Related "Dual Use" of Commercial Communications Satellite Systems 85
Evolution of New Digital Services and Applications 87
Limits to the Growth of Satellite Networks 87
Cross-References 91
Notes 91
References 92

Abstract

This chapter examines the ever increasing number of services and applications that are now provided by the commercial satellite industry. It explains that basic types of satellite services as defined by the ITU for the purpose of radio frequency allocations – particularly the Broadcast Satellite Service (BSS), Fixed Satellite Service (FSS), and the Mobile Satellite Service (MSS). This section further explains that regulatory, standards and policy actions by various international and regional organizations, plus commercial competition also leads to the development of different terms to describe new and emerging satellite

J.N. Pelton
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: joepelton@verizon.net

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_4, © Springer Science+Business Media New York 2013

services. Key to the development of satellite services and applications within the global telecommunications market is not only the development of new satellite technology, but also the competition between satellites and terrestrial wireless, coax, and fiber-optic networks. Satellites and terrestrial systems, despite being competitive, are nevertheless often complementary because they have particular strengths and weaknesses that do complement each other. Further these systems are also used to restore each other against outages – particularly during natural disasters. Satellites have evolved in their offerings for nearly 50 years and will continue to do so with the future including services to interplanetary distances and perhaps beyond.

Key Words

Amateur satellites services • American National Standards Institute • Broadcast satellite service (BSS) • Broadcast satellite services for radio (BSSR) • Business networks • Digital audio broadcasting services (DABS) • Direct to home (DTH) television • Enterprise networks • European telecommunications standards institute • Fixed satellite service (FSS) • International electrical and electronics engineers (IEEE) • International electrotechnical commission (IEC) • International telecommunication union (ITU) satellite service definitions • Internet engineering task force (IETF) • Internet protocol (IP)-based satellite services • Machine to machine (M2M) satellite services • Microsatellites • Mobile satellite service (MSS) • Nano satellites • Satellite telephony • Space communications for interplanetary and cislunar links • The International Standards Organization (ISO) • Very small aperture terminals (VSATs)

Introduction

The world of satellite communications services and application can seem complicated due to the fact that there are a number of service descriptors that come from quite different sources. Some of these service definitions come from the International Telecommunication Union while others come from other standards making bodies, national or international governmental regulatory agencies or perhaps most frequently from the various communication satellite industry markets around the world. Sometimes, the differences in the names of satellite services are regional in nature in other instance copyright or trademark restrictions lead to differences in terminology. Despite a difference in the names for various services, in many cases, the actual satellite service in question may exactly describe precisely the same offering. Figure 1 in ▶ Chap. 1, "Satellite Applications Handbook: The Complete Guide to Satellite Communications, Remote Sensing, Navigation, and Meteorology" indicates in a synoptic way the very wide range of commercial satellite services available today.

The International Telecommunication Union (ITU) is a specialized agency of the United Nations that has the prime responsibility for creating standards for satellite communications. It is also responsible for the allocation process for radio frequencies that are essential to satellite communications services. This organization establishes official definitions for a wide range of telecommunications satellite services and indeed does the same for the various other types of satellite applications described in later sections of this Handbook.[1] These official definitions of satellite communications services, however, are not always used by national regulatory agencies or the commercial satellite communications markets around the world. Further, there are many other standards making bodies that create standards that relate to the provision of satellite services. These bodies and their standards sometimes lead to confusion concerning service definitions and uncertainty as to whether one particular term is equivalent to another. This problem can be compounded by commercial organizations which because of copyright and trademark restrictions often resort to using different phrases to describe the same service.

There is a lot of complexity just in the global standards making arena. The ITU and its regulatory and standards making processes, plus regional standards groups, address satellite issues from the perspective of telecommunications efficiency. Then there is the Internet Engineering Task Force (IETF) that is organized under the auspices of the Internet Society to address satellite networking issues. The IETF, however, develops its standards and terminology not from the perspective of satellite communications but in the context of improving Internet connectivity standards. Other standards groups that also develop standards, and in the process sometimes develop services or service requirements, are the International Electrical and Electronics Engineers (IEEE), the International Electro-Technical Commission (IEC), the International Standards Organization (ISO), as well as national and regional bodies such as the European Technical Standards Organization (ETSO) and the American National Standards Institute (ANSI). The complications extend further in that military and defense-related organizations also often define their own terms for various specialized satellite services.

The commercial and defense-related terms for applications and services are often driven by marketing and sales personnel. These terms are thus much more dynamic and change not only year to year, but even month to month. These market-driven terms for satellite applications change for a variety of reasons such as trademark restrictions, a perceived market advantage, a new or altered national governmental regulatory restriction, a new way to set a tariff, or a dozen other reasons.

The various types of satellite communications services and applications will be first presented here in terms of ITU definitions and frequency allocations. This is simply because ITU terminology often serves as the “common language” of satellite communications and allows a basis for some global commonality when it comes to satellite communications services.

The remainder of this section provides an overview discussion and analysis of the commercial satellite communications services broken down by the various markets, including “dual use” of commercial satellite networks to support

military or defense communications services. Later in this chapter, the actual development and growth dynamics of these markets are addressed in greater depth with regard to the major ITU-defined service categories of Fixed Satellite Services (FSS), Broadcast Satellite Services (BSS), and Mobile Satellite Services (MSS). There is also a section with regard to the development and evolution of military and defense-related satellite services that are carried on commercial satellite systems.

The first presentation and service definition relates to FSS applications. This section thus describes how satellites of this type provide connection between ground antennas that are fixed in their location. These FSS satellites typically provide two-way voice communications, various types of information technology (IT) relay, data services and data networking, video and audio distribution services, and commercial defense-related satellite communications. These are all services that are provided between two "fixed" earth station antennas or a network of fixed earth stations. Later sections of this book provide background and analysis as to how FSS services and applications relate to other commercial communications that are sometime competitive and in other cases satellites play a complementary role to terrestrially based telecommunications systems. These sections thus address how coaxial cable, Ethernet systems, submarine cable, and fiber-optic networks relate to and sometimes compete with commercial satellite communications systems in terms of economics and global division of markets. Most FSS applications provide interactive communications in "real time" between two antennas or perhaps among a network of fixed earth stations. In some cases, however, the FSS applications can include so-called store and forward and upervisory control and data acquisition (i.e., SCADA) noncontinuous communications. For these types of services simple data messaging is sufficient since the satellites in this type of configuration or constellation are typically not able to "see" at the same time the various locations that are to be connected. In these types of systems, a satellite picks up a data message at one point and then delivers the stored message at another location. This is why it is called a "store and forward" system. These are also sometimes referred to as "machine to machine" (i.e., M2M) systems.

The second broad category of commercial satellite service relates to "BSS applications" and the direct broadcast of satellite television, high-definition television, and radio. In this instance, the service is uplinked to a high-powered satellite that sends a one-way signal to a large broadcast audience. These broadcast satellites are capable of sending a signal directly to the consumer where it is received by a home or office micro terminal or in the case of broadcast satellite service radio to a small receiver that is sufficiently small that it can be installed in an automobile or other type of vehicle.

The third presentation and service definition relates to MSS applications and the type of satellite service that provides interactive communications with antennas for users that are on the move and can perform mobile communications. These MSS satellites can be used for communications to and from ships, aircraft, or land mobile vehicles.

Satellite Communications Services as Defined by the ITU

The membership of the International Telecommunication Union (ITU), headquartered in Geneva, Switzerland, includes virtually every country and territory in the world. Although it is a specialized agency of the United Nations, the ITU membership is actually larger than the United Nations Organization itself. This is not surprising since today virtually everyone needs to use radio frequencies for a wide range of applications and there is a global need for information and communications technology (ICT) services. One of the many functions of the ITU is to allocate frequencies for satellite communication services as well as to help with intersystem coordination so that the various satellite networks do not provide excessive interference to one another. One of the ways that the ITU functions with regard to radio frequency allocations is to define different types of satellite services and provide for the use of different radio frequency bands for these services. The various defined services, as agreed through global meetings, now known as the World Radio Conference (WRC), are specifically identified in Table 4.1.[2]

About half of these various ITU-defined services relate to commercial communications satellite offerings. All forms of satellite activities, whether for defense- related applications, remote sensing, space navigation, satellite meteorology, radio astronomy, time synchronization, space research or space operations, need to operate active communications links to convey information to the Earth and to receive information and commands from Earth locations. ITU allocations of frequencies associated with these various satellite services is a complex activity. In most cases, there is a primary allocation of one or more frequency bands for these various services. There can also be a secondary and tertiary allocation for radio frequency spectrum (Pelton 1998).[3]

Further organizations can also use certain bands, on a noninterference basis. In a number of cases, countries will place an "asterisk" against a frequency band allocation for a particular service in their own national boundaries. This means that they do not accept this allocation within their own country. This will be explained in greater detail later in the section on ITU functions and, especially, its role with regard to frequency allocation.

All of the communications satellite services are defined by the International Telecommunication Union. In many cases, however, other terms are used to describe these same services as they are marketed within the commercial world.

Fixed Satellite Services

The "official" ITU definition of Fixed Satellite Services (FSS) is as follows: "A radio-communication service between earth stations at given positions, when one or more satellites are used; the given position may be a specified fixed point or any fixed point within specified areas; the fixed satellite service may also include feeder links for other space radio-communication services."[4]

This was the initial form of commercial service that began in 1965 with the Intelsat global satellite network for international links and the Molniya satellite

Table 4.1 ITU-defined satellite services with specific frequency allocations (Derived from Pelton 2006)

ITU-defined satellite services
Fixed satellite services (FSS)
Inter-satellite services (ISS)
Broadcast satellite services (BSS)
Broadcast satellite services for radio (BSSR)
Radio determination satellite services (RDSS)
Radio navigation satellite services (RNSS)
Mobile satellite services (MSS)
Aeronautical mobile satellite services (AMSS)
Maritime mobile satellite services (MMSS)
Maritime radio navigation satellite services (MRNSS)
Land mobile satellite services (LMSS)
Space operations satellite services (SOSS)
Space research satellite services (SRSS)
Earth exploration satellite services (EESS)
Amateur satellite services (ASS)
Radio astronomy satellite services (RASS)
Standard frequency satellite services (SFSS)
Time signal satellite services (TSSS)

network that provided domestic services within the Soviet Union. Fixed satellite services today provide applications that include telephone, facsimile, various data services, audio distribution, videoconferencing, video distribution, multi-casting services, and corporate enterprise services such as virtual private networks. One of the more rapidly growing data services on fixed satellite systems support IP networking for Voice over IP and broadband IP data services. This type of IP-based service via satellite can be broadband data transmission to support heavy "trunking" links between cities or countries or it can support relatively wideband Internet connections directly to end users at small office/home office locations (i.e., so-called SOHO connections). These commercial satellite services typically operate in the so-called C-band (6 GHz uplink and 4 GHz downlink), the Ku-band (14 GHz uplink and 12 GHz downlink), and most recently the Ka-band (30 GHz uplink and 20 GHz downlink). Additional frequencies are used for military satellite services. All these frequency band allocations by the ITU will be discussed in detail in the section of the Handbook addressing this issue.

Broadcast Satellite Services

The "official" ITU definition for broadcast satellite services (BSS) is as follows: "A radio-communication service in which signals transmitted or retransmitted by space stations are intended for direct reception by the general public. In the



Fig. 4.1 State-of-the-art Direct Broadcast Satellite-Echostar XIV – Manufactured by Space Systems/Loral (Graphic Courtesy of Space Systems/Loral)

broadcasting-satellite service, the term 'direct reception' shall encompass both individual reception and community reception" (See Fig. 4.1 above for current BSS type satellite).[5]

Despite the fact that this ITU definition of the Broadcast Satellite Service (BSS) is quite detailed, and despite the fact that specific frequencies are allocated by the ITU for this broadcast service, there are nevertheless ambiguities in terms of types of satellite services actually delivered in the marketplace. The problem is that there are overlaps that occur between BSS and FSS systems in practice when it comes to how satellites in orbit are utilized. Some very high-powered FSS systems have been used not only to distribute television signal to cable television head ends and other locations for redistribution to the public, but also to provide service directly to the consuming public. This is sometimes referred to as "direct to home" or DTH service.

The Home Box Office (HBO) system began delivering television programming to cable television "head ends" via satellite in 1975 using FSS satellites. Consumers in rural and remote areas responded to this "opportunity." They bought "backyard satellite dishes" and sometimes "decramblers" to see the video signal and hear the audio. Thus an informal type of direct broadcast satellite or direct to home service was born in the United States, Canada, and many other countries. The Astra system in Europe was the first FSS system to launch entire satellites based on the DTH business model of selling television service directly to consumers.[6]

In what might be called the "reverse situation," BSS systems have been used to deliver what have been traditionally considered the domain of FSS applications. In this case, individuals and distributed business offices have utilized BSS systems to obtain two-way high-speed data services using protocols such as Digital Video Broadcast with Return Channel Service (DVB-RCS) or DOCSIS (Digital Over Cable System Interface Standard) to deliver digital video and high-speed data "to the edge" of business enterprise networks. To obtain both television services and asymmetrical data distribution with thin route message return, the business user only needs to install very small aperture antennas (VSAAs). Such "asymmetric satellite ground systems" cannot only receive high quality digital audio and television as well as high-speed IP-based data, but they can also transmit low data rate return messages. This type of operation is almost ideally suited to support typical digital telecommunications services to an asymmetrical corporate-type data network with interactive services to the "edge" of an enterprise business hookup. Companies with national, regional, or even global business networks such as automobile manufacturers with highly distributed dealerships, oil companies with a large network of gasoline service stations, department stores with many outlets, etc. would use such satellite business networks to support communications to their corporate headquarters or to credit card authorization offices. Such entirely digital networks can be operated on either FSS or BSS networks.

In such cases, there will be a high-speed downstream blast of data at many megabits/s (typically at speeds like 30–75 Mbps) and then there would be a thin route uplink response capability. These types of asymmetrical data distribution service with a return channel link actually operate most typically on an FSS system, but can now also operate on BSS systems as well. In a telecommunications service world with data networking supporting video, audio, and high-speed data (that can be telephony or digital streaming), the division between FSS and BSS service tends to be increasingly blurred. One example is that the Boeing Corporation manufactured three high performance digital satellites known as Spaceway 1, 2, and 3. Two of these satellites are deployed by the Direct TV network to provide direct broadcast BSS services in the United States, but the third satellite in the series, the Spaceway 3, is utilized by the Hughes Network Systems (HNS) to provide high-speed digital services to customers under its Hughes Net offering. The three satellites are essentially the same, but the utilization is dramatically different.[7]

There is currently no ITU enforcement of how various satellite systems are used. Indeed, over the past decade various proposals have been made at various ITU conferences to allow more flexibility in service definitions. The argument behind these proposals is that if there were "multiple usage allocations" provided for satellite communications frequencies, this would encourage the most efficient and effective use of satellite spectrum. Currently, the reverse is true. ITU allocations for satellite services tend to be what might be called "overly precise." The ITU allocations and the frequency planning process for the BSS services tend to be geared to particular orbital locations for satellites and the available frequencies vary over the three ITU regions around the world. Detailed information on the frequencies available for this service is provided in later sections.

The ITU has also defined a service known as Broadcast Satellite Service Radio (BSSR) with frequencies allocated in a lower band to accommodate this service that is more narrowband than broadcast satellite-television services. This service can also be called simply Satellite Radio or Direct Broadcast Radio Satellite. Since most systems use digital broadcasting to support efficient transmission it can also be called Digital Broadcast Radio Satellite Systems or Direct Broadcast Radio Systems or Digital Audio Broadcast Satellite (DABS) System. In this type of satellite radio service, the audio broadcast signal is digitally encoded and broadcast to Earth-based receivers from an orbiting satellite directly to a receiver in a home, office, or vehicle. In some cases, it is sent to a "repeater station" that then rebroadcasts the signal to receivers. The Worldspace, XM Radio and Sirius Radio satellite systems operate in these bands. The BSS or direct broadcast satellite systems that operate in the higher Ku-band transmit a large number of video and high definition television channels, however, also broadcast high fidelity audio channels as well. The XM Radio and Sirius Radio satellites (now merged into a single system under the ownership of XM Radio) primarily transmit their programming to vehicles on the move equipped to receive their signals – with subscribers paying a monthly subscription fee. World space, which is nearing the end of its operations after experiencing financial and revenue stream problems, transmits its digital audio signals largely to developing countries where television service is limited or nonexistent.

Mobile Satellite Services

There are, in fact, a number of mobile satellite services that include Land Mobile Satellite Services (LMSS), Aeronautical Mobile Satellite Services (AMSS), and Maritime Mobile Satellite Services (MMSS).

The official definition by the ITU of Mobile Satellite Services (MSS) is as follows: "A radio-communication service (a) between mobile earth stations and one or more space stations, or between space stations used by this service; or (b) between mobile earth stations by means of one or more space stations. This service may also include feeder links necessary for its operation."[8]

Today there are some satellite systems, such as the ones operated by Inmarsat Ltd. or Thuraya, that have satellites with sufficient power and flexibility to support service to aircraft, ships, and land mobile units. Because these systems are equipped with enough power to operate to mobile units they can clearly also be used at fixed locations. This is because it is much easier to complete a satellite link to a fixed location where there is a clear line of sight to the satellite that to a constantly traveling mobile unit. In the case of mobile satellite service, blockages to a clear line of sight between the satellite and the mobile antenna can occur at any time due to tunnels, forestation, utility poles or buildings. Since this is the case, broadcasters that cover a wide variety of events from the field may set up a temporary "fixed" location and uplink and downlink from a mobile satellite system. Offshore platforms that are "fixed" often rely on mobile satellite systems as well.

In short, ITU definitions are precise and seek to separate the three main satellite communications services into the categories of FSS, BSS, and MSS. These exact divisions of the three primary satellite services are not always observed in practice. Since there are no "enforcement officials" or fines imposed these distinctions have no practical implications. In most cases, the prime use of the satellite system is consistent with the defined service and the spectrum band allocated to that service.

There is a new type of mobile satellite system that has originated in the United States and is under planning in other regions of the world. This is the mobile satellite service with so-called Ancillary Terrestrial Component (ATC). This type of mobile satellite service is known in Europe as a mobile satellite system with a Complementary Ground Component (CGC). The first of these satellite systems for service in the United States, namely the SkyTerra system and the Terrestar system (as pictured below), need to deploy extremely large satellite antennas. These gigantic antenna systems with huge apertures more than 15 m across allow a very large number of powerful beams to be generated from space. These systems are unique in that they are designed to operate in tandem with terrestrial mobile satellite systems. The concept is that terrestrial wireless mobile services provide high quality mobile connections within well covered urban areas, but mobile satellites are used to provide broad and complete coverage outside the city with the service seamlessly switching from terrestrial wireless to satellite as the consumer moves from one point to another.

The financial status and viability of businesses providing mobile satellite services in North America has changed dramatically in the past two years. Both Terrestar (formerly Sky Terra) and DBDS North America (formerly ICO Ltd.) experience financial set backs and both formally declared bankruptcy in 2010. Then, both were acquired by Dish Inc. in 2012 for $2.3 billion out of bankruptcy court in order to provide mobile satellite services on a consolidated basis. Light Squared has also experienced major difficulties because its frequencies and GPS frequencies have significant interference. The Terrestar 1 satellite is shown in Fig. 4.2.

The radio frequencies allocated for this service typically start at around 1,700 MHz and range up to 2,600 MHz in an assortment of bands. The lower radio frequencies allocated for this service (as opposed to FSS and BSS service) is important because of signal blockage. When mobile satellite users are on the move, there is a problem of signal blockage. Factors such as foliage from trees, buildings, signs, utility poles, and especially the roof tops of vehicles can serve to block a direct line of sight signal from mobile satellites to the user. The lower frequencies with longer wavelengths that can "bend" around obstacles are more tolerant of partial pathway blockage. These lower frequencies in the so-called L-band and the UHF frequencies can still close a link to a mobile satellite without having a direct line of sight between the satellite antenna and the user antenna device. In contrast, the higher frequencies used for FSS and BSS services typically need to have a direct line of sight to the satellite to complete a transmission.

Higher frequencies in the Ka-band are also allocated to mobile satellite services, but these are not typically used. This is because the frequencies in the

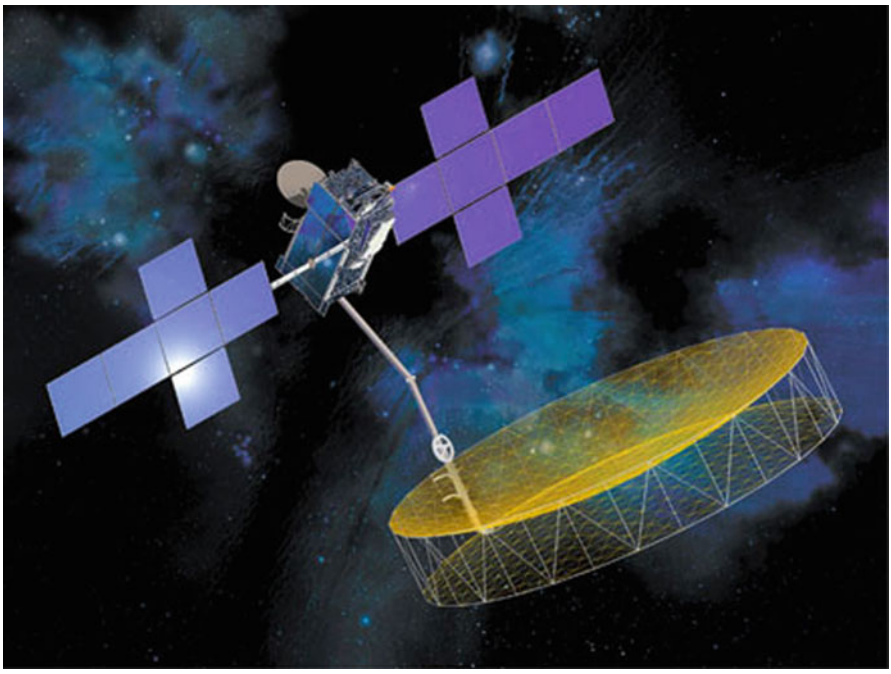


Fig. 4.2 The gigantic Terrestar 1 satellite pictured as deployed in space (Artwork Courtesy of Terrestar)

30 GHz band with quite tiny wavelengths are not tolerant of interference and essentially require direct line of sight to close a link between a satellite and a user's mobile antenna. Specific information on these allocations is provided later in Handbook.

Other Types of Commercial Satellite Services

The three most important commercial satellite communications systems are thus those providing fixed, real time broadband satellite services, broadcast satellite service, and mobile satellite services. These services represent the overwhelming amount of revenues associated with satellite communications. There are, however, some other forms of communications satellite systems that exist and provide services to various users around the world. In some cases these are thin route commercial services and in other cases these are satellite that are used to serve educational, medical, first responder and emergency communications, and amateur "ham radio" operators.

1. *Store and Forward Satellite Systems* (also known as M2M). These systems, that rely on microsatellites, are typically deployed to provide non-real time data relay. The most well known of the M2M commercial satellites systems is the Orbcomm satellite network that provides a global data relay satellite service for

what is often called "machine to machine (M2M)" communications. This is a commercial service that can relay data around the world with a minimum of delay. The M2M types of satellite system deploy ground transceiver units that can be positioned at fixed locations (such a remote mine or offshore drilling rig) or as vehicular mounted systems. These vehicular mounted units are designed not only to transmit and receive data, but they are also often paired with a space navigation system that can indicate the vehicle's location in real time as well. One of the more common application for this type of satellite network is for "supervisory control and data acquisition" (SCADA) services that are used for control of power stations, monitoring of oil, and gas pipe lines, and managing various types of mobile fleets from trucks, buses, or rail systems to ships at sea.

The Orbital Sciences Corporation designed and built the first generation satellites for this system and deployed 35 of these micro satellites into low earth orbit using its Pegasus and Taurus launch vehicles. These lightweight satellites could be "stacked" together in a launch vehicle and launched eight at a time (Fig. 4.3).

This network was first established as a part of Orbital Sciences but divested as a separate entity when the initial system experienced financial difficulties. Orbcomm Ltd. is now a separate company and independently owned. The second generation of this system involves a new constellation of 18 satellites that supplements the initial system. These satellites are under contract for launch by the Space Exploration Technology Corporation (known as Space X). This next generation of satellites were built by the Sierra Nevada Corporation and its Microsat subsidiary plus a team of contractors that included Boeing and ITT (Orbcomm 2008).

Applications associated with commercial store and forward satellite networks can be to send and receive information from trucks, buses, trains, dispatch vehicles, and even ships and airplanes, and also fix the location of the mobile unit so this can be relayed to a home office. Although different frequencies and different satellites are used for the data relay on one hand and for position location on the other the mobile unit is consolidated and relies on a common battery. These consolidated units that providing this type of service are popular with car rental agencies, bus, and rail systems and shipping lines. This machine to machine (M2M) type of service provides a reliable form of data relay that can often relay a message on a global basis in a matter of minutes and provide space navigation services on an instantaneous basis using GPS satellites. The M2M satellites typically operate in the 137–138 MHz and 400 and 435 MHz frequencies. Since these are very narrow bands, they have very limited service capacities that relay only short messages.

In addition to commercial M2M services such as Orbcomm provides, there are a number of small satellites that are used to support the mission of nongovernmental organizations operating around the world to for educational or research purposes. These satellites are often designed and built by the small satellite program at Utah State University in the United States or by the Surrey Space Center in the United Kingdom. In some cases, they are built by newly emerging national space programs that are built in cooperation with the Surrey Space Center as a way of launching their initial application satellites programs. There are also commercial companies



Fig. 4.3 One of the first generation of Orbcomm satellites that created a "fast" store and forward global LEO constellation (Graphic courtesy of Orbital Science Corporation)

that design and build so-called microsatellites, such as Microsat Systems Inc. and Space Dev who often design small satellites for defense-related missions. There are also number of small firms that manufacture nano satellites (i.e., 50–100 kg) including companies like ISIS, GomSpace, and UTIAS-SFL.[9]

An example of the small store and forward satellite systems, designed by the Surrey Space Center is the two-satellite low earth orbit "Lifesat" system. This microsat system was designed to provide medical information on demand via satellite to rural and remote parts of the world with a guaranteed response time of 2 h or less using Surrey-designed satellites. This type of microsatellite designed and built at the Surrey Space Center or Utah State University can be designed for many applications. Thus, in addition, to store-and-forward microsatellites for communications, these small satellites can also be designed to provide remote sensing, scientific missions, or other space operations. These satellites, which typically range from 10 to 100 kg, are often launched as a "piggy back" operation added on to the launch of one or more larger satellites. There are also very small "cube satellites" or "nanosatellites" in the 1–10 kg range that are typically built by students at universities to carry out short-term experiments that are capable of relaying small bursts of data communications, but because these are often restricted to battery power only, they have only a short duration lifetime. These satellites typically operate at 137 and 400 MHz.

2. *Amateur Satellites*: These satellites, generically known within the industry as Amsats but to the world as OSCARs. These are microsatellites designed to be

used by amateur "ham radio" enthusiasts around the world to relay messages. This is also essentially a "store and forward" satellite data service, but they also include voice transponders to support short bursts of conversations. Since these satellites are launched into low earth orbits, they cannot support continuous communications because the satellites come quickly in and out of range as they pass overhead of the ground based amateur satellite operators and their transmit and receive equipment. This represents one of the largest applications of this type of technology in terms of number of satellites launched and the number of the participants. These satellites are known as "OSCARs," which stands for Operating Satellites Carrying Amateur Radio. Figure 4.4 shows OSCAR I that was launched on December 12, 1961. This was the first micro satellite to be launched as a "piggy-back" operation. In short, it was launched in tandem with a much larger satellite as a much smaller ancillary or "secondary" payload. Since that time there have been many hundreds of launches wherein microsatellites were launched along with a larger primary mission. Over 70 OSCAR satellites have been launched since the original launch a half century ago.

These satellites are built typically to operate in the following radio frequencies: 28–29.7 MHz, 144–146 MHz, and 435–438 MHz (on a noninterference basis.)[10] Many thousands of amateur radio operators all around the world use these satellites to relay messages. This is not only as a hobby, but many times this has proved a very important way to relay messages during disasters and other types of emergencies where conventional communications systems may have failed or are temporarily out of service.

Overview of FSS Services: Telephony, Information Technology (IT) Services, and Enterprise Networks via Communications Satellite

The Fixed Satellite Services (FSS) began commercial service in 1965 with the Intelsat I (or Early Bird) satellite leading the way. These first commercial communications satellites were able to support telephone, telex, radio and audio service, and for the first time live international television service – although of quite low quality. Overtime satellite transmissions from space were sent with higher power and higher gain antennas and the quality of the television signal improved. These next generation satellites, as deployed a few years later, could support full color and reasonably high quality audio and television in all of the international standards then used, namely NTSC, SECAM, and PAL. The range of telecommunications and information services that are offered by FSS networks today is quite extensive. A listing of some of the most important types of services is provided in Table 4.2.

It should also be noted that commercial satellite systems also often fall into the categories of international satellite networks, regional networks, and domestic networks. International networks can, in fact, be utilized to provide not only global interconnectivity around the world, but also provide capacity, often on

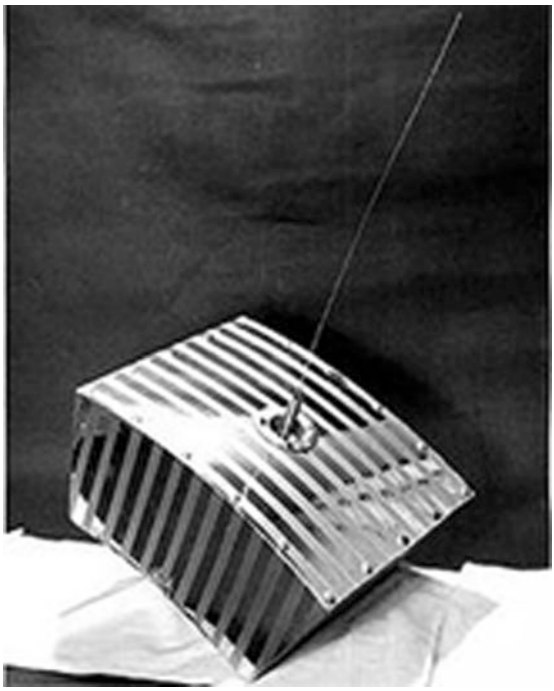


Fig. 4.4 The OSCAR 1 satellite launched in 1961

a transponder lease basis, for either regional services or domestic services. Examples of global networks are the Intelsat network and SES Global. Regional networks include Eutelsat, Arabsat, and domestic systems number in the many dozens of systems all around the world. A worldwide inventory of FSS systems is provided in the Appendices to this Handbook. When the first FSS systems were deployed they represented the widest band capability for transoceanic service and dominated the overseas markets for video, voice and data services.

The emergence of fiber-optic cable systems with their even greater throughput capabilities and lesser transmission delays in comparison to geosynchronous communications led to these terrestrial systems reestablishing market dominance for virtually all of the heavy telecommunications traffic streams around the world – especially for voice traffic. With the transition from analog to digital traffic and the growth of Internet Protocol (IP) transmissions, satellite systems have needed to make special adjustments. This is because satellite transmission delay can be "perceived" as system congestion and lead to slow recovery modes and thus transmission interference. Also IP Security (IP Sec) over satellite routing leads to difficulties when "header information" is stripped from a satellite transmission.

A great deal of progress has been made to develop new standards to allow efficient IP-based satellite transmission over geosynchronous satellites where the

Table 4.2 Examples of satellite applications divided by generic service category

FSS satellite–based telecommunications and information services
Voice- and telephony-based services
• Rural telephony
• Telephone connections for locations in hostile natural environments
• Transoceanic and regional telephony
• Remote connectivity to research stations, offshore drilling stations, mines, etc.
• Submarine cable and terrestrial telecommunications network restoration
Audio-based services
• High-quality audio and music downloads (8–32 kHz channels)
• Radio programming distribution
Video-based services
• Television distribution to cable television head ends (conventional, high definition, and 3D high definition)
• International television relay
• Remote newsgathering and transfer to television production centers
• Direct to Home (DTH) television (i.e., quasi-direct broadcast services)
• Television to ships at sea and commercial aircraft
• Video-based tele-education and tele-health services
Digital networking services
• Broadband Internet "trunking" or heavy route interconnection
• Internet services directly to the Small Office/Home Office (Soho)
• IP-based telephony (Voice over IP) and multimedia over IP
• Digital Video Broadcasting and network data distribution
• Digital Video Broadcasting with Return Channel Service (DVB-RCS) or DOCSIS services (i.e., high data rate distribution with narrow band return)
• Multi-casting IP-based services
• Enterprise networking via VSAT and micro-terminal networks
• Business television networks
• Remote office and plant management systems
• Scientific network connections
• Digital networking for defense and military applications (including Comms on the Quick Stop)

greatest transmission delay occurs, but these issues will be discussed in greater detail below. A significant number of FSS systems in the developing world continue to operate in the 6 and 4 GHz frequencies (known as the "C band"), while many of the networks that are providing services to enterprise networks and support large scale VSAT corporate networks today operate in the higher spectrum bands, namely the 14 and 12 GHz frequencies (i.e., the "Ku bands") and the 30 and 20 GHz frequencies (i.e., the "Ka bands").

Commercial FSS satellite systems can be used for virtually all types of telecommunications services that include among other things telephony, fax, data, terrestrial and submarine cable system restoration, interconnection of rural wireless telecommunications networks, television, audio, and IP-based data

networking. The satellite-based IP services can be used for both heavy route IP-based major system interconnection (or "trunking" services) and for services that are provided directly to individual subscribers and small businesses. On a global basis, FSS satellites are still employed for all these services and more. Despite this broad range of potential applications, satellites for communications are most efficient when they operate in the following modes and for the reasons provided in Table 4.3.

Overview of Video and Audio Broadcast Satellite Services

By far the predominant market in the field of satellite communications relates to video and audio services. A significant portion of the FSS market comes from satellite- television and audio distribution services. Some of the higher-powered FSS satellites actually transmit television and audio services directly to homes, offices, multi-dwelling units, and even to ships at sea. These FSS satellites sometimes also employed to send video programming directly to third and fourth generation broadband cell phones. The direct broadcast satellite market, or in ITU terminology, the BSS market, as indicated in Table 4.3, represents over 50% of the total satellite communications revenues worldwide. These types of direct broadcast services are (except for the exceptions noted above) are different from FSS services in that they typically connect to end-users and bill consumers directly, as opposed to the FSS services that are connecting business, governments, and other large scale users and usually are not offering services to general consumers on a "retail basis." The BSS services are thus offering retail services to millions of consumers and billing them directly. These BSS service providers are further up the "value added ladder" of commercial offerings than the FSS service providers. Rather than having millions of direct consumers as customers, the FSS providers have a much smaller pool of large commercial customers or governmental agencies that in turn relate to the actual public or end user.

Audio services are typically provided in tandem with television distribution and/or broadcast television services. In addition to the BSS television and related audio services, there are now a growing number of special digital audio broadcasting-satellite services, such as those associated with Worldspace, XM Radio and its Sirius subsidiary service. These direct radio or audio broadcast satellite services represent an additional and growing satellite communications market. In the case of this type of service, radio programming is often provided directly to vehicles and individual mobile consumer receivers. These audio broadcasting services are often offered along with associated security or road assistance programs such as the "Onstar" service provided via XM radio to the owners of General Motor vehicles that subscribe to the service. These systems are also often designed to combine the "radio" and "security related services" with a space navigation service as well. In the case of an accident, this allows emergency services to know exactly where to go. In the case of a criminal issue such as car theft, a stolen car can be disabled so it will no longer operate.

Table 4.3 Prime usages for fixed satellite services

The most efficient FSS satellite applications	
FSS application	Reason for satellite efficiency in this application
Television distribution over broad areas	Satellites can provide broad coverage over very large areas, including locations with difficult terrains or locations that are hard to reach via terrestrial media. An antenna for transmission or reception can be quickly added or removed at modest expense. Hundreds, thousands, or tens of thousands satellite receivers can be added to a television distribution system at virtually no additional cost to the satellite transmission system in space.
Communications services to rural and remote areas	Satellites, by covering large areas with a signal, do not require a concentration of users at a particular point to make a satellite transmission cost effective. In comparison, laying a fiber-optic or coaxial cable or building a microwave relay tower system to reach only a few users in isolated areas is often not an economic proposition. In some cases, satellites are used to interconnect rural wireless telecommunications networks.
Communications services to islands with small populations	Islands with only a few inhabitants are not economic for fiber-optic or coaxial submarine cable systems. Thus satellites that cover broad areas of the oceans are economic ways to serve islands and colonies with modest populations. Again, in some instances satellites are used to interconnect rural terrestrial wireless telecommunications networks.
Communications services to large scale networks over broad areas	Large scale networks for such applications as credit card verification for gasoline/petrol stations, grocery store or retail chain stores, or for networking together car dealerships, hotel/motel chains are cost effective satellite applications. Satellites work well for these applications because they can provide connectivity over broad areas and for many flexible locations at low cost and constant updatability of service locations
Multi-casting	This is quite similar to large scale networking, except that data messages can be selective sent and received at multiple addresses that vary within a large network. This is useful for functions like inventory control messaging, retail pricing by geographic area, etc. The advantages are thus the same as for large scale networking.
Short-term events or remote news gathering	Satellites do not require long-term installation of antennas. Temporary access to a videoconference, up-linking of television news or other short-term events do not require the laying of a terrestrial cable or other "permanent infrastructure" in the ground. A truck mounted satellite earth station, a fly-away terminal or other temporary antenna can be used for a brief period and then relocated to the next place where temporary access is needed.

Defense-Related "Dual Use" of Commercial Communications Satellite Systems

When commercial satellite services began, the first applications were essentially focused on governmental agencies, television, and radio broadcasters, and telecommunication companies seeking international and transoceanic services. Overtime these markets expanded to cover additional regional, domestic, rural, and island telecommunications services. Defense-related communications via satellite were quickly seen as major applications since satellites could provide broad coverage and transportable earth stations could be set up at remote locations. For similar reasons satellites were also seen a key way to respond to natural disasters and other emergency situations.

As early as 1965, at the same time as the first commercial satellite system, namely Intelsat, was deployed, the US military launched a low earth orbit constellation known as the "Initial Defense Satellite Communications System." It was designed to test the feasibility of having a dedicated satellite system for defense and security communications. The Soviet Union, in 1965, also deployed a system known as "Molniya." This three satellite system was designed to provide continuous coverage of the Northern latitude regions of the world. This was accomplished by launching three satellites in highly elliptical and highly inclined orbits (i.e., orbits that were highly inclined with respect to the Equator. Note: A satellite that travels in the plane of the Equator has a 0° inclination and a satellite that orbits around the Earth from the North to South Pole is essentially inclined 90° to the Equator.) This allowed each of the Molniya satellites to be visible above the horizon for at least 8 h a day. This Molniya system was used for a mixture of domestic governmental and civilian communications for the Soviet Union as well as for military and defense purposes.

The military forces of countries around the world quickly concluded that many non-tactical communications requirements related to defense operations, such entertainment for troops and personal communications for overseas personnel who wished to talk to families, etc., could be more easily provided and at lower cost by commercial satellite systems. Thus many defense-related communications of a non-tactical basis that required the use of satellite networks migrated to commercial systems. Such usage became known as "dual use" of commercial systems to provide civilian as well as defense-related communications.

Article XIV of the now defunct Intelsat Agreements that covered the coordination of the Intelsat system with other separate satellite facilities simply stated that: "This Agreement shall not apply to the establishment, acquisition or utilization of space segment facilities separate from the Intelsat space segment facilities for national security purposes" (Intelsat Organization 1973). These Agreements also explicitly indicated that Intelsat should not create "specialized satellite service facilities" for military purposes, but it was silent on what forms of satellite services might be carried on the Intelsat systems for military or national security purposes. In practice, Intelsat carried a range of "dual use" defense-related services but did not provide

"tactical military communications services involving the direct application of military forces in a combat environment." Other commercial satellite systems have tended to follow that same pattern.

Over time, however, commercial satellite systems have become more adept at providing "dual use" services since these types of applications represent a significant revenue stream. Particular ways that commercial systems have adapted to military-related types of services include:

- The ability to provide efficiently encrypted communications services
- The creation of special units within the commercial satellite operators that are especially designed to handle governmental or military communications satellite services
- The ability to accommodate specific military or defense-related requirements such as "communications on the move," "communications on the quick stop," and coverage of isolated or littoral areas
- Tailored lease arrangements so that communications satellite services can be "called up" and be available when required
- Arrangements for the quick launch of capacity when a particular part of the world requires additional satellite communications services

Despite these specific steps by commercial satellite system providers to accommodate military, defense, or emergency communications requirements, there have been new innovations in recent years to respond to specific requirements. These innovations fall into two new categories of commercial satellite services.

One is the case where specific defense-related satellite service requirements are procured from commercial contractors on the basis of a long-term lease, with any additional capacity be available for sale to regular commercial customers. The first such instance was the so-called Leasat, where capacity was made available to the US Navy and additional service was sold to maritime service customers. Today several European countries are obtaining defense-related services in this manner. The United Kingdom "Skynet" and other similar defense-related satellite service programs willl be discussed in ► Chap. 9, "An Examination of the Governmental Use of Military and Commercial Satellite Communications."

The other case is where commercial satellite consortia or companies design and build a communications satellite designed to operate exclusively in military or defense-related frequency bands on the basis of selling capacity – usually as a transponder lease – to various military organizations or units around the world. The first such project is known as X-TAR. This system is backed by the US-based Space Systems Loral Corporation and a consortium of investment companies in Spain, and the capacity from the two X-TAR satellites is being leased to US, European, and South American defense and emergency communications organizations. The specific characteristics of these types of defense-related satellite systems, which offer these services on the basis of transponder leases to a number of countries on a regional basis, are also described in further detail in a later chapter.

Evolution of New Digital Services and Applications

As described earlier, the evolution of communications satellite services over the past few decades has been marked by several key trends. One of the most important trends has been the continued subdivision of communication satellite markets into a growing number of service categories. Today there are fixed satellite services (FSS); land, maritime, and aeronautical mobile satellite services (MSS); broadcast satellite services (BSS), direct broadcast satellite (DBS) services, and Direct to Home (DTH) services. There are also store and forward satellite services and a wide range of military, defense-related, and emergency satellite services that range across all of the above service categories. Some satellite service providers even provide a hybrid type of service such as a combined store and forward messaging services plus a space navigational service to support shipping, trucking, bus, and railway customers. The key to the ability to provide all of these different types satellite communications services efficiently, reliably, and cost-effectively is the new digital services that have largely transplanted analog services around the world.

The use of digital modulation, digital encoding, and digital multiple access techniques, and especially Time Division Multiple Access (TDMA) and Code Division Multiple Access (CDMA) coupled with new highly efficient coding concept such as "Turbo-coding," has allowed satellite systems to grow and become more reliable and cost effective. The greatest challenge to the growth of satellite services in the digital age of the Internet has been adapting to the widespread application of the Internet Protocol (IP) across the planet. The transmission delay associated with connecting between the Earth and geosynchronous orbit created a special problem for networks operating on the Internet Protocol. This is because the original design of IP networks was based on terrestrial networks and perceived delays were registered as "network congestion" rather than transmission delay. This led to a "slow recovery" process that undermined the efficiency of satellite transmission. Even satellites in medium earth orbit (MEO) had this sort of problem. The development of IP over Satellite (IPoS) standards has led to new efficiencies in satellite transmission. Another problem for IP-based satellite services have involved the creation of virtual private networks (VPNs) within business enterprise networks. The main problem in this regard that IP Security (IP Sec) strips off "headers" as a VPN is created. This also requires an innovative solution for VPNs delivered via satellite. The optimization of satellite transmission to operate efficiently within IP networks and to accommodate corporate VPNs is described in section "Satellite Communications Services as Defined by the ITU."

Limits to the Growth of Satellite Networks

The development of satellite communications services and applications over many decades since the start of such services in 1965 has been impressive. The most

important stimuli to growth have been diversification. The field of satellite communications has expanded by constantly finding an ever-wider circle of applications. This began with just fixed satellite services, but services quickly diversified into such areas as video and audio broadcasting, maritime, aeronautical and land mobile satellite services, and even store and forward and supervisory control and data acquisition (SCADA) applications. Many of these applications today are "hidden" from the consumer since the satellite network is remotely orbiting thousands of kilometers out in space. When a consumer makes a purchase at a gas station, a grocery store, or retail outlet, a credit card is most likely validated via satellite connection. Most television signals, even those received via cable television, most likely originated or traveled by one or more satellites at some point in their transmission.

The second most important source of growth has been the prodigious expansion of digital satellite communications services and the adaptation and optimization of satellite for IP-based transmission. Many countries connect to the worldwide Internet via backbone or trunk satellite transmissions. Increasingly satellites can support high-speed broadband Internet connections to the small office/home office (SOHO), especially where fiber, cable or terrestrial broadband wireless is not available.

Despite the growth of satellite communications services, there are limits to this expansion as terrestrial networks (fiber, coaxial cable, or broadband terrestrial wireless systems) are installed across the planet. This suggests two paths forward. One path is the increased integration of satellites with terrestrial systems. The other is the off-world use of space communications for interconnection with scientific satellites and in time even colonies in space.

The integration of satellite and terrestrial systems is seen in various ways. Fiber-optic systems and satellite networks have often been planned in tandem and satellites have been used as backup to terrestrial cable systems in case of emergency outages or natural disasters. In large and geographically diverse countries, satellites have frequently been used to cover large and thinly populated areas such as mountain ranges, deserts, and forested areas or wetlands in conjunction with terrestrial cable systems. In recent years, the new trend is to use satellites in conjunction with terrestrial broadband wireless systems. There are now combined satellite and land mobile systems that provide fully integrated wireless broadband Fourth Generation (4G) service in the United States and North America. These systems and their current financial status were described above. The official generic name for this service, as defined by the US Federal Communications Commission (FCC), is land mobile satellite service with "ancillary terrestrial component" (ATC). In Europe, such type of integrated satellite and terrestrial wireless services is known as mobile satellite service with "complementary ground component" (CGC).

Another similar effort to provide integrated satellite and terrestrial broadband services for the developing world is system being planned under the name O3b. This unconventional name stands for the "Other Three Billion" people who largely live in the equatorial regions of the planet and have limited access to potable water, health care, educational services, electricity, and communications. The origin of this idea is that the density of broadband Internet traffic in continents such as Africa

and large parts of South America and Asia still do not yet justify the installation of broad band terrestrial fiber or coaxial cable systems. Instead, however, a broadband wireless terrestrial systems linked to an IP-optimized satellite system would be a way to provide service to the underserved areas of these continents. Indeed the logic that initially suggested that this might be a viable solution for Africa, led to the further idea of a global satellite system optimized for the developing countries largely concentrated in the equatorial band of the planet. This new type of satellite system could be deployed to meet the traffic requirements of the three billion plus people in developing countries that would like to have economical and reliable access to broadband terrestrial wireless.

Although the O3b system is intended to be optimized for this service, other satellite systems such as Intelsat, SES Global, Asiasat, etc. can also support this type of rural connectivity architecture. Consumers with broadband wireless Internet access could then connect via satellite to national and even global service – especially if the service was truly low cost and suited to the affordable pricing realities of developing economies. The O3b system is currently in planning and capital financing is still being put in place. The economic difficulties that have been experienced by satellite systems geared to significant dependence on developing country markets, such as the "Worldspace" audio broadcasting system, the initial Iridium and Globalstar systems however does suggest that such types of satellite systems will face significant economic challenges. Nevertheless, the concept of combining satellite wide area coverage with terrestrial wireless networks in town and city areas has a great deal of logic behind it, in terms of serving regions with a lower density of broadband IP-based traffic.

The future of satellite services and applications does seem to be constrained by major parameters and yet stimulated by specific opportunities suggested by satellite technology with its broad areas of coverage.

Satellites thus seem to have particular service opportunities for broadcasting and multi-casting services (both audio and video coverage) for a good time to come. This is the largest source of revenues to commercial satellite services today with well over 70% of revenues coming from this source. Satellites, for similar reasons of broad coverage at economic rates, are well suited for large scale corporate networks, often referred to as enterprise networks. Since these networks tend to be very dynamic with nodes often being added or subtracted all over a country, a region or indeed the world, satellite links remain quite well suited to such services. Satellites likewise remain very well suited to mobile communications or communications on the move, especially where the density of traffic is low or traffic needs suddenly appear such as in the case of an emergency, natural disaster, or area of armed conflict.

The reverse condition works in favor of terrestrial fiber-optic or coaxial cable. In areas where there is a high density of users within a developed economy, the installation of such terrestrial infrastructure tends to occur. Also fiber-optic submarine cables are also installed between high-density international telecommunications traffic routes. In these cases, satellite services tend to provide backup capability in case of cable breaks or natural disasters. The new land mobile

satellite systems (with ATC or CGC) represent an excellent example of how satellite communications design engineers have adapted to these broad trends related to the strengths and weaknesses of satellites and terrestrial telecommunications networks to design systems that are optimized to both forms of technologies.

Finally, the other major trend in commercial communication satellite services known as "technology inversion" is also expected to continue. This trend is the consumer mandate to develop, for the individual user, ever more compact ground communications units that require less operating power, are lower in cost, and easier to operate. As the number of users of satellites has expanded from thousands to millions to billions, the economics of satellite communications have changed dramatically. The volume of users has allowed large investments in powerful and sophisticated satellites that allow user transceivers to shrink from huge 30 m earth stations with large operating crews to quite small hand held units. Although there might be some small satellites for experimental purposes, to support ham radio relay, or message relay, the main commercial trends will continue to support mass consumer needs and very small and low cost user terminals and transceivers.

The truly longer-term future of satellite telecommunications services will relate to the need for truly long distance communications through space to the Moon and lunar colonies and eventually even to other planets and beyond. Today the various space agencies, particularly NASA (USA), ESA (Europe), JAXA (Japan), CNSA (China), Roscosmos (Russia), CSA (Canada), and ISRO (India) have had need to create space communications systems to support exploratory missions. NASA has developed and expanded a very sophisticated Deep Space Network (DSN) for decades to receive signals from scientific missions to the Moon and beyond. The important element to note is that technology developed for the demanding requirements of sending and receiving signals over the vast distances of space have often led to technical or service innovations that can be used to improve commercial communications satellite systems here on Earth. The three axis body stabilized satellites that are now in common use around the world, that provide higher levels of pointing accuracy and better solar illumination of solar cell arrays, were first developed at the US Jet Propulsion Labs in terms of improved communications requirements associated with planetary missions. In coming years, commercial satellite communications systems may indeed evolve to provide interplanetary and cislunar communications services.

Conclusion

The commercial satellite communications market has proven to be very dynamic and new services and applications have diversified and grown as the field has matured over the decades. Digital communications technology and Internet Protocol based services have helped to accelerate this diversification. The dynamic interrelation of satellite technology to terrestrial telecommunications overtime has been another key factor that has influenced both the size of commercial satellite markets and the shape of the market in terms of services offered. Coaxial cable

systems, fiber-optic networks, and most recently terrestrial broadband wireless systems have both limited the growth of commercial satellite communications markets and services, and also defined new opportunities for satellite systems to complement terrestrial telecommunications systems.

Satellites continue to have particular market opportunities in terms of broadcasting and multi-casting services, offerings to rural and remote areas, to island countries and developing countries – particularly when the terrain or topology of these countries create significant barriers to deployment of conventional terrestrial telecommunications systems. In short, countries with a large number of islands (i.e., Indonesia, the Philippines, or Micronesia), countries with significant mountainous terrains (i.e., Chile or Nepal), countries with extensive jungles (Brazil, the Republic of Congo, Malaysia, and Thailand), or countries with major deserts (i.e., Algeria, Libya, or Mauritania) will find satellite technology and services well adapted to their needs. The ability of new more powerful satellites to operate to smaller and smaller user terminals of lower cost and greater mobility will also continue to stimulate the growth and diversification of commercial communications satellite markets. Finally, the next frontier for satellite telecommunications services in the decades ahead will relate to communications across the solar system and beyond as human exploration scientific studies and practical space applications extend further and further beyond Planet Earth. For further readings related to materials covered in this section please consider the materials covered in the endnotes.[11]

Cross-References

- ▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ▶ Economics and Financing of Communications Satellites
- ▶ Fixed Satellite Communications: Market Dynamics and Trends
- ▶ History of Satellite Communications
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends
- ▶ Satellite Communications Video Markets: Dynamics and Trends
- ▶ Satellite Orbits for Communications Satellites

Notes

1. The International Telecommunication Union www.itu.int/net/about/index/aspx
2. World Radio Conference, of the International Telecommunication Union, Final Acts of the 2007 Conference www.itu.int/ITU-R/go/index.asp?wrc-07/en
3. ITU Frequency Allocation Table, www.itu.int/ITU-D/tech/spectrum-management./SMS4DC_AM_TM_4.pdf
4. ITU Radio Regulations, Article 1, Definition of Radio Service, Section 1.21, http://www.ictregulationtoolkit.org/en"practiceNote.aspx?id=2824

5. ITU Radio Regulations, Article 1, Definition of Radio Service, Section 1.25, http://www.ictregulationtoolkit.org/en"practiceNote.aspx?id=2824
6. A global overview of Direct to Home Television services www.electronics.ca/publications/products/Direct%252dto%252dHome-(DTH)-Satellite-Television-Services-%252d-A-Global-Strategic-Business-Report.htm
7. The Boeing Company Manufactures and Deploys the Spaceway Satellites www.boeing.com/defense-space/spaceway/spaceway.html "Direct TV's Spaceway F1 Satellite Launches New Era in High Definition Programming" www.comspacewatch.com/news/viewpr.html?pid=16748
8. ITU Radio Regulations, Article 1, Definition of Radio Service, Section 1. XX, http://www.ictregulationtoolkit.org/en"practiceNote.aspx?id=2824
9. For more information on micro satellites and nano satellites see www.satellite-links.co.uk/links/satman.html
10. The International Amateur Radio Users Guide to Frequency Coordination provides a full listing of approved frequency bands under the International Telecommunication Union Radio Regulations for all three ITU Regions. Provision 5.282 of the ITU Radio Regulations specifies the following: "In the bands 435–438 MHz, 1,260–1,270 MHz, 2,400–2,450 MHz, 3,400–3,410 MHz (in Regions 2 and 3 only) and 5,650–5,670 MHz, the amateur-satellite service may operate subject to not causing harmful interference to other services operating in accordance with the Table (See 5.43) [i.e., The ITU Allocations Table] Administrations [of the ITU] authorizing such use shall ensure that any harmful interference caused by the emissions of a Station in the amateur-satellite service is immediately eliminated in accord with the provisions of the "Spectrum Requirements for the Amateur and Amateur-Satellite Service, International Amateur Radio Union, August 2008" http://www.iaru.org/ac-08spec.pdf
11. For further reading concerning satellite telecommunication services See J.N. Pelton, A. Bukley (eds.), *The Farthest Shore: A 21st Century Guide to Space* (Apogee Books, Burlington, Canada, 2010). Particularly Chapter Six on satellite applications. D.K. Sachdev, *Business Strategies for Satellite Systems* (Artech House, Boston, 2004). R. Jakhu, *National Regulation of Space Activities*. Space Regulation Library Series (Springer, Dordrecht, 2010). And J.N. Pelton, *The Basics of Satellite Telecommunications* (Professional Education International, Chicago, IL, 2006).

References

Intelsat Organization, *Article XIV of The Agreement of the Intelsat Organization, as it entered into force as of Feb. 12* (Intelsat, Washington, DC, 1973)

Orbcomm, Orbcomm Signs Next Generation Satellite Constellation Contract for 18 Satellites (Orbcomm, Fort Lee, NJ, 2008)

J.N. Pelton, *The Satellite Revolution: The Shift to Direct Consumer Access and Mass Markets* (International Engineering Consortium, Chicago IL, 1998).

Satellite Orbits for Communications Satellites

5

Joseph N. Pelton

Contents

Introduction 94
Different Orbital Configurations for Different Communications Satellite Services 96
Geosynchronous or Geostationary Satellite Orbits 97
Medium Earth Orbit (MEO) 98
Low Earth Orbit 100
Various Types of Communications Satellite Constellations 100
Molniya, Highly Elliptical Orbits (HEOs), Extremely Elliptical Orbits (EEOs), and Loopus Orbits 104
String of Pearls Orbit 104
Quasi-Zenith or Figure-8 Orbit 106
Super Synchronous 107
Earth Station and User Terminal Design for Different Orbits 108
Relative Economics of Different Satellite Orbits 110
Cross-References 113
Notes 113
References 113

Abstract

One of the key elements of a communications satellite service is the ability to launch satellites into precisely defined orbits and to maintain them in the desired orbit throughout the lifetime of the satellite. This systems control and oversight of satellite orbits involves not only the technical ability to launch and maintain the orbit, but also the ability to attain the proper legal authority at the national and international level to transmit and/or receive radio signal from these orbits. This regulatory process means a number of specific steps associated with registering for the allocated frequencies from the International

J.N. Pelton
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: joepelton@verizon.net

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_5, © Springer Science+Business Media New York 2013

Telecommunication Union through a national governmental administration, obtaining assignments of those frequencies in the required orbits in accord with national licensing procedures, and coordination of the use of the specific frequencies through intersystem coordination procedures.

There are a wide range of different orbits that are currently used in communication satellite services although the most common are geosynchronous Earth orbits (GEO), medium Earth orbits (MEO), and low Earth orbits (LEO). This chapter explains the various orbits that can be used and the advantages and disadvantages of each of the orbits most often employed for satellite communications. This analysis indicates some of the primary "trade-offs" that are used by satellite system engineers in seeking to optimize a satellite systems performance both in its design and subsequently over its operational lifetime. The activities involved in selecting an orbit, designing and achieving an operational satellite network, and optimizing its technical, operational, and financial performance over the systems lifetime involve a wide range of issues. These start with selecting a desired orbital framework, obtaining authorization for orbital access (including the registering and pre-coordination of the satellite and its orbit with other systems), launch, deployment and test, systems operation, and end-of-life disposal of a satellite from its orbit.

Keywords

Antenna gain • Antenna pointing • Command and control of satellites • Figure-8 orbit • Geostationary earth orbit • Geosynchronous satellite orbit • Inclined orbit • Loopus orbit • Low earth orbit • Medium earth orbit • Molniya orbit • Omni antennas • Polar orbit • Quasi-Zenith orbit • Radio astronomy • Satellite constellations • Space weather • String of pearls orbit • Sun spot activity • Sun-synchronous orbit • Super synchronous satellite orbit • Tracking of satellites • Van Allen belts

Introduction

Sir Isaac Newton first discovered the laws of gravitational attraction and created understanding of the planets revolve around the Sun and why satellites revolve around the planets. Newton even recognized that it would be possible to "shoot" artificial satellites into orbit around the Earth or other planetary bodies. His universal law of gravitational attraction is still fundamental to understanding basic orbital mechanics. This law is expressed as follows:

$$F_g = GMm/r^2$$

The above formula expresses Newton's Universal Law of Gravitation and shows how to calculate the force of attraction between the Earth and another object. This mutual attraction (although the Earth's pull is obviously very much greater) is determined by taking the mass of the Earth (M) and then knowing the distance

between the center of the Earth's mass and the center of gravity of the orbiting mass. This shows that the gravitational pull decreases as the object moves further from Earth. Indeed this mutual attraction decreased by the square of the distance it is away from the Earth's center. In order for an object to be in Earth orbit, it must have sufficient velocity or centripetal force to overcome the gravitational pull (Pelton 2006).

In the early experimental days of satellite communications there was a wide range of opinion about what types of orbits might be used most effectively for telecommunications services. There were a number of perceived advantages and disadvantages envisioned for different orbital configurations. Once the experimental Syncom 2 and 3 satellites established the viability of launching and operating a satellite in geosynchronous (GEO) orbit, however, the worldwide practice – in a short period of time – concentrated on this orbit. This was because of several factors such as the simplicity of not having the Earth station systems to track the satellite and the high gains of those types of antennas over those with low-gain omni antennas that could receive signals across the full open horizon. Also the fact that satellites in GEO could cover over one-third of the World's surface was a strong economic factor (Pelton 1974).

Over time, other applications rekindled interest in other orbits. These factors included interest in mobile satellite systems, defense satellite systems with the need for communications on the move, store and forward applications, and the desire to achieve low latency (or less delay in satellite transmission time). These elements and more served to revive interest in other times of orbits.

This is not to say that GEO presented the only orbital option. Special conditions such as the northern latitudes of the Russian landscape allowed a special highly elliptical orbit, known as the Molniya orbit, to be utilized (Pelton 1974, p. 55). The limitation of available radio frequency allocations and the crowding of the geosynchronous orbit that over time served to move the "spacing" of communications satellites in GEO orbit closer and close together affected the technology of the satellites and Earth stations as well as the active consideration of different orbits. There is, however, a technical coordination difficulty when satellite systems that use the same allocation of frequencies for similar services attempt to use disparate orbital constellations.

For instance, when constellations of satellites in low and medium Earth orbits and using the FSS frequencies cross over the equatorial plane, they can cause substantial interference to the geosynchronous satellites that utilize the same radio spectrum bands. Further LEO and MEO satellites with high-powered beams can also cause significant interference with radio astronomy – particularly when they are using frequencies for mobile satellite services (MSS).

The following discussion presents the various types of orbits and their uses and applications including the advantages and disadvantages that are involved from a technical, operational, financial, and regulatory perspective. This is followed by consideration of particular issues that are involved with Earth station design, technical coordination between different types of orbital configurations, as well as terrestrial and other services (such as radio telescope services, terrestrial microwave, and high

altitude platform systems (HAPS), and finally disposal of satellites at end of life). Most of the issues involving technical coordination and registration of radio frequencies for satellites are addressed in the chapter that explains ITU allocation procedures, but the issues particular to different orbital configurations are addressed here.

Key technical concerns that accompany the selection of satellite orbits include the extent of their geographic Earth coverage and the so-called path loss that is determined by how far the satellites are away from the Earth's surface. Another major concern, however, is the problem of destructive radiation that can disable or even completely end the useful life of communications satellites. The Van Allen belts contain intense radiation which actually helps to shield the Earth from radiation from the Sun or the stars but these "structured belts" around the Earth can be destructive to satellites that must fly through them. In addition, radiation from the Sun, so-called space weather, is also a concern to maintaining satellites effectively in orbit. Indeed during intense "sunspot activity" or solar storms, satellites are typically shut down to prevent failure to satellite electronics (Charles et al. 2009). Further, satellites in orbit are also subject to cosmic radiation and are especially vulnerable to the solar wind and to the most intense radiation from the Sun that comes with occasional solar storms that follow a multiyear cycle that peaks during what is called the Solar Max period.

Different Orbital Configurations for Different Communications Satellite Services

There are a large number of orbits that can be used for satellite communications and an infinite number of constellations that can be created using a multiple satellites. The most common orbits are geosynchronous or geostationary orbit (often called Clarke orbits in honor of Arthur C. Clarke), medium Earth orbits, and low Earth orbits.

Deployment of satellites into these various orbits require progressively more energy as they are positioned further away from Earth since they are continually overcoming the Earth's gravitational pull as they ascend to higher orbits. Also highly specialized orbits, such as a GEO orbit, that require the satellite to be placed into a circular plane above the equator and at great height require greater orientation and positioning capability as well as the need for nearly constant station-keeping maneuvering. In the simplest terms, low Earth orbits that involve "direct insertion" are the easiest to attain. Polar orbits and medium Earth orbits are the next easiest to achieve. GEO orbits are the most difficult. At the outset of the satellite industry, satellites were deployed into very highly elliptical orbits with the apogee at the desired height of 35,870 km (22,230 miles). After this orbit was firmly established then something called an "apogee kick motor" (AKM) was fired at the appropriate apogee in order to push the satellite from transfer orbit into a perfectly circular orbit around the equator at the desired longitude.[1]

Today's rocket systems with advanced propulsion systems can insert the satellite into the final GEO orbit without the hazard of firing a solid motor rocket. In the future, ion engines might achieve GEO orbit in entirely different ways by

continuously firing "electronic thrusters" that slowly spirals the satellites outward from Earth to achieve the desired circular GEO orbit after weeks of firing these much lower energy thrusters. The subject of rocket launchers and orbital deployment is discussed later in this handbook.

The following discussion describes the characteristics of the various orbits utilized and their various strengths and weaknesses for various types of communications services.

Geosynchronous or Geostationary Satellite Orbits

The most common orbit for satellites providing fixed satellite services and broadcast satellite services and quite a few mobile satellite services are those that are called the geosynchronous, geostationary, or Clarke satellite orbit. This is a unique orbit where the orbital velocity is sufficient to maintain the satellite in this circular equatorial path with the centripetal force away from Earth exactly overcomes the pull of gravity at this altitude. The "g" force or gravitational pull at this orbit is approximately one-fiftieth (1/50th) that experienced at the Earth surface. This is to say that at a distance of 22,230 miles or 35,870 km the accelerative pull of the Earth's gravity is (0.22 m/s^2 rather than 9.8 m/s^2). What makes this orbit so special is that the orbital velocity that creates the angular momentum (and thus the centripetal force) needed to overcome the pull of gravity just happens to constitute the exact speed needed to complete a revolution around the world exactly every 23 h and 56 min and 4 s. The "odd missing 4 min" of a 24 h day represents rather exactly the 1/365th of the time the Earth uses to revolve around the Sun. In short, in celestial (or sidereal) time, a spacecraft in GEO Earth orbit revolves exactly once around the world every day. It thus appears as if it were indeed a very, very tall tower in the sky with the satellite at the top of the imaginary tower (Pelton and Madry 2009).

This special orbit identified in 1928 by Herman Potočnik (who wrote under the German pseudonym Hermann Nordung) as a location for an inhabited space station, and more famously identified by Sir Arthur Clarke as a location a "geostationary communications satellite" has been in continuous use by artificial spacecraft since 1965.[2] There is always a difference between theory and practice. The geostationary orbit of theory would keep a satellite moving west to east exactly with the rotation of the Earth as it remains stable exactly about the Equator. There are gravitational affects of the Moon and Sun that tend to tug a satellite in this orbit to move North above the equatorial plane and the tug it back South of the equatorial plane. It turns out that the "station keeping" required to keep a spacecraft exactly in geostationary orbit in terms of North and South migrations demands approximately ten times more energy (i.e., firing of station-keeping jets) than maintaining the East–West stabilization. In short a spacecraft to be maintained within its assigned "GEO orbit box" – in terms of its desired longitude (East–West location) and its desired "0" degree latitude requires active station keeping. The excursions North and South both above and below the equatorial plane look like a very low amplitude sine

wave. These are the most difficult to control and require by far the most fuel use in the firing of thrusters.

The buildup of the excursions to the North and South of the equator are called the inclination of the satellite orbit. Most "GEO satellites" tend to move perhaps a degree or so North and South of the equator each day. As long as the spacecraft does not move more than plus or minus a very few degrees above the equatorial plane, this does not create more than a very minor problem for an Earth station on the ground pointing to the satellite receiving the signal. The slight variation in the gain for the Earth stations on the ground are most pronounced in the equatorial regions, but since the locations at the so-called subsatellite point receive the strongest signal (i.e., the least path loss), this is not a particular problem. In practice, therefore, satellites in "GEO" orbit are thus more or less "geosynchronous" but not really "geostationary" because of these small excursions off the equatorial plane. Toward the end of life, satellite operators tend to let "GEO" satellites build up their inclination (i.e., movement North and South of the equator) because this saves station-keeping fuel and allows the extension of the satellite's practical lifetime (Williamson 1990; Pelton 2006, p. 76). The ITU that maintains the global registry of satellites and their location recognizes the registration up to 5° inclination above and below the equatorial plane. (Note this is of significance in terms of problems of interference between GEO satellites and those in low and medium Earth orbits when operating in the same frequency bands.)

A final note on orbits is that concerning eclipses of satellite in different orbits. Satellites in polar or near polar orbit in low Earth sun-synchronous orbits will in a 90 min orbit be behind the Earth and shielded for the Sun for about 35 min for each revolution. Batteries must provide power for this part of the orbit. In the case of geosynchronous satellites, the issue of eclipses represents a more complicated issue. The Earth is "tilted" 23.5° on its axis and as such during the solstice times a GEO satellite and its solar cell arrays are fully illuminated. GEO satellites either "see the sun" from over the North Pole or under the South Pole during the winter and summer months. Some 22 days before the equinox period, however, GEO satellites will experience a small eclipse that builds up to a maximum of some 70 min a day during the spring and fall equinoxes. These eclipses then dissipate in the same manner until 22 days after the equinox. The issue of an eclipse from the shadow of the Earth is again addressed by satellite operators by the use of batteries during these two periods of semiannual darkness. Since GEO satellites have nearly 300 days a year of total illumination, the issue of eclipse is much greater for LEO systems where a satellite can be in eclipse over a third of the time. Satellites in LEO constellations often use the time when they are over polar regions or over oceans where traffic demand is low for time to recharge batteries (Table 5.1).

Medium Earth Orbit (MEO)

A medium Earth orbit (MEO) satellite constellation can be configured in many different ways to achieve global coverage. The main constraint that impacts the

Table 5.1 Advantages and disadvantages of the GEO orbit

Advantages and disadvantages of GEO satellites for communication services (Pelton 2001)	
Advantages	Disadvantages
Three satellites in GEO orbit provide essentially global coverage except for the polar regions. This means that global coverage can be achieved at a lesser cost than MEO satellites (10–18 satellites for global coverage) or LEO satellites (48–60 satellites for global coverage). This is because the closer a satellite is to the Earth the less the satellite is able to "see" of the Earth below. Even one GEO launch can create a full-service capability for a region, while with MEO and LEO satellites a full constellation must be in place to create a fully functional system.	The satellites in this orbit are almost one-tenth of the way to the Moon and thus there is a very large path loss between the satellite and ground antennas. Since path loss (i.e., diminished signal) is a function of the square of the distance, the satellite is away from the Earth. This is a substantial factor in satellite design and the ability to "close a link" between a GEO satellite and the Earth.
A satellite in GEO orbit allows continuous connection with high gain Earth stations without constant tracking of the satellites. This allows for a simpler and less expensive antenna design. Or it requires the ground antennas for LEO or MEO satellites to be much lower gain devices that are essentially "omni-devices" (i.e., ones that can receive a signal from all different directions).	The great distance the spacecraft orbits away from Earth creates delay or latency in the transmission. This latency is on the order of a quarter of a second for the entire pathway from the Earth to the Satellite and the return. This creates problems for telephone communications and in Internet connections. At a low elevation angle from the antenna to the spacecraft coupled with a low elevation angle for the return transmission, the path can be over 75,000 miles or 120,000 km.
A GEO satellite with large high gain antennas can create spot beams continuously pointed to desired geographic locations on the Earth.	Inter Satellite Links (ISLs) between GEO satellites are much harder to establish and require much higher capability, power, etc. than is the case with LEO or MEO satellites.
A GEO satellite is relatively easy to maintain in orbit and can sustain in-orbit lifetimes of 15–20 years which is longer than medium Earth orbit (MEO) and especially longer than low Earth orbit (LEO) satellites.	To use a GEO satellite for mobile satellite service requires a very high power and huge aperture multi-beam satellite to connect with a simple hand-held user terminal with reasonable reliability. (This is largely a function of significant path loss).
A GEO satellite can easily have its orbit raised out of GEO orbit at the end of life. This is much easier to accomplish than spending a great deal of fuel (40% of all fuel) to de-orbit a MEO or to de-orbit a LEO satellite that creates various types of risks to other satellites.	Each satellite tends to be larger, more complex, has longer production schedules, is more expensive to launch and insure, and allows less economies of scale than MEO or LEO satellites. This is usually more than offset by the economies achieved by the need to launch many fewer of this type satellite to complete a full system.

planning of a MEO system is to launch the system so that the satellites are essentially flying above the Van Allen belts. As noted earlier the radiation in the Van Allen belts contains very high-speed particles such as high-energy neutrons. This radiation can do damage to satellite electronics and even with spacecraft

shielding of the electronics and glass coating on solar cells, the lifetime of the spacecraft will be significantly shortened if it must fly within the Van Allen belts.

Although a number of communication satellite systems have been proposed that would utilize global MEO constellations, such as the original ICO and the Odyssey systems, these were never deployed. There have also been concepts for using MEO systems for high capacity broadband systems for the Ka-band, but such a system for a variety of reasons was never deployed. Currently the system known as O3b (for the "other three billion" people) contemplates using a MEO system to support a high-speed Internet service to broadband wireless users in developing countries with a focus on the equatorial countries of the world.

It is possible to operate just one or two satellites in MEO orbit for store and forward services or machine-to-machine (M2M) connectivity.

In many ways, a MEO constellation provides a compromised between the advantages of a GEO system on one hand and a LEO constellation on the other (Table 5.2).

Low Earth Orbit

The orbit of choice for a number of land mobile satellite systems in the past decade or so has been the LEO Constellation. The Iridium and Globalstar systems have deployed and operate LEO Constellations for mobile communications services for nearly 15 years. Further, there have been a significant number of store and forward satellite networks that have included the commercial Orbcomm system as well as many Surrey Space Center class and Utah State University and Oscar (Amateur Radio) small satellites that have been launched going back many years. LEO constellations represent the opposite extreme from GEO satellites with their strengths being GEO satellites weakness and *vice versa*.

Indeed the advantage of being close to the Earth and thus allowing transmissions to experience less path loss is also a disadvantage because many more satellites are needed to achieve global coverage. Figure 5.1 below shows, in cartoon fashion, the geometrical profiles of GEO, MEO, and LEO orbits with respect to the Earth. This figure is not to scale since a GEO orbit can, in fact, be some 40 times further out in space than a LEO orbit and this cannot be easily shown to exact dimensions.

The following table presents the relative pros and cons of a LEO satellite constellation (Table 5.3).

Various Types of Communications Satellite Constellations

There are literally an infinite number of constellation designs that can be devised for low and medium Earth orbit satellite systems. In order to design the optimized constellation, there are a number of key threshold questions that a satellite system operator will typically address. Selection of a particular parameter to be optimized in a satellite system orbital configuration will likely dictate the number of satellites

Table 5.2 Advantages and disadvantages of the MEO orbit

Advantages and disadvantages of MEO satellites for communications services (Pelton 2001, pp. 228–230)	
Advantages	Disadvantages
A MEO constellation provides global coverage and with significantly less path loss and transmission delay than a GEO system.	More satellites than a GEO system must be purchased and launched to create a global network.
A MEO constellation can achieve global coverage with the launch of as few as 10 satellites and the tracking, telemetry, and command (TT&C) system needed to support the system is much less than a LEO constellation.	Full system must be deployed and completely checked out to operate network, unlike a GEO satellite that can operate as a complete network by itself with wide regional coverage.
Inter Satellite Links (ISLs) and mission control are much easier to accomplish than with a LEO system, but more difficult than with a GEO system.	There have been very few MEO constellations deployed for communications services and thus there is limited experience with the operation of these systems; optimizing the construction of satellites for MEO operation; or knowledge about special design aspects such as radiation shielding, etc.
These orbits provide a very good trade-off between total number of satellites, complexity of system control and TT&C requirements, requirement for satellite on-orbit "spares," wide area geographic coverage, power of spot beams and geographic coverage, and reasonable path loss and transmission delay.	De-orbiting a MEO satellite requires a great amount of thruster fuel and this recreates a cost disadvantage and adversely affects the lifetime of the satellites and of the overall constellation.

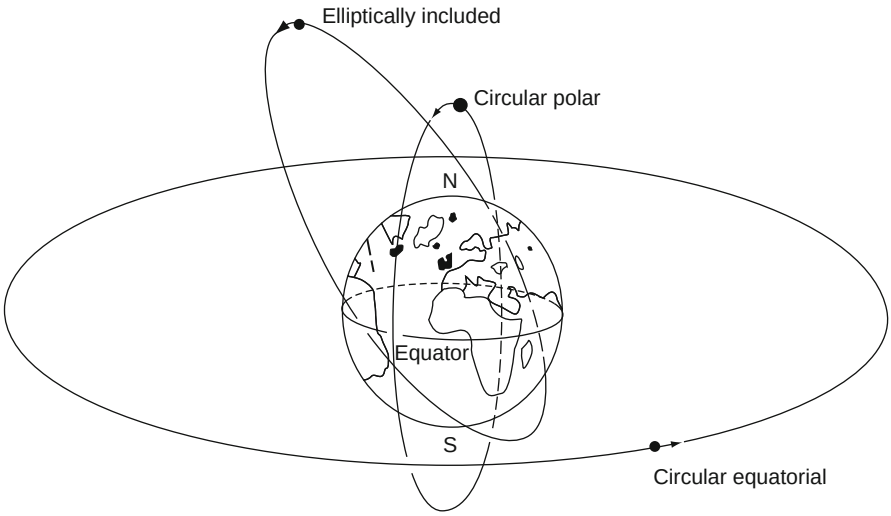


Fig. 5.1 A "cartoon" depiction of a LEO, MEO, and GEO orbit (Graphic courtesy of the Author)

Table 5.3 Advantages and disadvantages of the LEO orbit

Advantages and disadvantages of LEO satellites for communications services (Pelton 2001, pp. 228–230)	
Advantages	Disadvantages
Low Earth orbit (LEO) satellites are up to 40 times closer to the Earth's surface and thus experience up to 1,600 times less transmission path loss.	There is a need for a large number of satellites to complete a global constellation network (i.e., 50 satellites and up). This increases systems costs for a global system because there are many more operational and spare satellites, many more launches, and a more complex TT&C network for system control.
Low Earth orbit satellites experience up to 40 times less latency or transmission delay than GEO satellites. This is simply because a LEO satellite orbit is 20–40 times closer to the Earth's surface than a GEO satellite.	The system requires more difficult overall system controls, complex billing and authentication systems, and network implementation, including more active spares and system restoration procedures. This can in part be overcome by installing Inter Satellite Links (ISLs) on all satellites, but this also increases costs and satellite complexity.
LEO satellites, because they fly more directly overhead and cover the lower and higher latitudes more effectively than GEO satellites, typically will have lower "masking angles" to user receivers and particularly provide more effective coverage at upper latitudes and can even provide service to the polar regions.	One cannot use high gain ground antennas constantly pointed toward the satellites because the spacecraft is rapidly moving across the sky with only a few minutes of visibility before moving below the horizon and thus needing to be replaced by another satellite in the constellation.
LEO constellations are particularly well suited to mobile satellite services because of the lower path loss, lower masking angles, concentrated beam coverage, modest transmission delay, and the desire to provide users with lightweight, compact, and low-cost antennas with small, relatively low-gain antennas.	The satellites are being attracted much more strongly by the Earth's gravitational field and there is more fuel needed for station keeping and thus the lifetime of the LEO networks are less. The operational lifetime of LEO satellites are typically about 7 years. GEO satellite lifetimes can be 12–18 years.
Orbital designs for LEO constellations can be adjusted to concentrate coverage at lower latitudes (from 0° to 70° North and South). (Unfortunately, necessary coverage of all longitudes provides coverage of the Atlantic, Pacific, and Indian Oceans where there are limited customers.)	There is a much higher probability of LEO satellites being hit and partially or completely disabled by space junk because satellites and space debris are more closely spaced.
Detailed computer programs can be designed to make LEO systems "smarter" and "dynamically flexible." This means that satellites can be programmed to increase or decrease power or performance in specific beams in specific locations. Increased power can be used for ringtones or message-waiting signals.	LEO constellations, because they often cover the entire Earth's surface with cellular-like beams and most often are utilized for mobile satellite services (and thus associated frequency bands), tend to have a more significant problem of coordination with radio astronomy services. Also user transceivers can interfere with one another.

deployed, determine the maximum and minimum elevation angles for the overhead satellites, and indicate the feasibility of intersatellite links.

The Iridium land mobile satellite system constellation design was heavily dependent on the concept that this global network would provide intersatellite links or cross-links among and between all four of the closest satellites in the network. This led to the decision to have the satellites in polar orbits (nearly 90° inclination to the Equator) so that the system would be highly symmetrical and cross-links easily established among the two closest satellites North and South and the two closest satellites East and West. Other elements in the trade-offs in the constellation design were satellite power versus number of satellites in the constellation and orbital elevation versus typical elevation angles to the nearest overhead satellite in the constellation.

In contrast, the Globalstar satellite constellation decided not to have ISLs or cross-links and decided instead to have all LEO orbits to have less than 70° inclination North and South so as to concentrate satellite "overhead coverage" to the populated regions of the world and to avoid the polar regions. This approach provided a better look angle for everyone below 65° elevation North and South and simplified the number of TT&C stations that had to be put in place for system control.

The Orbcomm store and forward (i.e., machine-to-machine [M2M]) system chose a low Earth orbit constellation design that was able to minimize satellite size, power, and manufacturing, and launch costs. Yet the Orbcomm system contained enough spacecraft in the constellation to complete global data messaging within a very few minutes.

The once proposed and now defunct Teledesic "Mega LEO" satellite system for broadband Internet services opted for a design with an exceeding large number of satellites to be deployed (originally over 800 satellites plus a huge number of spares). This design was conceived so as to insure a very high elevation angle to support instant high data rate broadband communications via very narrow and high gain pencil spot beams. These features of many, many satellites in a low Earth constellation to support minimal transmission delay, very high elevation (or so-called masking) angles, plus high power transmissions from all orbital spacecraft were unique aspects for this proposed system. This was because, unlike the systems for mobile satellite services that are designed to operate in the radio frequency range around 2 GHz, the Teledesic system was intended to operate in the Extremely High Frequency (EHF) or Ka-band frequencies (with a 30 GHz uplink and a 20 GHz downlink). Satellites that operate in these frequency bands require a direct or uninterrupted line-of-sight connection between the satellite and ground antenna systems to complete a transmission link.

There are certain important similarities between the Teledesic system and the more recently conceived "other three billion" (O3b) satellite system whose designers have opted for a very high-powered medium Earth orbit constellation design with its spacecraft orbiting some 8,000 km above the Earth. The original constellation will consist of eight satellites but will expand to eventually include up to 20 satellites to populate the full MEO constellation.[3]

Molniya, Highly Elliptical Orbits (HEOs), Extremely Elliptical Orbits (EEOs), and Loopus Orbits

There are a family of orbits that are variously described as Molniya orbits, highly elliptical orbits (HEOs), extremely elliptical orbits (EEOs), and "Loopus" orbits. These orbits all have in common the following elements – a very high apogee and a low perigee orbit. In the most extreme configuration, the shape of these orbits can be thought of as being "cigar shaped." The advantage of this type of orbit is that it can have a very long effective "hang" time especially about high latitude countries such as Russia where this type of orbit was first used. In particular, this Russian system employed the Molniya orbit. This orbit had a 12 h period with 8 h of the orbit being above the horizon in the Northern latitudes above the Russia subcontinent. This meant that three satellites placed in three separate Molniya orbits could be deployed like the petals on a flower to provide continuous service to the entire country throughout a 24 hour day. Also the very high elliptical orbit with the very, very high apogee meant that the satellite did not "seem to move" as it ascended and then descended along a very narrow track in the sky.

In recent years as the geosynchronous orbit became more and more populated with satellites providing various communications services, the concept of using HEOs or EEOs once again became an attractive idea. The Sirius Radio broadcasting satellite system was initially deployed in this type of extremely elliptical orbit and several broadcasting satellite systems have been proposed for this type of orbit. A very particular type of HEO or EEO orbit is the so-called Loopus orbit that is depicted in the graphic below. Long duration visibility is available in Northern latitudes for this special orbit during positions 1, 3, 5, 7, 8 and 9. This type of orbit can be utilized for fixed and broadcast applications and Earth stations require only limited pointing capabilities (Fig.5.2).

The main application for these types of orbits is thus essentially for broadcast types of services. This renewed interest in these types of orbits is fostered by the fact that it is no longer easily available to obtain new orbital slots within the GEO (or Clarke orbit). The long periods over which a satellite "appears" to be in the same location can thus serve to emulate a satellite in GEO orbit. Nevertheless, there is a need to have at least three and probably four satellites populating this type of orbit since they typically will only maintain this "apparent" location during their "highest apogee phase" for a period of 6–8 h.

String of Pearls Orbit

Another orbital concept that has been considered by a variety of different system planners over time is the so-called string of pearls orbit. The concept involves the deploying of a number of satellites, such as six to eight spacecraft, in a medium Earth orbit around the equator. The concept here is that the satellites would be equipped with zonal beams that would cover more than one-sixth of the Earth's

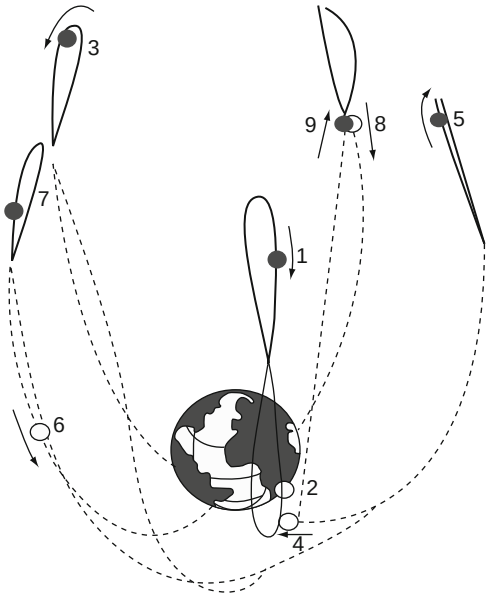


Fig. 5.2 The "Loopus orbit" shown in its movements relative to the Earth's rotation during a 24-h period. (Graphic courtesy of the Author)

circumference (for the six satellite configuration) or more than one-eight of the Earth's circumference (for the eight satellite configuration) so that as one satellite moved below the horizon a new zonal beam from the "ascending spacecraft" would provide an equivalent coverage. One would need to have a significant overlap of coverage to provide a seamless handoff between the "departing" and "arriving" satellite so that the handoff would be entirely seamless and so that the ground antenna systems would not need to track a particular satellite and so that the quality of the signal would not be significantly degraded during the handoff process.

The value of this particular orbital configuration is that there can be continuous coverage to the entire equatorial region (i.e., 3–4,000 km above and below the equatorial belt) where some 2.5–3 billion of the world's population is concentrated.

Part of the time a particular satellite might be over Brazil, or Columbia, or Ecuador, or Peru, or the Congo, or Kenya, or Uganda, or Indonesia, or India, or Southern China, or Laos, or Thailand. Much of the time, however, each and every one of the six to eight satellites would be over heavily populated areas. This is simply a result of the world's land mass geography. Such a satellite network would be well suited to providing domestic services to the various countries of the equatorial region. It would not, however, be well suited to providing international services since it provides limited interconnectivity between equatorial countries and even less with the rest of the world. There are also practical difficulties with how such a system would be financed and a logical system devised for paying for the derived services actually used. Some countries such as Brazil, India, Indonesia, and China might use such a system heavily, but other countries such as Sri Lanka or

Guyana might find it useful to a much smaller degree than much larger nations. Brazil had once thought of deploying such a system to meet its own domestic needs and offer the facility as a "gift" to other equatorial countries of the world.

Quasi-Zenith or Figure-8 Orbit

Another orbit that is now being utilized for mobile satellite communications is that which can be variously described as "Quasi-Zenith" or the "Figure-8 orbit." This is essentially a GEO orbit that is inclined 45° to the equatorial plane and then populated by three or more satellites so that a country near 45° latitude such as Japan (near 45° North latitude) or Australia (near 45° South latitude) always has one satellite "overhead" with a steep look angle to the subsatellite point. Japan was one of the countries to first identify this type of orbit to utilize for mobile satellite communications. Japan has a number of cities with very tall skyscrapers such as Tokyo, Osaka, or Yokohama, and thus satellite communications to user transceivers can be easily blocked by buildings that rise high into the sky. The Quasi-Zenith or Figure-8 orbit provides one satellite overhead with something like an 80° look angle down into the cities and thus a much better look angle than a GEO satellite. One of the unique features of this orbit is that satellites create a pattern that appears like the figure 8 with the top of the orbit being at 45° North and the bottom being at 45° South. Thus, a system deployed over the Pacific Ocean would create excellent coverage for Japan in the North while equally providing coverage with the other satellites in the orbit for Australia in the South. Japan has designed an experimental Quasi-Zenith satellite, named Michibiki (see Fig. 5.3). The Michibiki experimental satellite was designed by the National Institute for Communications and Information Technologies, fabricated by the Mitsubishi Electric Corporation (MELCO) and launched by the Japanese Space Agency (JAXA). The purpose of the experimental satellite is to test out this type of orbit for use for space navigation as well as for mobile communications satellite coverage.[4]

If an operational satellite system were to be deployed then some form of commercial arrangement would logically be created to provide coverage not only in the Northern latitudes but in the South as well. In this instance, it could be used not only for mobile satellite communications but for space navigation as well.

Because of the high altitude represented by the QZSS orbit, it can indeed easily also support a space navigation function. It is for this reason Japanese experimenters decided to equip these satellites with atomic clocks and the ability to transmit space navigation beacons. Satellites equipped in this manner can thus be effectively used to augment the existing Global Positioning Satellite network and thus to provide augmented space-based reference points to allow more accurate data for navigation and mapping. Because the GPS availability enhancement signals transmitted from Quasi-Zenith satellites are compatible with modernized GPS signals, and thus interoperability is ensured. The Michibiki satellite not only has a highly accurate atomic clock but will also be able to transmit the L1C/A, L1C, L2C, and L5 signals that are compatible with the GPS space navigation system

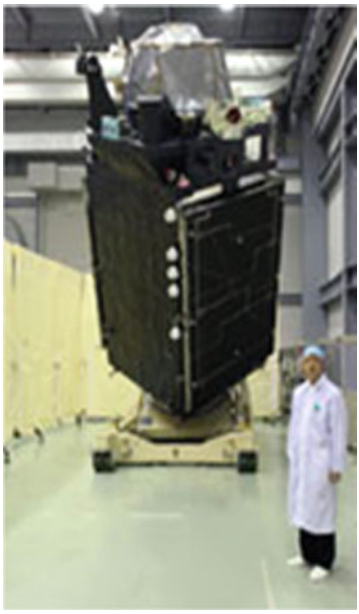


Fig. 5.3 Michibiki experimental satellite to Test Quasi-Zenith orbit for mobile satellite communications services and space navigation (Graphic courtesy of JAXA)

(The Quasi-Zenith Satellite System is a project of the National Institute for Information and Communications Technologies (NICT) in cooperation with the Japanese Space Exploration Agency (JAXA) http://www.spacecom.nict.go.jp/control/efsat/index-e.htm).

Super Synchronous

The above discussed orbits are the main ones used for satellite communications around the world today. The satellites are easier to track, command, and control the closer they are to Earth. The same is true of space debris that is concentrated in these various orbits as well. It would be possible to deploy a satellite in orbits that are farther away from the Earth than Geosynchronous orbits for communications or other purposes. These reasons that one might do this would be to avoid detection, such as for military or defense-related purposes, or to establish a relay point for communications to the Moon, Mars, asteroids, or scientific satellites. One of the often discussed such locations are the so-called Lagrangian Points that exist in space as discovered in 1772 by Joseph Louis Lagrange, the Italian–French mathematician. These are relatively "stable" locations where a satellite once located in these positions are trapped in these orbital positions by the competing gravitational effects of the Earth and the Sun (and to a lesser extent the Moon and the other planets). Thus, there are five such points as shown in the following diagram. It has been suggested the L-1 Lagrangian Point could be used as a suitable point for

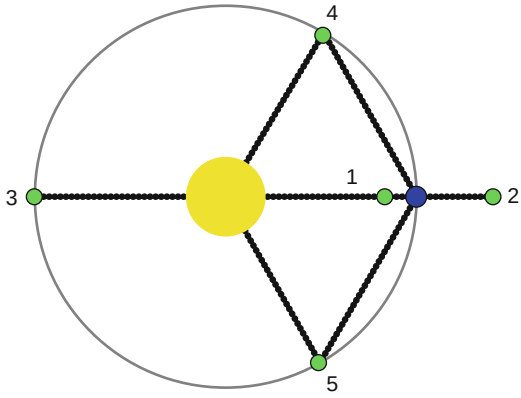


Fig. 5.4 The five Lagrangian points as shown in relation to the Sun's and the Earth's orbit

observation of the Earth's atmosphere (i.e., the Triana Project by NASA to observe the Ozone Hole and other atmospheric phenomena), a Space colony or a translunar communications link between the Earth and the Moon. It has also been suggested that other Lagrangian points such as 3, 4, and/or 5 might be used as communications relay positions for broadband communications satellites to provide links between Mars and Earth (Fig. 5.4).

These points are highly desirable for very long-term satellite positioning since once a satellite or space colony reaches one of these locations it will remain there indefinitely – trapped by the gravitation of the Sun and Earth.

Earth Station and User Terminal Design for Different Orbits

The different orbits described above require different types of ground antennas to operate. The different types of satellite Earth station antennas and terminals are addressed in detail in later chapters. There are some basic concepts that are important to note with regard to antenna designs as they relate to different types of orbits used for satellite communications and particularly with regard to how antenna designs can be optimized. The GEO orbit allows high performance or high gain antennas to be exactly pointed toward a satellite with a minimum of "tracking." Thus a large dish (i.e., a parabolic antenna with a large aperture) to point toward a GEO satellite and remain virtually stationary without tracking a satellite as it moves rapidly from horizon to horizon. Even in the case where a GEO satellite is building up inclination North and South of the equator, a relatively simple mechanical system can be added to the antenna steering system to move in tandem with these small migrations North and South during a 24 h period that is highly predictable. Ground antennas working to GEO satellites for fixed satellite services (FSS) have a higher sensitivity because they can point a focused beam constantly at a "stable satellite." This means that satellites providing FSS can have smaller antennas in space and lower power because the ground antennas have the ability to send a more focused beam to the satellite and

receive a more focused beam from the satellite without a high-speed tracking. This is in contrast to the lower gain end user antennas that typically work to medium Earth orbit or low Earth orbit satellite constellations. These lower gain user antennas are likely to be omni antennas that can capture a signal from any direction or squinted beam omni antennas that can capture signal from anywhere above the Earth's horizon. Thus, these are ground antenna systems that are designed to capture signals from a satellite that moves rapidly across the horizon.

As in most cases, there are exceptions to the rule. There are especially desired and more expensive higher gain antennas designed to support tracking, telemetry, control, and monitoring functions that have high-speed tracking capabilities that can support the operation of LEO and MEO orbit satellites. These antennas have large aperture dishes but that are also able to track even a low Earth orbit (LEO) satellite that can cross from horizon to horizon in a few seconds. Such ground antenna installations are quite expensive and thus are built and operated only to support the safe operation of satellite networks. These antennas are too complex and expensive to be utilized by actual satellite system users. The broad band consumer users, particularly those equipped with hand-held transceivers or micro terminals to support mobile satellite services (MSS), have simple omni or near omni antennas or quite small antennas with limited tracking capabilities.

The other exception comes when one seeks to use a GEO orbit satellite to support MSS type operations. The higher gain antennas for FSS markets or for direct broadcast services work quite well when the satellite is stable and the ground station antenna is stable. If you attempt to support mobile satellite services from a GEO orbit there is immediately a problem, the satellite is stable, but the user terminals are typically moving. The users and their ground antennas could be moving through a forest, an underpass, or through the middle of a city. In this case, you do not have the satellite antenna constantly pointed toward a higher gain dish antenna on the ground. This means that you are forced to design much high-powered and larger antennas on the GEO satellite to compensate for the much smaller and lower performance "omni" or "near omni" type antenna that the user on the ground must utilize.

The key to designing and engineering a successful satellite network involves what is called "systems engineering" or "system optimization." One must have a sufficient level of power and a focused degree of transmitted radio frequency signal to "close a link" between a satellite and a ground receiver (on the ground, the sea, or in the air). The calculation of the antenna gain (or size) and the power (on the ground and on the satellite) is called a "link budget." Additional power and antenna gain above the minimum needed to complete a "link budget" is called "link margin."

If you were to design a system – a "cable in the sky," where there is only one single high-capacity pathway – let us say between New York and London, then you could afford, within your system engineering, to have two very large, high gain antennas that send their signals back and forth between a simple satellite with relatively low power and small spacecraft antennas. For such a cable in the sky this might result in the lowest overall systems cost.

If one takes the opposite extreme and wanted to design a system to send a direct broadcast television signal to every home in Europe, the system optimization

process would be dramatically different. Now one would want a very high-powered satellite indeed to send a signal to millions of very small, low-cost, and easy-to-install antennas on the ground. This in many ways accurately describes the process of system engineering that has characterized the development of satellites over the last 30 years. This is sometimes called "technology inversion." This is the evolution from very small and low-powered satellites that worked to quite large, powerful, and expensive Earth stations to the reverse situation where there is a very large and powerful satellite that distributes signals to low cost user terminals.

This means that within the process of "technology inversion" the number of users on the satellites on the ground has grown from dozens, to hundreds, to thousands, to now millions. The investment of a large amount of money in increasingly powerful satellites with very sophisticated and large in-orbit antenna systems makes economic sense if this allows the overall cost of the entire system to go down. If one takes the example of a direct broadcast satellite, the logic goes as follows. Even it costs many hundreds of millions of dollars to build and launch a high-power DBS satellite, this still becomes economic if it can reduce the individual cost of "millions" of consumer antennas on the ground to hundreds of dollars rather than thousands or tens of thousands of dollars. This is because there is just one satellite or one satellite and spare, but there are a huge number of user antennas on the ground. If one can reduce their costs by just $100 and there are ten million users, the cost for the ground part of the overall system goes down by $1 billion.

Satellite system engineers actually spend a lot of time trying to figure out how much money will be spent on the satellites manufacture, launch, and satellite operations on one hand versus how much money will be spent on the other hand by the consumers on ground antenna systems. In most satellite systems today, the bulk of money will be spent on consumer-based antennas to receive television signals, mobile satellite communications services, or high-speed broadband data services or telephone circuits. This is simply because there will be so many consumer users – typically numbering in the hundreds of thousands or even millions. Systems engineers then try to design an "optimized" total system that represents the overall lowest cost system. Sometimes they get it quite right and the satellite network is successful and attracts the projected number of users and the lowest possible cost system. In other cases they get in wrong and the satellite, the ground antennas, or some other aspect of the system is badly designed for the market and the system fails and goes bankrupt. Examples of where the projected number of users on the ground failed to be achieved were the Globalstar, Orbcomm, and Iridium satellites that subsequently entered into bankruptcy.

Relative Economics of Different Satellite Orbits

Most satellite planners and systems engineers start out by considering the service that is desired to be provided and the type of orbital configuration that can best provide the desired service at the lowest net overall cost and with the highest level of reliability and service quality. As can be seen in the graphic below, one can cover

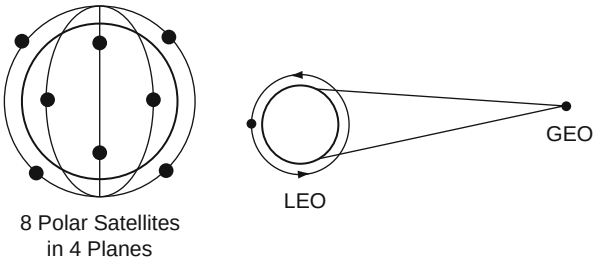


Fig. 5.5 Different Earth coverage by satellites at different orbital heights (Graphic courtesy of the Author)

more than a third of the Earth from GEO (or Clarke orbit). One can reasonably cover the Earth with about eight satellites in medium Earth orbit (MEO), but when the satellites get very close to the Earth one, in low Earth orbit, it turns out that one need forty or more satellites to provide total coverage of the globe. One only needs to shine a flashlight on a round balloon or basketball at varying distances to see why the satellite coverage capability varies as one nears the Earth (Fig. 5.5).

The problem is that the tradeoffs between and among satellite and satellite launch costs, ground antenna costs, service quality, and tracking, telemetry, command, and monitoring costs and other costs such as marketing, advertising, billing, and regulatory services are not easy to project before a system is designed and deployed. Sometimes, when the service is almost entirely new, the ability to project market and consumer acceptance can be dramatically off – as was the case of the three satellite systems first conceived to provide land mobile satellite services around the world. These three initial systems, namely, Iridium, Globalstar, and ICO all ended up in bankruptcy without it clearly being established as to exactly what went wrong. Possible explanations include that the new market was over estimated, the cost of the system too high, the wrong type of orbital configuration was chosen, or the satellite service design or ground antenna unit for the consumer were not well matched to the market demand. What is clear is that GEO-based satellite systems that require only a few satellites to begin operations and collect revenues are often a lower risk business propositions.

Some analysts suggest that it is a "very steep climb" to seek to deploy a large-scale LEO constellation system at the outset of a new service. In short, a LEO satellite system requires a long lead time and a very large investment to build and launch many dozens of satellites. This large expenditure becomes particularly difficult as a start-up business because there is no established revenue stream. In short, there is enormous challenge in designing, building, deploying, and testing a large LEO constellation with no incoming revenues or established market base. Certainly the bankruptcies of the Iridium System with 72 satellites plus spares, the Globalstar system of 48 satellite plus spares, and the Orbcomm satellite network of 48 satellites plus spares indeed all seem to constitute a strong caution against deploying LEO systems as a totally new start-up business. The Teledesic satellite system that originally envisioned deploying some 840 satellites plus 80 spares was the ultimate example of new LEO system that required the launch of way too much hardware prior to the realization of any revenues

against a huge capital debt. In the case of ICO and Teledesic, these projects folded before actual satellite launches began.

What is clear from an economic sense is that a GEO satellite system can be initiated with a single satellite in orbit. Indeed one can lease one or more transponder from an existing GEO satellite system and increase capacity as markets and revenues grow. In short, GEO systems allow the strongest case for organic growth of satellite services for localized, national, or regional services and in many cases the ground antenna systems can be shifted from leased satellite capacity to a dedicated satellite network as market demand grows. A medium Earth orbit (MEO) system can be started with far few satellites and perhaps as few as six to eight satellites, although many MEO constellations do require a larger number of satellites. From an economic standpoint, the MEO constellation, in many ways, represents a compromise between a GEO and a LEO system in terms of number of satellites to manufacture and launch, and complexity of TT&C operations. The actual design and implementation of a communications satellite system, however, involves far more than the orbital configuration and the desired national, regional, or global coverage that a satellite system provides to a specific market.

These factors must take into account service requirements such as transmission latency, reliability, quality/bit error rate, coverage, and look angles; design, performance, and cost of user antennas/terminals; operational cost and complexity (including TTC&M design and costs); as well as overall cost efficiencies, capital financing, regulatory constraints, and strategic business case.[5] Nevertheless, one of the key starting elements in any satellite system design will typically be the orbital configuration to be utilized. This thought process will often start with the feasibility of obtaining access to one or more GEO satellite locations or the lease of capacity on an existing satellite network. MEO or LEO constellation designs thus represent a "step beyond" in terms of pursuing a business plan that will typically involve an element of greater technical, financial, and regulatory risk. These risks will in many cases be considered to be acceptable in exchange for improvements in high service quality (i.e., lower transmission delay and lower path loss), ability to attain access to orbits and allocated frequencies that may be available, or improvements in user antenna compactness, complexity, and cost.

Conclusion

The ability to attain access to allocated radio frequencies to operate a satellite system continues to be an ever more challenging activity. The difficulty grows as more and more satellites are launched into Earth orbit and very few new opportunities exist for satellite system operators without engaging of closer and closer spacing of GEO satellites, or the use of the quite demanding frequencies in the millimeter wave bands, or possibly opting to deploy satellites into orbital configurations beyond the most "conventional choice" of the GEO orbit. The challenges of opting for other orbital configurations have actually spurred the trend toward closer and closer spacing of satellites in the GEO orbit and the implementation of GEO satellites that utilize the higher frequency bands. Today,

the problem of frequency coordination has become even more difficult than ever before. This is simply because there are more and more satellites that are operating at higher and higher power levels and spaced ever more closely together.

In addition, the problem of orbital debris has increasingly emerged as a problem for LEO, MEO, and GEO, and polar orbits and indeed for the general sustainability of all space efforts near Earth in the future. Efforts to coordinate among the various operators of satellite networks to minimize the possibility of collisions to among spacecraft are also being intensified through coordinative efforts. This has resulted in organizations reducing frequency interference (now known as simply the Satellite Interference Reduction Group (sIRG)) and also led to the creation of the Space Date Association that has created a coordinated global data base that monitors the orbits of various satellite systems such as those of Intelsat, SES, Inmarsat, and Eutelsat so as allow avoidance techniques to be followed in case of impending satellite collisions. It is hoped that this initiative will expand to include more and more operators and will include more and more orbits.

Cross-References

- ▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ▶ Fixed Satellite Communications: Market Dynamics and Trends
- ▶ History of Satellite Communications
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends
- ▶ Satellite Communications Video Markets: Dynamics and Trends
- ▶ Space Telecommunications Services and Applications

Notes

1. Op cit. (Pelton 2000, pp. 73–87)
2. Hermann Nordung, the Slovenia Scientist, www.astronautix.com/astros/noordung.htm
3. "Agreement signed with Arianespace for Initial O3b Satellite Launches" O3b Networks http://www.o3bnetworks.com/Media_Centre/press_release_details.aspx?id=60
4. The Quasi-Zenith Satellite System is a project of the National Institute for Information and Communications Technologies (NICT) in cooperation with the Japanese Space Exploration Agency (JAXA) http://www.spacecom.nict.go.jp/control/efsat/index-e.htm
5. Op cit. (Pelton 2001, pp. 1–31)

References

C. Charles, F. Ciovanni, F. Lauren, G. James, M. Mikhail, Mc.Kay. Chris, R. Michael, S. Isabelle, The Universe and Us Chapter 5, in *The Farthest Shore: A 21st Century Guide to Space*, ed. by J. N. Pelton, A. Bukley (Apogee Books, Burlington, 2009), p. 157

J.N. Pelton, *Global Communications Satellite Policy: Intelsat, Politics and Functionalism* (Lomond Books, Mt. Airy, 1974), p. 48

J.N. Pelton, *Research Report: Satellite Communications – The Transition to Mass Consumer Markets, Technologies and Networks* (International Engineering Consortium, Chicago, IL, 2001), pp. 228–230
J.N. Pelton, *Basics of Satellite Communications* (International Engineering Consortium, Chicago, 2006), p. 73
J.N. Pelton, S. Madry, Satellites in Service to Humanity, in *The Farthest Shore: A 21st Century Guide to Space*, ed. by J.N. Pelton, A. Bukley (Apogee Books, Burlington, 2009), p. 220
M. Williamson, *The Communications Satellite* (IOP Publishing, Bristol, 1990), pp. 76–83

Fixed Satellite Communications: Market Dynamics and Trends

6

Peter Marshall and Joseph N. Pelton

Contents

Introduction 116
Evolution of FSS Services and Competition from Terrestrial Communications Systems... 117
Digital Satellite Communications and the Move to Higher Frequency Bands 120
Decentralization of FSS Services as Small Ground Systems Move to the "Edge" of Global Networks..... 124
Regulatory Shifts Concerning FSS Systems to Make Them Openly Competitive 128
Evolution of FSS Markets from Global Networks to Regional and Domestic Satellite Systems 131
Cross-References..... 134
Notes 134
References..... 134

Abstract

The history of fixed satellite services (FSS) systems, in terms of technological and institutional development, has been previously provided in ▶ Chap. 3, "History of Satellite Communications" of this handbook to a very large extent. Thus, this chapter addresses the market trends related to FSS systems and also discusses how a variety of new types of satellite services has evolved out of the initial FSS networks over time.

The market dynamics and trends of FSS systems are particularly addressed in terms of four main factors: (1) the competitive impact of high-efficiency fiber-optic terrestrial and submarine cable communications networks; (2) the

P. Marshall
Royal Television Society, England, UK
e-mail: pmsatellites@aol.com

J.N. Pelton (✉)
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: peltonjoe@gmail.com

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_6, © Springer Science+Business Media New York 2013

conversion of FSS systems from analogue to digital services that allowed FSS systems to be more cost-efficient and use spectrum more efficiently as well as migrate to spectrum in higher bands more effectively; (3) the move of FSS systems toward deployment of smaller and lower cost ground systems (variously called VSATs, VSAAs, USATs, and micro-terminals) that allowed services to migrate closer to the "edge" of telecommunication user networks (i.e., satellite services directly to end user facilities); and (4) a shift in regulatory policy that allows FSS systems to compete directly for services that has generally served to reduce cost and spur innovations in services and applications.

These four trends have combined to contribute to what has been previously described in ▶ Chap. 3 as "technology inversion." This "technology inversion" has thus seen FSS systems in space become larger, more complex, longer-lived, and more powerful as ground systems have become more user-friendly, lower in cost, and designed to interface directly with users at localized office facilities or even small office/home office (SoHo) VSATs or micro-terminals. These technological, regulatory, and market-based trends have shaped the FSS networks and related market dynamics. All four of these trends have dramatically reshaped the nature of FSS services for both commercial markets and defense-related satellite networks around the world.

The historical trend in FSS markets has been the initial development of global networks since global connectivity was the highest value market and the most underserved by terrestrial telecommunications networks available in the 1960s. Over time, satellite technology matured and the economical viability of regional and domestic satellite systems evolved in the years that followed. Today there are some 300 FSS satellites, essentially all in GEO orbit where these systems provide a complex combination of global, regional, and domestic satellite services. Although broadcast satellite services have outstripped FSS in terms of market value and sales, the FSS is still a very large and growing multibillion dollar industry.

Keywords

Analogue to digital conversion • Bit error rate • C-band • Digital satellite services • Domestic satellite systems • Fixed satellite services • Frequency bands of satellite service • International Telecommunication Union • Internet protocol (IP) over satellite • Ka-band • Ku-band • Micro-terminal • Quality of service • Regional satellite systems • Satellite markets • Spectrum allocations • Spectrum efficiency • Submarine cable • Ultrasmall aperture terminal (USAT) • Very small aperture antenna (VSAA) • Very small aperture terminal (VSAT)

Introduction

This chapter notes how this first type of communications satellite service was defined by the International Telecommunication Union as fixed satellite service (FSS). With the maturation of satellite technology over the years that followed, the development of lower cost and easier to use ground systems, together with

regulatory shifts, allowed the further development of direct broadcast satellite services, mobile satellite services, and even store and forward data relay or machine-to-machine type services. FSS services, as the oldest and most mature of the satellite services, is the father and in some cases the grandfather of all the various satellite communication services that have followed since the start of commercial services in the 1960s. Both mobile satellite services, which evolved in the 1970s, and direct broadcast satellite services that date from the 1980s, have benefited from the initial technology first developed for commercial FSS systems (Chartrand 2004).

The development of these additional services as well as defense-related satellite services is discussed in detail in ▶ Chaps. 7, "Satellite Communications Video Markets: Dynamics and Trends," ▶ 8, "Mobile Satellite Communications Markets: Dynamics and Trends," and ▶ 9, "An Examination of the Governmental Use of Military and Commercial Satellite Communications."

The key market dynamics for FSS are discussed in this chapter in terms of five predominant trends that can be concisely stated as: (1) evolution of service capabilities and related competition from terrestrial communications systems; (2) digital satellite communications and the move to higher frequency bands; (3) decentralization of FSS services as small ground systems move to the "edge" of global networks; (4) regulatory shifts with regard to FSS systems to make them openly competitive around the world; and (5) the shift of FSS systems from primarily serving global markets to more and more satellite networks serving regional and domestic markets. These five interrelated trends are discussed and analyzed in terms of their impact on the FSS markets.

The pattern for FSS markets was for networks designed for global services to evolve first because this was the highest value type service. Regional and domestic FSS systems followed thereafter. This was a logical consequence as satellite technology matured and market demand allowed these regional and domestic systems to become economic around the world, particularly as lower cost satellite antennas and terminals became available. The development of military and defense-related traffic has represented yet another dimension of the market for FSS networks around the world. These market trends and dynamics are addressed separately in ▶ Chap. 9, "An Examination of the Governmental Use of Military and Commercial Satellite Communications."

Evolution of FSS Services and Competition from Terrestrial Communications Systems

The evolution of fixed satellite services (FSS) in the earliest days of satellite communications was largely the history of the Intelsat satellite system in the period from 1965 through the early 1970s. The first Intelsat satellite, known as "Early Bird," was essentially a "cable in the sky" that could only connect point-to-point service. Then came the Intelsat II series which was able to provide multi-destination service and connect several points at once. This satellite was designed and built with US government funding to support the US manned space program Gemini so that ships at sea could maintain communications with the crew in the space capsule.

The Intelsat III series that was launched in the 1968/1969 time period were the first commercial satellites to provide a full range of satellite services similar to today's satellites in terms of providing multi-destination services to many points with the capability to provide voice, data, color television, and high-quality radio channels.

It was this Intelsat III series that in July 1969 was able to provide global coverage of the Moon landing by Apollo 11. It was only a few weeks before the Moon landing that true global connectivity via satellite was established. As of 1970, satellite communications had become the predominant form of international communications as this technology provided broader band and lower cost connectivity than the coaxial submarine cables of the day. Further, multi-destination satellites were able to connect any country in the world to a globally interconnected network by constructing and operating only a few Earth stations. As the first Director General of the Intelsat and former head of Entel Chile once said:

> Communications satellites changed almost everything for our country. For the cost of one Boeing 707 airplane, we could build and operate a satellite earth station that could allow Chile to be fully connected to the rest of the world. (Interview with Santiago Astrain 1974)

The cost of international telephone calls from remote areas of the world could exceed $50 a minute prior to the advent of satellite communications. However, since the arrival of global satellite systems and ever more efficient submarine cable systems, the cost of an international call has dropped to a level that is little different from the cost of a long distance call within a country. Prior to the advent of satellite communications, the global delivery of live television was simply not possible. Today some over 18,000 video channels are available worldwide via satellite connections (Pelton 2006).

For over a century there has been an ongoing rivalry between terrestrial submarine cables and wireless communications systems to provide better, lower cost, and more reliable communications for overseas communications. In the middle of the nineteenth century, telegraph submarine cable systems began to provide limited international communications service. These cables had limited capacity and were subject to disruptions and failures due to storms, trolling fish vessels, and other factors. The invention of shortwave radio provided a way to provide overseas voice and data services at lower cost and with greater throughput capability. However, shortwave radio was subject to disruptions as the result of space weather interference with the ionosphere. The invention of coaxial cable systems capable of carrying voice traffic in the 1940s and 1950s moved international voice and data traffic back toward terrestrial technology. The resulting submarine cable systems, even with 3 KHz telephone channel spacing and the so-called time assignment speech interpolation (TASI), still had very limited capacities of only 72 voice circuits in the mid- to late 1950s. The advent of satellites such as the Intelsat I with 240 voice circuits in 1965 and then Intelsat III with 1,200 voice circuits plus two color television channels sharply shifted international telecommunications traffic to satellite connections. Satellite circuits were lower in cost and allowed much more voice and data traffic to be provided between the continents and enabled international television transmissions to be provided, both technically and economically (Pelton and Alper 1986).

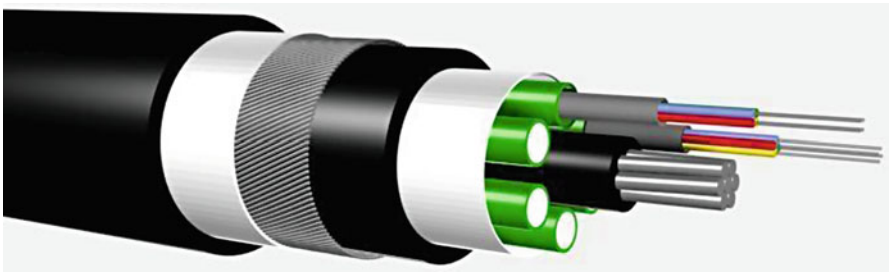


Fig. 6.1 Fiber-optic submarine cables have become predominant for the heaviest transoceanic routes

Beginning in the 1980s with the advent of new fiber-optic submarine cable technology, the international telecommunications market shifted focus once again. Fiber-optic submarine cables became more and more cost-effective, broadband, and higher quality and thus quickly began to reclaim international telecommunications services, at least on the heaviest transoceanic links (see Fig. 6.1 above). This shift from satellite telephone and data back to submarine cables, particularly for trans-Atlantic and trans-Pacific Ocean traffic through the 1990s and up to the present time, was hastened by several factors:

- *Quality of service*: Transmission via fiber-optic submarine cables, as measured in bit error rate, could be as very low and typically could be in the range of only 10^{-10} or even 10^{-12}. This was an unprecedented level of transmission quality. Further transmission times were typically less than 100 ms as opposed to the 250 ms of transmission delay associated with geosynchronous satellite transmission. This shorter latency or transmission time made fiber the preferred choice for telephone service.
- *Cost of service*: The very heaviest routes, such as between the United States and Europe, could be considerably lower than the costs associated with international satellite connections. Satellites remained cost competitive for more remote locations with thinner routes of traffic or to locations not served directly via fiber-optic networks. Satellites also remained cost competitive for television distribution services.
- *Structure of service provision and ownership*: Submarine cable services were provided as if they were actually owned and capitalized by telecommunications service providers under what were called "indefeasible rights of use" (IRUs) that made provision of service more cost-effective and profitable under current regulatory policies then in effect.

The cost efficiencies of both fiber-optic submarine cables and satellites have continued to plummet as both of these technologies have matured. The improvement of the technology, the extension of the lifetimes of these systems and more have now been so dramatic that the "capital cost" of a single voice circuit might be as low as $5 per voice circuit on a submarine cable and under $50 a voice circuit on an advanced communications satellite.

The economics are such that other costs associated with international telecommunications, such as marketing and sales, advertising, billing, and operations, now tend to be predominant over the creation of the international link itself. Thus, issues such as quality of service, lack of transmission delay, redundancy of service links, network design and complexity, and the ability to establish links to particular locations with great speed often tend to dominate the decision as to whether or not a link is established via satellite or submarine cable. In general, it can be said that most heavy route traffic between countries or even within countries today are carried by fiber-optic networks. Satellite communications networks thus tend to carry medium to thin route voice or data traffic to supplement fiber-optic networks and a variety of different types of television services where distribution to widely distributed audiences of business networks may be involved.[1]

The need to create integrated global telecommunications networks to serve the "enterprise networks" of multinational enterprises, national governments, international organizations, and military systems has seen a growing trend toward forming combined and seamless networks. These combine fiber and coaxial fiber networks, broadband terrestrial wireless networks, and satellite systems under unified ownership. This is, in part, the result of the growth of Internet, intranets, virtual private networks, and digital networks that provide broadband to support voice, data, video, and audio services on demand. The digital satellite revolution and the provision of voice and other services over IP, is discussed immediately below.

Digital Satellite Communications and the Move to Higher Frequency Bands

The provision of satellite services for the initial two decades was essentially via analogue-based services. Analogue services and multiplexing systems using frequency division multiple access (FDMA) were inefficient in several ways. The amount of information that was sent via satellite was inefficient in terms of information transmitted per Hertz (or information sent per cycle per second). Also, there were just a limited number of carriers of set size for everything from small routes to very large routes. The information throughput density was progressively lower for smaller and smaller carriers for thin routes of traffic because of the need to separate the carriers with guard bands and because the carriers were only efficient when completely filled with actual active voice traffic. Once a carrier was filled with traffic, however, there was a need to jump to a larger fixed carrier size to accommodate growth. In all of these ways, the analogue satellite service was inefficient. In the 1980s and 1990s, there was a digital revolution in satellite communications and most space traffic was converted from analogue transmission using FDMA multiplexing to either time division multiple access (TDMA), code division multiple access (CDMA), or a special system developed for very thin routes of traffic known as the SPADE system that allowed single channels to be used on the satellite on a demand-assigned basis.

Some ways by which digital satellite communications can be considered superior to analogue satellite service include the following:

- Greater ability to operate at higher transmission speeds.
- Improved quality of service – especially in a high noise environment.
- Greater compatibility with terrestrial digital switches – that now predominate.
- Greater compatibility with digital fiber-optic systems.
- Easier to allow accommodation of encryption/decryption systems.
- Easier accommodation of digital signal compression techniques.
- Easier accommodation to onboard digital switching and onboard signal processing to overcome rain attenuation and other forms of interference.
- Greater compatibility with all forms of digital services – whether digital television, digital voice, voice over IP, multimedia over IP, and all forms of data services (Lewis 1988).

In terms of market efficiency, the conversion to digital satellite services allows very high new efficiencies to be achieved. A 72 MHz transponder using analogue technology for high-quality television was typically able to derive two color television channels of reasonably high signal to noise (S/N) quality while operating to very highly sensitive Earth stations of 18 m or larger. In contrast, using digital TDMA technology, on the order of 14 high-quality television channels could be derived from a 72 MHz transponder while also operating to smaller satellite antenna systems of 10 m diameter aperture size. The improved throughput for voice channels and data transmission was not as dramatic as was the case for digital television, but there were nevertheless considerable gains.

The gains in efficiencies were approximately four to six times depending on a variety of factors such as the volume of traffic, the size of Earth station antennas, etc. These gains created market disruptions during the transition because the dramatic gains in efficiencies offered by digital services could not easily be reflected in pricing policies without creating a shortfall in revenues.

Also, because the ownership of the satellites and the space segment was divided from the ownership of the Earth stations within the structure of the Intelsat organization, there was a division of interests involved in terms of seeking the benefits from digital satellite services. The owners of the ground stations, especially those with low volumes of traffic, questioned why they should invest in the new digital equipment after having invested in analogue equipment only a short time before. Their position was that the benefits, which would flow from digital efficiencies, went primarily to the largest users and owners of the space segment and not to the smallest users – particularly if they continued to use analogue equipment. Many of the smallest users of the space systems, especially the developing countries, thus had the least incentive to convert to the more efficient digital equipment.

The resulting decision that ensued from this dilemma of what might be called conflicting interests of conversion to the more efficient digital technology was a compromise within the Intelsat Board of Governors. This compromise decision was to phase in the "efficiency pricing" for digital services over a series of years. In short, the plan was to phase in the new pricing for digital services and not to seek to

reflect immediately all of the gains achieved by rapid, high-efficiency digital throughput all at once, but to gradually reflect the digital efficiency as TDMA systems were introduced. This was known within the Intelsat organization (the organization that dominated international satellite communications up through the 1980s and were the first to introduce commercial digital services) as the decision in the "spirit of Chang Mai." This was so-named because the Intelsat Board meeting that reached this compromise decision was held in Chang Mai, Thailand, where the local markets were known for their intensive bargaining over price.

In the years that followed, digital conversion continued apace in international, regional, and domestic satellite systems. In many of these systems, networks began with digital systems and thus there was not a question of analogue to digital conversion or the need for a transitional pricing scheme as the switchover occurred.

The competitive processes that were set in motion with the divestiture of AT&T in the United States in 1984 and the liberalization of telecommunications competition within Europe and Japan in the following years helped to speed the conversion to the more efficient digital technology in the form of TDMA and SPADE and subsequently CDMA and spread spectrum services. (This relationship between and among the technology, the market dynamics, and the regulatory process are discussed later in this chapter.)

Ironically, the greater efficiencies of digital satellite services and the reduced cost of service led to a rapid surge in demand. International satellite communications and international submarine transmission capabilities in the 1960s and 1970s were miniscule in comparison to national telecommunications networks. The dramatic decrease in cost for telecommunications and IT systems that occurred in the 1980s, further driven by competition and the spread of multinational enterprises, led to a dramatic increase in demand for international communications. Thus the digital satellite revolution that was thought would create excess satellite system capacity had almost the reverse effect. The net result was that the communications satellite spectrum that had been the mainstay of the satellite industry in the 1960s, 1970s, and 1980s was almost saturated even with the efficiencies that digital communication satellite services engendered. In key locations for geosynchronous satellites, providing for relay over the Atlantic, Pacific, and India oceans, the C-band spectrum was fully consumed.

The new growth of regional and domestic satellite networks further compounded the problem of limited available spectrum for geosynchronous FSS services. The result was to push forward to exploit higher frequencies and also to seek more efficient designs for FSS satellites to allow more reuse of available frequencies. Both solutions were needed to keep up with rapidly growing market demand.

Technical innovation led to the creation of more efficient designs with FSS satellites deploying many more spot beams that allowed frequency reuse. Spot beams that were sufficiently isolated from one another allowed the same frequencies to be used over again, just as was being done with terrestrial cellular systems. Digital switches on board the satellites allowed these various spot beams to be interconnected. A signal could thus go up to the satellite in a spot beam at one location and then be switched to another spot beam for the downlink connection.

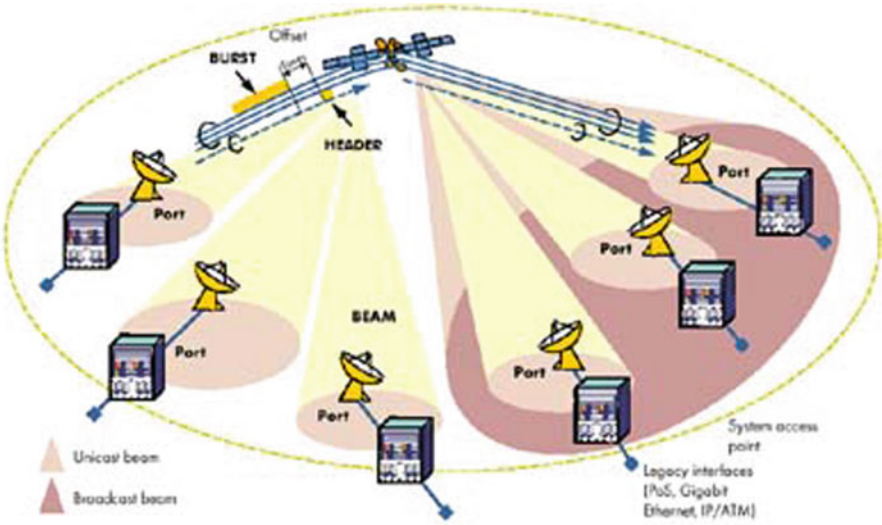


Fig. 6.2 Multi-beam ARTES satellite showing digital spot beams and onboard beam interconnection between beams (Illustration courtesy of ESA)

If these spot beams illuminated different parts of the Earth's surface, then the same spectrum could be used without interference. This type of multi-beam satellite that uses high-speed onboard spot beam digital switching thus can provide interconnection as illustrated in Fig. 6.2 above. This illustration shows the Artes satellite, which has been developed by the European Space Agency to provide flexible interconnection of many different VSAT ports. This design allows the efficiency of multiple reuse of the same spectrum and the high-efficiency spot beams can support more rapid throughput within the high-powered beams. These higher-powered spot beams allow smaller aperture antennas to operate to the satellite and also allow for more margin against rain attenuation.

The migration of more and more traffic from the "C"-band spectrum (i.e., 6 GHz uplink and 4 GHz downlink) to the "Ku"-band spectrum (i.e., 14 GHz uplink and 12 GHz downlink) thus accommodated new growth associated with more and more regional and domestic systems and more demand for international services. The Ku-band was in many ways well suited to spot beam operation since the higher frequencies and thus smaller wavelengths were suited to creating higher and higher gain spot beam antennas that could be smaller yet have higher gain just because of the physics of radio waves. A Ku-band antenna could be four times smaller in aperture size but has the same gain as a C-band antenna.

The transition to higher frequencies was not without its difficulties. The new and more demanding higher frequency transmission equipment (on the ground and in space) was more difficult to engineer and build and was thus more expensive. Further, rain attenuation problems that were minimal at C-band increased as one moved up the microwave band to the higher frequencies. The closer the wavelength

of radio waves approaches the size of raindrops, the greater the problem of heavy rain acting as a sort of lens to distort the pathway of radio wave transmissions to and from the satellite. Thus, more power margin had to be added to overcome these rain attenuation problems at the higher frequencies.

Most recently, the demand for additional satellite capacity has driven satellite services toward even more powerful and narrow spot beams interconnected by digital switching technology to allow even more frequency reuse. Market demand has also supported the move upward to the still higher "Ka-band" frequencies (i.e., 30 GHz uplink and 20 GHz downlink). The rain attenuation issues associated with "Ku-band" are even more present with "Ka-band" frequencies and the much higher frequency equipment is even more difficult to design and build. Thus, the cost of the Ka-band equipment is still higher than the Ku-band equipment. There is also a need for greater power margins to protect again heavy rainfall (i.e., rain attenuation).

One might ask why not accommodate traffic growth and new market demand by simply allocating new frequencies in lower bands? The problem is that the demand for terrestrial mobile wireless communications has outstripped all other demands for over a decade. There is no realistic hope of new satellite communications allocations for FSS requirements in lower frequency bands. The likelihood of new allocations for FSS services in the microwave band for instance is almost none. This is particularly true since broadcast satellite services (BSS) and mobile satellite services (MSS) are also seeking new allocations as well. The BSS systems, because they provide direct-to-home services to millions of customers, and MSS systems because also serve millions of customers directly at locations on land, the sea, and the air, are likely to receive priority for obtaining new frequency allocations over FSS systems because of considerations related to rain attenuation and consumer costs.

The bottom line, as noted in more detail above, is that digital services are more efficient than analogue systems in being able to overcome noise and interference. They are certainly better suited to rapid switching of digital traffic between numerous spot beams on the satellite. This factor alone has been critical to the growth of both FSS and MSS satellite systems. Digital satellite systems have also been critical to the effective use of small VSAT and micro-terminals on the ground. The efficiency of digital satellite services and the resulting reduction in the cost of services stimulated the rapid growth of global, regional, and domestic demand and has also seen a shift of space-networked FSS offerings to ever higher frequencies. Thus FSS offerings are now in the C-band, Ku-band, and Ka-band and there could conceivably be use in future years in even higher bands such as the so-called Q, V, and W bands in the millimeter wave frequencies.

Decentralization of FSS Services as Small Ground Systems Move to the "Edge" of Global Networks

The early days of satellites were controlled by the large telecommunications monopolies that saw fixed satellite services as a means to interconnect national

communications with overseas countries because of the limited capacities of the submarine cables of the day. In this early satellite market, large national Earth stations connected to satellites of still limited capacity and therefore it was the national telecommunications terrestrial networks that controlled all international traffic. The subsequent emergence of national satellite systems and national television satellite distribution changed not only the market structure, but also spurred the rise of new satellite systems to compete with national terrestrial networks.

Once this trend started, it created increasing pressure to design smaller and more cost-efficient satellite Earth stations that could bring traffic connectivity closer and closer to the headquarters of large businesses, to satellite broadcasters, and to cable television networks. It likewise created the demand to design and build very low cost, small receive-only satellite ground stations for consumers to get television and radio programming. This trend started with the early national satellite systems in Canada, the Soviet Union, the United States, and Indonesia and then spread to dozens of countries around the world. In turn, this spawned what might be called the VSAT (or the very small aperture terminal) revolution. Instead of hundreds of Earth stations to connect the countries of the world, there were, over time, hundreds of thousands of transmit and receive small satellite antennas located at businesses and eventually many millions of television receive-only (TVRO) terminals.

This trend started during the analogue era of satellite communications but mushroomed with the dawn of the era of digital satellite communications. Digital transmission, with its more efficient use of limited satellite bandwidth and allocated frequencies, made the system efficiencies of satellite communications that connected much smaller antennas ever more attractive. Instead of 30 or even 10 m Earth stations, there were now two-way transmit and receive VSATs that were 3 m or smaller in diameter. As satellites became larger and more powerful, the system economics and evolving technology encouraged the building of even smaller micro-terminals which were also cost-efficient. Thus, there was a series of technological advances such as 3-axis stabilized satellites with higher gain antennas, more powerful satellites, the deployment of satellites in new higher frequency bands such as Ku-band and Ka-band, the conversion to digital satellite services, and onboard intelligence, switching, and processing. These advances not only allowed higher capacity satellites but also satellites capable of working to even smaller and more cost-effective ground antenna systems. The NASA program in the United States to develop the Advanced Communications Technology Satellite (ACTS) that demonstrated the use of Ka-band frequencies and onboard processing helped to move this process along during the late 1980s (Fig. 6.3).

The most remarkable aspect of the new technology made possible by experimental satellites such as the ACTS satellite in the United States, the ETS VI satellite in Japan, and the ARTES satellite in Europe was actually realized on the ground. These new satellites demonstrated very broadband capabilities could be accomplished to small and compact ground antenna systems. The ACTS ultrasmall aperture terminal (USAT) pictured below had only a 60-cm (about 2 ft) aperture yet could receive data rates of 45 Mbits/s with a lower upstream return rate of 1.5 Mbits/s (Fig. 6.4).

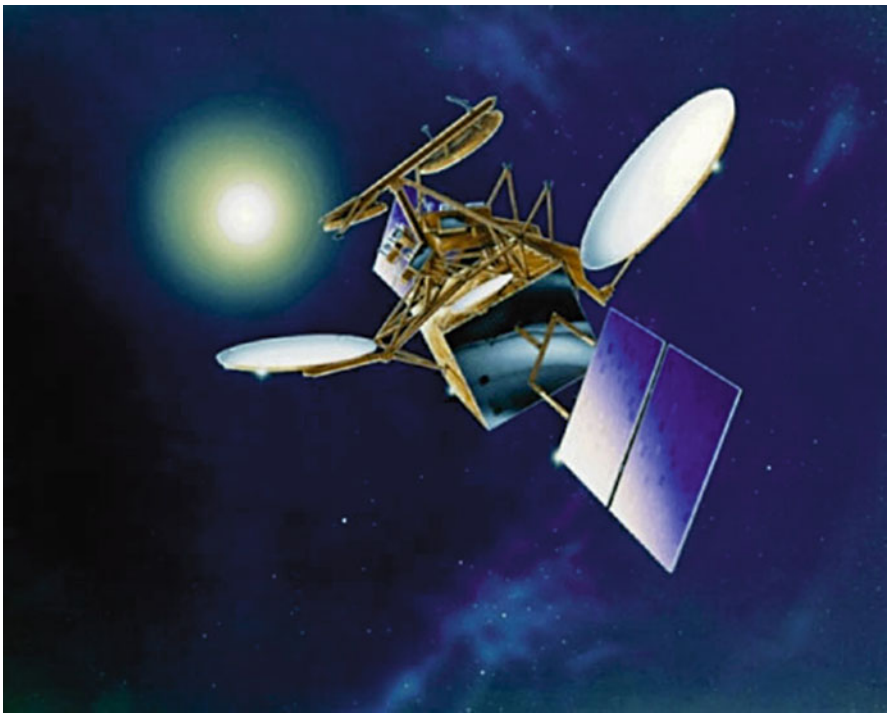


Fig. 6.3 The advanced communications technology satellite developed by NASA in the 1980s to promote new digital capabilities and Ka-band utilization (Graphics courtesy of NASA)

This research program hastened the conversion of the FSS industry to digital video broadcasting services. The digital broadband distribution function could send high-speed data to support television, voice or data services, which could be used to download new computer software, validate a credit card, or update a global corporation's inventory at thousands of outlet stores. In the age of the Internet, this has perhaps been the most significant stage in the evolutionary process for today's FSS digital networking services. The latest stage in this evolution has been the increasing shift by businesses and private users to employ Voice over IP (VoIP) services regardless of whether the data stream might be going over satellite, fiber, coax, or microwave relay.

The international standards to allow this digital broadcast service to be interactive with downstream rates have now fully evolved. This digital broadcast service is often in the 36–72 Mbits/s range downstream, with thinner stream response uplink rates that originate from 1 m micro-terminals. This shift to digital broadcast services have thus served to move FSS services closer and closer to end users. Large multinational enterprises with enterprise networks can thus use such digital broadcast satellite networks to connect efficiently with thousands of node points. For example, large oil companies can use these networks to link with all their service stations and automobile companies can link to all their dealerships. Global



Fig. 6.4 The ACTS smallest user terminal was only 60 cm in diameter (Graphics courtesy of NASA)

department stores, insurance companies, banks, and airlines can also connect with great flexibility to thousands of locations worldwide.

The two most predominant standards that allowed the development of this type of asymmetrical global satellite digital networking (i.e., a heavy stream of data out from corporate headquarters and thin route return data service) are known as: (1) Digital Video Broadcasting with Return Channel Service (DVB-RCS); and (2) Digital Over Cable System Interface Standard (DOSCIS). In the case of DOSCIS, this service was first developed for cable television networks on terrestrial systems, but then adapted to use on satellite networks as well.

This new type of digital broadband satellite service has truly allowed satellites to support global networks with thousands or even tens of thousands of nodes very cost-effectively.

The shift to large-scale digital networks via satellite has, however, presented a great challenge to the fixed satellite service (FSS) industry. The problem is that most large-scale digital networks today operate using the Internet Protocol. However, the original IP interface connections were established on the basis of terrestrial networks where the issue of satellite transmission delay and the IP Security (IP Sec) procedures did not take into account the particular characteristics of

satellite transmission. These two issues initially made it very difficult to use satellite-based digital networks using TCP/IP (Transmission Control Protocol/Internet Protocol) efficiently. Satellite transmission delay was mistakenly interpreted as network congestion and led to slow recovery procedures. In time, the clock for detecting congestion was reset to take into account satellite transmission times and the so-called spoofing methods compensated for geosynchronous satellite-related transmission times. Also the problem of IP Sec procedures, that stripped off key routing header information needed for effective satellite transmission, has also been largely rectified by the Internet Expert Task Force (IETF) Request For Comment (RFC) processes. The result is the now widely adopted Internet Protocol over Satellite (IPoS) transmission standard. Thus, today large-scale digital satellite networks using IP-based interface standards can operate with much higher efficiency and are typically within 80% of the efficiencies achievable on terrestrial networks (Kadowaki 2005).

Regulatory Shifts Concerning FSS Systems to Make Them Openly Competitive

The regulatory environment in which telecommunications and IT services are provided on a global basis has shifted dramatically since the 1980s where "liberalization" or competitive service offerings began to replace the so-called rate-based competition of monopoly carriers in the United States, Europe, and Japan and then around the world. The first step in this process began in the United States in the 1970s when the MCI Corporation challenged the monopoly status of AT&T in courts by claiming anticompetitive actions.

It was in the early 1970s during the Nixon administration that the US Justice Department opened an investigative process against both AT&T and IBM, charging that there was evidence of anticompetitive practices by both firms. In time, the proceeding against IBM was dropped but the action against AT&T continued. In fact, there were two different but related proceedings. There was the MCI suit against AT&T seeking damages for anticompetitive practices that was ultimately successful. And then there were the antitrust charges brought by the Justice Department which continued through the Ford and Carter administrations until the very waning days of the Carter administration. At that time, Federal Judge Harold Greene adjourned the proceedings on January 16, 1981 to let the Justice Department and AT&T to see if they could reach a final negotiated settlement. After months of negotiations that went many months into the Reagan Administration a negotiated final settlement was reached between the Justice Department and AT&T and approved by Judge Harold Greene.

Under the terms of this negotiated final agreement, planning was undertaken to begin the restructuring of AT&T with the divestiture of AT&T actually occurring as of January 1, 1984. Under this negotiated final agreement, the divested AT&T would continue its long distance and international services, but it would give up

ownership of its various local Bell Operating Companies, which were reconstituted as a series of new regional corporations. AT&T, as of 1984, faced competition for its telecommunications services for long distance and overseas services and it also faced competition in the design and installation of telecommunications facilities. In order for AT&T and its Bell regional operating companies to reliably interconnect the FCC established rules called "open network architecture" (ONA). This allowed these various systems in the United States to interconnect to common digital standards (MacAvoy and Robinson).

The negotiated agreement changed the entire regulatory structure for telecommunications in the United States. In the past situation, the Federal Communications Commission regulated AT&T by explicitly approving new facilities for telecommunications services that went into an "official rate base." This rate base of approved facilities allowed AT&T to make a certain amount of profit or rate of return on this "officially approved" investment. Critics of this arrangement included those who were heavy users of telecommunications such as banks, insurance companies, airlines, etc. They argued that this "rate base" system for regulating monopolies created the wrong incentives and that it led to wasteful investment in unnecessary telecommunications facilities (switches, microwave relays, coaxial cables, satellites, Earth stations, etc.) and thus stymied innovation and cost efficiency. Under the new FCC regulatory regime, US telecommunications providers were given incentives to make higher profits if they could lower investment costs and lower their prices to business and consumers.

In Europe, the newly formed European Union was beginning to wrestle with a different but somewhat parallel problem. Its objective was to create an integrated telecommunications system that could allow all of the networks within Europe to be compatible with one another and connect seamlessly as if it were one system. The concept for digital communications under development at that time, called Integrated Services Digital Network (ISDN), allowed largely compatible digital networking and served to provide part of the solution. The major breakthrough was to adopt what they called "Open Network Provisioning" (ONP). The bottom line in Europe was that ONP not only allowed national networks to interconnect seamlessly, but it set the stage for national telecommunications to start competitive telecommunications networks in neighboring countries.

In Japan, there was also interest in the competitive approach to regulation of telecommunications and they sent observers to the United States to monitor the divestiture of AT&T. The result was that the Diet passed two new telecommunications laws – one dealing with domestic telecommunications and one dealing with international telecommunications. These laws authorized competition for telecommunications services in Japan but restricted ownership of competitive networks to Japanese-owned entities. Thus from 1984 through 1992 there was a major shift in many of the so-called developed economies to "liberalize" telecommunications regulation and create a regulatory process under various types of open network standards to allow the efficient interconnection of competitive networks.

The situation for satellite communications was complicated in that the Intelsat Agreements that acted very much like a treaty among all member countries and

territories (almost 200 in number) specified that there should be a single global satellite network with a mission to provide services at low cost to developing nations. These Intelsat Agreements had been set up under US initiatives starting from the Kennedy Presidential administration. The United States was caught in a difficult situation. The single Intelsat Global Satellite System had been the brainchild of the United States and the Communications Satellite Corporation (Comsat) that had been created by the 1962 Communications Satellite Act by the US Congress. The United States was the predominant member and owner of the Intelsat system and from 1965 to 1975 Comsat had been the system manager.[2]

In 1983, several filings were made to the FCC to build and deploy new satellite systems that would provide international links in competition with Intelsat. This left US policy makers caught up in a dilemma. Article XIV of the Intelsat Agreement specified that any member country of Intelsat that wished to deploy and operate a separate satellite system must technically coordinate it with Intelsat and if it wished to carry international traffic then it must "economically coordinate" with Intelsat under Article XIV(d) of the Agreements to show that such removal of international traffic did not create "economic harm" to Intelsat.

This economic coordination was successfully carried out by the "Eutelsat" organization for regional traffic essentially within Europe and involved traffic that Intelsat was for the most part not carrying. The Reagan Administration favored competition and believed that competitive satellite systems would serve to reduce prices to businesses and consumers. It nevertheless proceeded slowly. It authorized several competitive systems to proceed, but on the basis that the competition would only be to serve large corporations on dedicated "enterprise networks" and not to be competitive for publicly switched telephone traffic. In time, the emergence of regional satellite networks such as Eutelsat, Arabsat, and proposals for an Africasat that proved to be economically viable, as well as a growing number of domestic satellites, created a groundswell of opinion within governments around the world to abandon the concept of monopoly satellite systems owned by governments. There were meetings of the Intelsat Assembly of Parties that allowed the Agreements that had been negotiated originally in 1963–1965 and adopted in definitive form in 1983–1986 to be abandoned with the result that Intelsat was "privatized" and part of the monopoly system spun off as the New Skies organization of The Hague, The Netherlands. This shift to "privatize" Intelsat and take away its intergovernmental status affected not only Intelsat. The Inmarsat organization for mobile satellite communications and Eutelsat for European and other international services proceeded toward privatization as well. In fact, Inmarsat, of London, United Kingdom was the first to complete the privatization process. Today everything concerning Inmarsat has been "privatized" except for a small unit to assist with public safety for ships and aircraft and a unit to provide assistance for developing countries to obtain satellite services (GAO Telecommunications 2004).

There are several ironic results with regard to the global privatization process for satellite services and the opening of international satellite services to competition. The prime competitor to Intelsat in the 1980s was the so-called Panamsat organization. In the aftermath of privatization, Intelsat has now totally acquired Panamsat

through merger arrangements. Thus, the competitor that played a prime role in forcing the privatization of Intelsat has essentially disappeared while Intelsat is as large as ever with ownership and investment in some 80 satellites and is earning the largest amount of revenues ever in its history. The entity named New Skies that was spun off to compete with Intelsat has been acquired by the SES Global organization of Luxembourg and thus has also essentially disappeared as well. Privatization, followed by a number of acquisitions and mergers in the past decade, has seen the reemergence of just a handful of dominant carriers.

The good news for consumers is that this global competitive process has largely seemed to accomplish the goal of lowering the cost of television, data, and voice services via satellite. The price of international connections via both fiber-optic cable and communications satellites are at an all time low. The very largest carriers, namely Intelsat, SES Global, and Eutelsat, have also tended to form alliances with regional carriers in many instances (Pelton 2005). The Appendices to this Handbook indicate the various international and regional systems and the many alliances and partnerships that now exist around the world in the field of satellite communications.

Evolution of FSS Markets from Global Networks to Regional and Domestic Satellite Systems

As described earlier, the first major FSS system was the Intelsat global network that was established to provide international connectivity in 1964 with the first satellite going up in 1965. Intelsat first provided connectivity across the Atlantic Ocean and then followed with connections across the Pacific Ocean. Global connectivity across all three major oceans was not established by Intelsat until 1969, just before the Moon landing.

The success of these early international satellite services stimulated all other uses. The enthusiasm to employ satellites for regional and domestic national services thus also grew apace. By the early 1970s, Intelsat began to lease capacity to countries for domestic services. Even in the late 1960s, dedicated national satellite systems began to be deployed. The Soviet Union and Canada led the way and then the United States adopted an "open skies" policy. This new policy adopted during the Nixon Presidency urged the development of national satellite systems. Shortly thereafter, multiple national satellite networks began to emerge in the United States for fixed satellite and especially for satellite television distribution services. In time, other nations and regions allowed multiple satellite systems to be financed and deployed as well even though the US market remains the most dynamic in this respect.

The United States shifted quickly toward more competitive telecommunications markets and the so-called liberalization process also ensued in Europe, Japan, and elsewhere around the world, particularly within the OECD. This process has continued under a competitive process backed by the World Trade Organization

(as discussed in the previous section) and these factors all served to spawn more and more satellite systems at the international, regional, and national levels. This openly competitive process, however, has also led to consolidation. Mergers, competitive failures, and/or outright acquisitions of other satellite systems have also served to narrow the range of competition. Today, there appears to be a narrowing range of global networks as Intelsat has acquired its chief competitor Panamsat and SES Global has acquired New Skies and bought an interest in many other regional systems. Today Intelsat, SES Global, and Eutelsat are the most prominent globally ranging systems, although there are also many vibrant regional systems and of course an even larger number of domestic systems.

In the appendices to this Handbook, there is an extensive listing of the various national, regional, and international communications satellite systems that exist around the world today. The shift in FSS markets in the past nearly 50 years have been dramatic in terms of the range and volume of services. The Early Bird or Intelsat I satellite that started commercial satellite services had but 240 voice circuits of capacity using analogue technology and had both very low power and low gain antenna. Today's satellites using digital technology and deploying as many as 100 transponders (such as on the Intelsat 8 satellite) can have the capacity of millions of voice circuits or over a thousand video channels. The remarkable thing is that not only do the satellites now have tremendously larger throughput capacities, but the ground antennas are no longer huge, multiton structures, but can be only 1 m or less dishes. Despite their small size these dishes – thanks to digital video broadcast standards – can still support fast, broadband data rates. As of 2012, upward of 18,000 video channels are available via FSS networks for television distribution around the world on a 24 h a day and 7 days a week basis. These FSS networks can be used in very flexible ways to support corporate enterprise networks, data networking, and multi-casting, as well as a flow of traffic that can dynamically shift from voice, data, audio, video, or videoconferencing depending on consumer demand.

Conclusion

The FSS satellite systems that started the satellite communications in the mid-1960s nearly 50 years ago were the "grand-daddies" of the satellite industry. As the technology matured and the range of services that satellite could deliver expanded, new types of satellite services were developed and systems were adapted to this growing market in a diversity of ways. Today, FSS has spun off direct broadcast satellite systems (known in ITU as BSS networks), mobile satellite systems (known as MSS networks), store and forward data relay (or machine-to-machine networks), and specialized defense and strategically oriented satellite networks. These latter two types of satellite networks actually use different spectrum bands. Even within the mainstream FSS services there are global, regional, domestic networks, and even within these there are networks that specialize in data networking, video distribution, or emphasize highly connective "mesh networks" versus those that use a star (or hub and spoke) architecture. This market specialization tends to affect the technical design of the satellites, the user

antennas and terminals, and the interface standards. These specializations can at times complicate the ease and quality of interconnectivity with terrestrial or even other satellite networks. The evolution of IP-based standards, however, continues to serve as the key "glue" that allows all forms of global communications and IT systems to connect together as seamlessly as possible.

The long-term progress made in satellite communications seems likely to continue, but there remain key challenges for the future. The challenges that are discussed throughout the chapters of the Handbook and that consider satellite communications and related spacecraft and launcher needs include:

- Expanding or at least preserving satellite communications spectrum allocations, and the need for effective migration to the use of higher frequencies in the millimeter wave band in overcoming precipitation attenuation issues in these new bands.
- Access to adequate GEO orbital positions and minimizing intersystem interference.
- Coping with the problem of orbital debris.
- Technical standards to achieve seamless connectivity between FSS and terrestrial networks and even other types of satellite networks – especially related to completely fluid IP interfacing.
- Coping with the issue of satellites constantly changing role as a complement to terrestrial networks, as a potential restorer of terrestrial networks, and at times a direct competitor. (The satellite use of CDMA and TDMA multiplexing vs fiber-optic systems using DWDM creates an ongoing compatibility issue beyond that of satellite latency and IP Sec-related disruptions.)
- Developing improved satellites, lower cost and more compact ground antenna and lower cost launch systems to keep the cost of satellite networking moving to even more competitive levels.

The remarkable growth of computer and IT systems and fiber-optic networks worldwide has been so dramatic that they have at times overshadowed the rapid expansion of satellite technologies and markets. Few industries in the history of humankind have expanded more than a 1,000-fold in less than a half century, but the satellite industry in general and the FSS networks around the world have exceeded even this rate of expansion. Now, something approaching 20,000 satellite television channels have replaced the single low-quality black-and-white television channel that Intelsat I was able to achieve in 1965. Instead of satellites with hundreds or thousands of equivalent voice circuits, there are today satellites equivalent of millions of voice circuits. Just one of these massive satellite networks could transmit the equivalent of the Encyclopedia Britannica in a few seconds and the equivalent of the Library of Congress in a matter of hours. New capabilities such as inter-satellite links, onboard processing, active rain attenuation response capabilities, extremely high-gain multi-beam antennas, exploitation of additional spectrum in the Ka-band frequency bands, and new digital interface standards will allow satellites to improve their performance to even higher levels during the twenty-first century to keep pace with new user and institutional demand for communications and IT services.

Cross-References

- ▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends
- ▶ Satellite Communications Video Markets: Dynamics and Trends

Notes

1. Op cit, Chatrand, pp. 9–20
2. Op cit, Pelton and Alper

References

M.R. Chartrand, *Satellite Communications for the Nonspecialist* (SPIE, Bellingham, 2004), pp. 27–42

Interview with Santiago Astrain, September 1974

GAO Telecommunications, GAO telecommunications report to congressional requesters: Intelsat privatization and the implementation of the ORBIT Act. US GAO 04–891. (GAO Telecommunications, Washington, DC, 2004). www.gao.gov/new.items/d04891.pdf

N. Kadowaki, Internet and the new broadband satellite capabilities, in *Satellite Communications in the 21st Century: Trends and Technologies*, ed. by T. Iida, J.N. Pelton, E. Ashford (AIAA, Reston, 2005), pp. 62–72

G.E. Lewis, *Communications Services via Satellite* (BSP Professional Books, London, 1988), pp. 82–83

P. MacAvoy, K. Robinson, Winning by losing: the AT&T settlement and its impact on telecommunications. Yale J. Regul. 1, 1–42 (1983/1984). http://heinonlinebackup.com/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/

N. Pelton, *Future Trends in Satellite Communications Markets and Services* (IEC, Chicago, 2005), pp. 1–20. and 109ff

J.N. Pelton, *Basics of Satellite Communications* (IEC, Chicago, 2006), pp. 41–72

J.N. Pelton, J. Alper (eds.), *The INTELSAT Global Satellite System* (AIAA, New York, 1986)

Satellite Communications Video Markets: Dynamics and Trends

7

Peter Marshall

Contents

Introduction 137
The Early History of Satellite Television 137
The Evolution of Global Satellite Television Regulation and Tariffs 141
The Need to Understand the Key Concepts of Television "Contribution," Television "Distribution," and Television "Direct Broadcast" 146
The Special Role of CNN in Global Satellite Television Development 149
Daily News by Satellite 150
How Television via Satellites Influenced Global Politics 152
Sports Programming on Communication Satellites 153
The Growth of "Direct-to-Home" Satellite Television 154
Satellite Radio Broadcasting 156
High-Definition Television (HDTV) via Satellite 157
Future Trends 158
Notes 160
References 161

Abstract

The advent of satellite communications brought a new era to the TV industry. In the early years, however, these transmissions were quite costly and limited by the modest capacity of the first commercial communications satellites. The evolution of satellite technology and development of satellite aggregators such as Brightstar, World Communications, Bonneville, IDB, Keystone, and Globecast, allowed costs of satellite television to decrease sharply. The development of full-time, annualized satellite transponder charges – as opposed to per minute fees – were also critical to the development of much lower satellite television fees.

P. Marshall
Royal Television Society, England, UK
e-mail: pmsatellites@aol.com

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_85, © Springer Science+Business Media New York 2013

The evolution of digital television services was another important breakthrough. Instead of the one or two television channels per transponder with analog systems it became possible to derive up to 18 channels per transponder. Digital transmission speeded up the evolution of domestic television satellite systems and played a key role in the growth of direct-to-home satellite broadcasting.

The development of satellite based video systems was not seamless and encountered periods of major market difficulties. One of the most prominent market development issues was the failure of early direct broadcast satellite systems (or BSS in the terminology of the ITU).

However, DBS (or BSS) satellite markets are now well established in international, regional, and domestic markets. They are not only highly successful but represent by far the largest single satellite market in terms of revenues. Most recently there has been a rapid growth of high-definition television (HDTV) service via satellite.

These satellite services compete with coaxial cable and fiber-optic-based CATV systems. The economics of satellite television are quite different from terrestrial networks because once a direct broadcast satellite system is launched and operational there is very little incremental cost beyond the consumer terminal needed to receive service. The advent of new digital interface standards known as digital video broadcast with return channel service (DVB-RCS) and Digital Over Cable Service Interface Standard (DOCSIS) have allowed the rapid development of digital television over satellite and cable systems. DOCSIS is now widely used for both satellite and cable television systems. These digital standards, together with high-power fixed satellite systems and broadcast satellite systems – that by definition have high power – allows not only the distribution of a large number of video channels to consumers but also low-cost distribution of high-speed digital data service to both home consumers and businesses. These digital satellite video systems can be – and indeed are – used to provide broader band digital services to the "edge" of digital networks at low cost. Thus DBS and higher powered FSS satellite systems are now being used to provide commercial broadband data services to business as well as broader band digital services to remotely located consumers.

Keywords

3DTV (three-dimensional television) • American advanced television systems committee (ATSC) • Broadcast satellite services (BSS) • Broadcast satellite services for radio (BSSR) • Cable news network (CNN) • Direct broadcast service (DBS) • Direct to home (DTH) • Federal communications commission (FCC) • Fixed satellite services (FSS) • High-definition television (HDTV) • Intelsat • Olympic Games • Relay experimental satellite • Society of motion picture and television engineers (SMPTE) • Syncom experimental satellites • Telstar experimental satellite • World cup soccer

Introduction

The idea of using artificial satellites for broadcasting services was central to the original concept of why a telecommunications satellite system might be deployed in the first place. Today the prime usage of communications satellites is for video services. This can be for video and audio distribution to support CATV and national television networks (i.e., a so-called fixed satellite service application) or to support direct broadcasting services (a so-called broadcasting satellite service) that is provided directly to home consumers or business users. This prime application is reflected in terms of the delivery via satellite of well over 15,000 satellite channels around the world. This predominance of the satellite video market is even more apparent in terms of total revenues derived from satellite television and audio services, which exceed 70% of all satellite communications–related income.

Television and audio distribution and broadcasting services have grown steadily from the outset of commercial satellite communications offerings in 1965 and have diversified in their nature in the past few decades. Although the advent of fiber-optic cable services have tempered the growth of fixed satellite services for telephony and data services, satellite usage for audio and television has continued to expand around the world. This, in large part, is due to the fact that satellite networks (particularly those in geosynchronous orbit) have the ability to cover such very broad areas and thus allow an extremely cost-effective one-to-many service.

The advent of digital television, high-definition television, and the spread of television across the world have strengthened the satellite television market. The innovative new applications related to broadcasting satellite service for radio (BSSR) and digital audio broadcasting services (DABS) have created a new type of satellite market that not only distributes high quality and diverse programming for radio, but has evolved a range of two-way applications, such as those related to antitheft services and communications with emergency services in cases from natural disasters to traffic accidents.

The Early History of Satellite Television

The history of satellite communications and the role that Sir Arthur Clarke played in that history were addressed in an earlier chapter of the Handbook. However, there are some key elements of that history that particularly apply to satellite broadcasting services that are important to recall here.

It was Arthur C. Clarke who explained the potential for using satellites for broadcasting purposes first and did it most clearly. This “first” explanation was provided in the British journal *Wireless World* in February 1945, even prior to his better-known October 1945 article. This famous article in the fall of 1945, explained the technical aspects and potential uses of satellites in geosynchronous

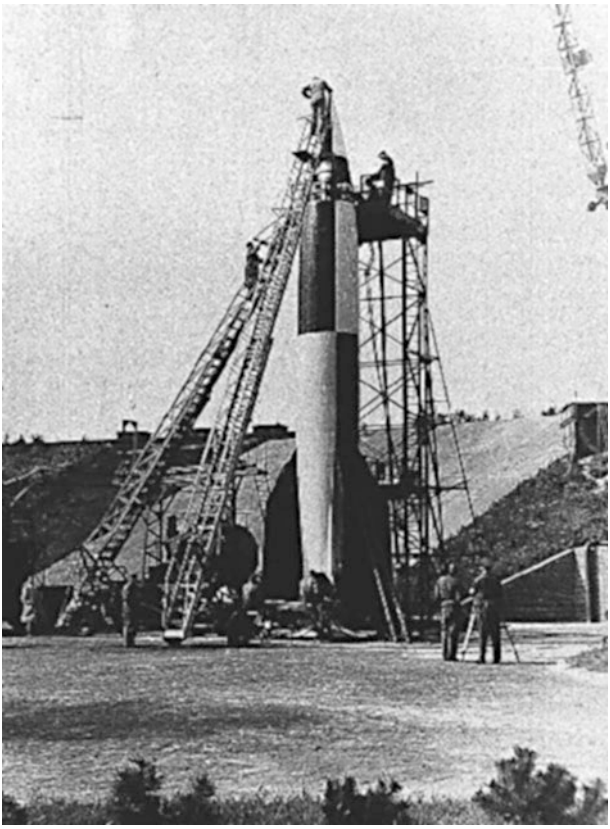


Fig. 7.1 A German V2 rocket being prepared for launch (photo courtesy of NASA)

orbit. But before he presented the technical explanation, he indicated the most powerful application – the "why" of launching an artificial satellite. In a letter to the editor entitled "Peacetime Uses for V2" he wrote:

> *An "artificial satellite" at the correct distance from the earth would make one revolution every 24 hours; i.e., it would remain stationary above the same spot and would be within optical range of nearly half the earth's surface. Three repeater stations 120 degrees apart could give television and microwave coverage to the entire planet. I'm afraid this isn't going to be of the slightest use to our post-war planners but I think it is the ultimate solution to the problem* (Clarke 1945a).

The V2 referred to in the headline was the German rocket-propelled launcher developed by Werner von Braun and his team of scientists. Hundreds of these weapons were launched against the UK in 1944 and 1945 as instruments of war, but Arthur C. Clarke was the first to present clear technical ideas as to how such systems could be instruments of peace (Fig. 7.1).

It was through Clarke's writings that the thought of "broadcast towers in the sky" became a reality. In less than two decades the concept went from being science fiction to becoming scientific reality. By the mid-1960s, distribution of global television news "live via satellite" would become accomplished fact (Clarke 1945b).

The Telstar and Relay satellites in 1962 demonstrated that it was possible to send and receive radio signals from space and indeed to transmit live television pictures via this new telecommunications media. In contrast, the capacities of submarine cables of the 1960s were too limited in transmission capacity to send "live television" programming. Slowing down the transmission of a television signal and then recreating it as a "full motion transmission" at the other end was too expensive for commercial use (Pelton 1974).

The experimental Telestar and Relay satellites, however, were in a low earth orbit. This meant that they were "visible" at ground earth stations for only an 18–19 min window out of every 90-min orbit. Nevertheless great excitement was created by the first television signals to be sent and received between the AT&T earth station in Andover, Maine, and the British Telecom earth station at Goonhilly Downs in the south west tip of England.

The successful 1963 launch of the first geosynchronous telecommunications satellite, the Syncom 2, after the launch failure of Syncom 1, created even more excitement. This satellite raised the solid prospect of continuous satellite transmission over an ocean. The three Syncom satellites were designed and built by Hughes Aircraft Company and launched by NASA (Pelton 1974, p. 48).

Television engineers were eager to experiment with the relay of television via the Syncom 2 satellite that was continuously available 24 h a day. These first experiments were conducted with the cooperation of the Columbia Broadcasting System (CBS) in the USA and the British Broadcasting Corporation (BBC) in the UK. This experiment consisted of the BBC anchorman Richard Dimble by introducing grainy and flickering black-and-white pictures of familiar scenes from the USA, including Mount Rushmore, the Mormon Tabernacle Choir from Salt Lake City, and Wall Street, New York (Fig. 7.2).

By October, 1964, Syncom 3 was in orbit above the Pacific Ocean and the Japanese government learned that it would be in a position to transmit live coverage of the Tokyo Olympic Games to the USA. The Japanese government lobbied for extensive live broadcasting to American viewers. However, NBC Sports had already acquired the exclusive rights from NHK, the host broadcaster in Japan, and were not prepared to change their program plans. They had already sold the commercial advertising time on the assumption that its coverage would be produced on videotape and flown to Seattle for transmission over landlines for delayed distribution across the nation. After considerable pressure from the US government, NBC agreed to a compromise for the opening ceremonies to be transmitted live by satellite. And so, at 1:00 am New York time on a Saturday morning, American viewers saw live black-and-white coverage of Japan's Emperor Hirohito and Empress Nagako presiding over the event and the parades of participants from the Meiji Olympic Stadium (Marshall and Wold 2004).

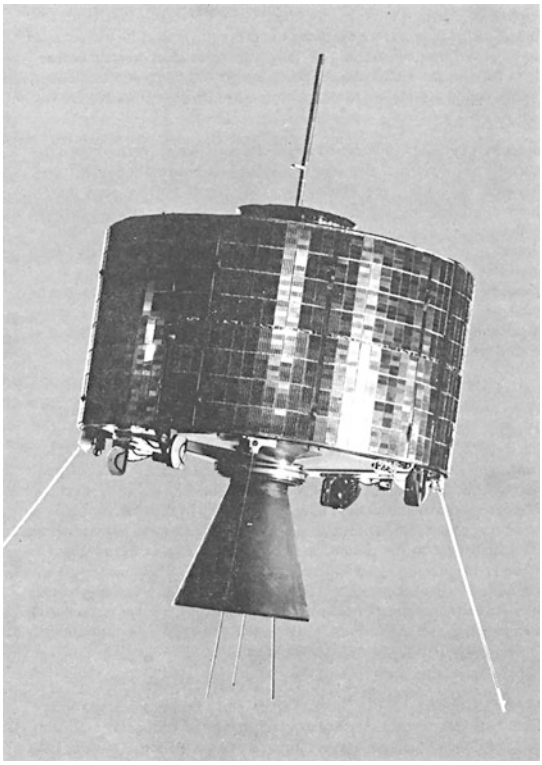


Fig. 7.2 The Syncom-2 satellite (courtesy of NASA)

The New York Times TV critic, Jack Gould, wrote: "Live television coverage of this morning's opening ceremony of the Olympic Games in Tokyo was of superlative quality, a triumph of electronic technology that was almost breathtaking in its implications for global communications" (New York Times 1964).

Although NBC's viewers were denied any further live coverage of the Olympic events, the opportunity was not entirely lost because Canadian and Mexican broadcasters arranged to receive the satellite signals for some of the events in California and then re-transmit them across the borders North and South. And so, although the outcome was limited by NBC's commercial pressures, this was still a historic occasion. The world had suddenly shrunk and the era of satellite television broadcasting had truly begun.

In the following year, on April 6, 1965, the world's first commercial communications satellite, Early Bird was launched into geosynchronous orbit and placed into commercial service. This satellite, manufactured by the Hughes Aircraft Company under contract, was an upgraded version of the first geosynchronous Syncom satellites. The Early Bird satellite (officially known as Intelsat I (F-1)) could transmit 240 voice circuits or one low-quality black-and-white TV channel.

However, it was not able to transmit television services and telephone or data traffic at the same time, since even a low-quality television signal commanded the full transmission capacity of the satellite.

The Intelsat I (F-1) remained in continuous, full-time service for nearly 4 years. During this period, there were occasional opportunities for the major broadcasters on both sides of the Atlantic to book time on the satellite for special events such as the first launches of NASA's series of Apollo spacecraft. The Intelsat I (F-1) and the Intelsat II satellites, launched largely to support NASA launch requirements associated with the Gemini program, were too limited in capacity to support a high-quality television transmission.

Then in 1968 and 1969 came the Intelsat III series of satellites. These satellites, with a much higher gain antenna and more power, had a capacity of two high-quality color television channels, plus 1,200 voice circuits. The Intelsat III series got off to a difficult start in that the "spinning antenna" that allowed continuous transmission toward the Earth initially "froze" and would not spin. The next Intelsat III satellite was planned to become known as "Olympico" and was slated to provide live coverage for the Mexico Olympics, but it turned out to be a launch failure. Fortunately the following series of launches were successful. The redesigned bearing and power transfer assembly solved the antenna "freezing" problem and spun around smoothly to allow the "despun" antenna to constantly point toward Earth. Thus eventually these Intelsat III satellites provided coverage for the Atlantic Ocean and the Pacific Ocean and finally the Indian Ocean regions, thereby completing the implementation of a truly global system in June 1969.

It was this network of Intelsat III satellites that enabled a worldwide audience of over 500 million people to see the Moon landing of Apollo 11 on the lunar surface and later to see the first space walk by astronaut Neil Armstrong in July 1969. The signal was sent from the Lunar Excursion Module (LEM) to a radio astronomy telescope in Australia. The signal was then transmitted to the Australian Intelsat Earth Station and from there the signal was relayed around the world.

Events such as the Moon Landing and the 1968 Mexico City Olympics created an ongoing demand for live satellite television coverage. But in spite of these momentous occasions, it was still another 20 years or more before the full potential of satellite broadcasting began to be realized and daily television live via satellite relays became truly routine. Years of technological progress and especially digital television transmission were necessary for satellite television to reach its potential. But in the 1970s and 1980s there were other issues in the political, regulatory, and economic arena that certainly served to slow the growth of satellite television – at the national, regional, and global level.

The Evolution of Global Satellite Television Regulation and Tariffs

Political, regulatory, and cost factors limited the growth of satellite television in the 1970s and 1980s with extremely high tariffs serving to act as a significant brake on

live global television coverage. Ultimately, it was the spread of competitive telecommunications services in many parts of the world in the 1980s that stimulated satellite television growth on a global basis.

During the first two decades of commercial satellite television distribution around the world, from the mid-1960s to the mid-1980s, it was only the most well-capitalized TV networks of the USA, Canada, Europe, Japan, and Australia which had the financial strength and high production value budgets to be able to afford the use of satellite transmission facilities. But even the most affluent network broadcasting companies around the world, such as ABC, CBS, NBC, BBC, FR3, RAI, NHK, CBC, etc., were limited in their ability to use satellite transmission except for special, high-budget occasions such as the Olympic Games, World Cup soccer tournaments, and news events of global importance such as armed conflicts, royal weddings, or US presidential elections.

A prime barrier to the expanded use of satellites by the television industry was the high tariffs imposed on the use of Intelsat's capacity which restricted the growth of international transmissions. At that time, video service was "sold" on a per minute tariff basis with a 10 min minimum. Even more importantly, the structure of global telecommunications was set up on the basis of each country having a single monopoly telecommunications provider. Intelsat was owned by these telecommunications monopolies (or signatories) who then bought Intelsat's services at what was to them a "wholesale rate." They each calculated the cost of their earth station operations and what seemed to them a reasonable profit and then charged their broadcasting organizations a retail rate that was equivalent to thousands of dollars an hour for satellite television coverage. This rate could, of course, be significantly higher in countries with limited satellite traffic, and thus the broadcasters ended up facing a very large bill for anything other than a very short transmission.

The signatory tariffs on top of the Intelsat tariffs were thus the product of an international government-owned monopoly structure where there was no great incentive to lower television broadcasting rates that kept pace with technological innovation. This was also partially an artifact of the "focus" of the Intelsat organization and its signatories. Intelsat was originally created to improve global telephony and telecommunications services and most signatories were ministries of posts and telecommunications. Thus to most of them, global television or radio transmissions were seen as not only a "sideline" but also a way to help lower the cost of international voice and data services (or in effect to subsidize what was seen by these telecommunications entities as the most vital service) (Fig. 7.3).

The total tariffs imposed on the global broadcasting industry were therefore a combination of the short-term (per minute) rates charged by Intelsat to the national carriers (usually government-owned ministries) for the use of satellite bandwidth, plus the 30%, or sometimes even greater, markup added by the national carrier, and then additional charges these same organizations billed for using their terrestrial earth stations and the landlines to the studios of the broadcasters.

Also, there were certain governments around the world who saw potential threats to national sovereignty or security in the use of the Intelsat system for broadcasting



Fig. 7.3 The INTELSAT building in Washington DC (Photo courtesy of Intelsat)

to the rest of the world television news coverage concerning local events. Countries with authoritarian rule and with closely managed news saw the free exchange of satellite news stories by a "free world press" as being against their own best interests. Such governments were therefore not displeased to see high tariffs acting as a barrier for international satellite television.

Despite persistent lobbying over many years by the broadcasters and their regional organizations such as the European Broadcasting Union (EBU), the North American Network Broadcasters Association (NANBA), and the Asian Broadcasting Union (ABU) tariffs remained quite high and discounted rates for volume usage were not available.

Most of the Intelsat signatories were reluctant to recognize the simple economic fact that lower tariffs would release a pent-up demand for satellite transmission services. In short, it would eventually be demonstrated that when prices were cut, there was price "sensitivity" associated with international television utilization so that when prices finally were reduced overall revenues dramatically increased.

Change began to occur in the 1970s. At the urging of the Nixon White House, a new US "open skies" policy was put into place under the auspices of the Federal Communications Commission. This policy opened up the design, manufacture, and deployment of US domestic satellite communications systems to competitive

applications. This action was to set the stage for competitive international satellite systems, but not until the 1980s.

Meanwhile, there was the launch of the Intelsat IV satellites in the early 1970s, each with a capacity of 4,000 telephone circuits plus two color television channels. This greatly expanded satellite capacity suggested the need for change in Intelsat's charging policies. The officials within the Communications Satellite Corporation (COMSAT), which served as the manager for Intelsat, set up an Intelsat IV Charging Policy Group. This group explored how satellite services might be sold in new ways. The emergence of new digital technologies in the 1970s further stimulated thoughts about new ways to sell satellite services. In the Intelsat IV era the Intelsat Board agreed, among other innovations, that Intelsat would sell spare capacity to signatories who wished to lease satellite transponders to establish domestic telephone, data, and television distribution services. In the early 1970s, Algeria became the first country to lease such capacity for this purpose. They used the capacity during the day for voice and data services, but from 5:00 pm they switched over to distributing television programming to regional capitals across the country. In a few weeks the centuries-old tradition that the bazaars closed at sunset changed with the shops closing at 5:00 pm because everyone went home to watch television.

It was with this first Intelsat transponder lease of capacity to Algeria that the idea was born of leasing capacity on a bulk basis, rather than only selling services on a per minute or per voice circuit charge. Dozens of other Intelsat member countries followed suit to sign domestic leases for telecommunications and television services. This was later followed by agreement by the Intelsat Board to lease capacity to the Australian television broadcaster Kerry Packer to provide regional television distribution in the Pacific region.

The true breaking of the logjam created by high per minute rates for television came with the move to end monopoly telecommunications services and to introduce competitive opportunities in the USA, Europe, and Japan. The year 1983 proved to be pivotal.

In the USA, a US Justice Department suit was brought against the AT&T Corporation regarding its "near monopoly" role and various alleged abuses against new telecommunications competitors such as the MCI Corporation. Judge Harold Greene "settled the suit" by obtaining consent from AT&T to his final rule making and order. This order led to the divestiture of AT&T into various parts that could broaden telecommunications competition. This Federal court ruling, that provided new levels of competition at all levels of telecommunications throughout the USA, also served to provide a surge of filings for new domestic satellite systems under the "open skies" policy. This in turn also hastened the development of satellite distribution to cable networks and further set the stage for competition.

Finally, in 1983, in a move that was a surprise to the world telecommunications community, a newly organized company in the USA named Orion filed with the FCC for permission to launch and deploy a satellite system. This was not for domestic service, but instead this new entity proposed to own and operate an international satellite system that would compete with Intelsat for international

telephone, data, and television services. In a short period of time other new companies filed to do the same. Panamsat sought permission to compete directly with Intelsat on international routes and indicated that they could provide services at lower prices.

The idea of other international satellite systems coexisting with Intelsat had of course already been established. Since 1971, the Intersputnik system had maintained services for Soviet Bloc countries. Next there had been the creation of Inmarsat in 1979 to provide international maritime and aeronautical satellite services. Further, the creation in 1977 of Eutelsat, an intergovernmental organization in Europe to provide regional satellite services, clearly had set the stage for this additional step toward international competition from the private sector. Despite these precursor steps, the direct attempts to bring commercial competition to the "single global satellite network" were entirely unexpected by Intelsat officials since the USA had heretofore been the organization's strongest supporter and President Kennedy and his administration had championed the idea of a single global system operated for the good of all. Even during the Nixon Administration, when the domestic "open skies" policy had been announced, the US government had supported the idea of a continued single global system during the negotiation of the permanent Intelsat Arrangements in 1969.

What followed these 1983 filings with the FCC by Orion and Panamsat for separate international systems was years of careful international consultations. In these consultations the US government, and the Reagan Administration in particular, continued to support the Intelsat Organization's standing intersystem coordination provisions as contained in Article XIV of the Intelsat Agreement. Yet, the Reagan Administration also quietly worked to find a way to promote international telecommunications competition. From 1980 to 1988 the Reagan presidency thus encouraged telecommunications competition. It actively sought a process whereby the Intelsat Agreements could be amended to allow competition. It took more than a decade and many Congressional representatives pushing for competition, but ultimately Intelsat was "privatized" and capitalized as another commercial entity that would compete for its traffic like any other commercial entity. No longer would this new entity be an International Organization, organized under an international treaty, but simply would compete on the basis of the strength of its technology and commercial offerings. This was, in fact, a sea change in the way international satellite services were to be provided. Inmarsat, located in London, UK led the way and was the first of the "monopoly" international satellite carriers to become "privatized". It quickly converted to become a commercial carrier owned by private equity firms. Eutelsat, located in Paris, France was also privatized and became owned by shareholders rather than by governmental ministries.

Internationally, banking, financial agencies, insurance companies, airlines, and other major users of global telecommunications services actively supported this process. With the privatization of Intelsat, Inmarsat, and Eutelsat, multinational enterprises and businesses were ultimately able to realize substantial savings on their telecommunications bills. TV news agencies and media networks who actively used satellites to service their many clients around the world, not

surprisingly worked in concert with this "liberalization process" to open up competition and bring down the cost of international satellite television services along with the lower cost of global telecommunications.

The difficult aspect of this history to assess is what impact technology made on this process, beyond the regulatory and policy shifts that allowed competition for international satellite services. On one hand new fiber-optic submarine cables began to be deployed in the 1980s. Certainly this new form of technological competition, would certainly have served to drive down the cost of satellite communications services without the regulatory and policy changes that led to the privatization of Intelsat, Inmarsat, and Eutelsat. Further digital satellite communications also began to be introduced in the 1980s and into the 1990s. In the analog era, only one or possibly two television channels could be derived from a 36 MHz transponder. But by the 1990s up to 18 television channels could be derived from a 72 MHz transponder. In short, almost ten times more television channels could be transmitted via a single satellite transponder as a result of the transition from analog to digital technology. This transition to digital technology, and its breakthrough efficiencies, clearly drove down costs as well. The latest in digital compression technologies, as represented by the Motion Picture Expert Group (MPEG) standards such as MPEG 2, MPEG 4, and MPEG 6, enabled the most efficient throughput of video channels, through fiber-optic cables, coaxial cables, or satellite transmissions. These have certainly served to boost the growth of video services around the world and helped to greatly reduce the price of television distribution and television broadcast.

The Need to Understand the Key Concepts of Television "Contribution," Television "Distribution," and Television "Direct Broadcast"

To help understand the evolution of satellite broadcasting, it is useful to explain at this stage the three separate and distinct terms used in the field of satellite communications – "television contribution," "television distribution," and "television broadcast".

Contribution is the use of a satellite link to transmit TV coverage from where an event is occurring to the broadcasting center or news agency for the television program production. This is usually a point-to-point, unilateral transmission with the technical arrangements being made by the broadcaster or agency concerned in order to get the original television signal to their production center. This process is often referred to as "satellite news gathering." The ubiquitous characteristics and widespread coverage provided by satellite footprints means that it is technically possible to uplink a signal from virtually any point on the earth's surface. In the 1980s, the only limitations were accessibility for the necessary uplink vehicles with antennas of 3 m diameter – although there were many examples of intrepid teams taking equipment to the jungles of South America and the Antarctic.

As technology developed – including the gradual transition from the mid-1980s to the 1990s from analog to digital signals – the size of the transportable uplink equipment was reduced, first to "flyaway" units which could be shipped in boxes as regular airfreight and then to portable one-man backpacks. At the same time, the arrival of each new generation of satellite news gathering (SNG) equipment was matched by a reduction in satellite transmission costs and improvements in technical quality. Thereby, miniaturization and mobility has now made possible a whole new era of "instant news" coverage, with the latest being the use of videophones and even video over Internet.

Another aspect of this trend toward miniaturization has been the use of multiple miniature cameras for special events (with or without satellite transmission) bringing a new dimension to sporting events and major outside broadcasts. Cameras can now be located discretely in race cars, on ocean-going yachts and even on sports fields and sometimes on players themselves.

New systems, reduced costs, and ingenuity have all contributed to a transformation in the possibilities available for "contribution" by satellite in the past 10 years or so and the trend will undoubtedly continue.

Distribution describes the use of satellites to transmit TV programming or program channels from a broadcasting center to viewers at home, either by "direct broadcast" to an individual rooftop antenna (as will be discussed shortly) or to a cable-TV operator for onward transmission to local subscribers or to an over-the-air local TV broadcaster. There is the possibility of using "distribution service satellites" for the so-called direct-to-the home satellite service in such a way as to "approximate" a direct broadcast to the consumer service. This can involve so-called backyard dishes to receive signals (with or without authorized decoders) in rural and remote areas. Or it can involve very high-powered FSS satellites that provide a service that "mimics" direct broadcast satellite services to very small dishes, but in fact is delivered in the FSS frequency bands. The classic case of the latter type of service is the SES Astra system in Europe. Thus distribution covers both indirect television distribution via cable television providers or over-the-air television broadcasters and direct broadcasting directly to end users (including direct-to-home television services that do not technically use direct broadcast satellites but still come directly from the FSS-type satellite to the end user).

Satellite television distribution, therefore, is provided in the form of a point-to-multipoint transmission, where the signal from a single uplink can then be received by suitable antennas at any point within the "footprint" of the satellite. These broadcast signals may be "free to air" or they can be encrypted to allow access only by paying subscribers. Clearly there is a "gray area" between where "distribution" ends and "direct broadcasting" (or broadcasting satellite service) begins. The distinction is based as much as anything on the frequency band utilized rather than the functional operation of the service.

Direct broadcasting by satellite (DBS) or the broadcasting satellite service (*BSS*): The first implementation of DBS services developed in the 1980s much at the same time that international satellite communications was being privatized. The

plan was to utilize a new type of high-powered satellites with a specific new global allocation of downlink frequencies and orbital locations to provide this type of service. In some instances this was seen as a new opportunity to meet national ambitions to have a new truly national television service. However, such systems were high-cost ventures with limited channel capacity (since only high cost and limited capacity analog transmission technologies were available at that time.) The early operators in the USA, Europe, and Japan had limited success. These ventures essentially failed due to the fact there was limited programming available at reasonable cost. Also, the satellites were not sufficiently powerful and analog technology could not provide sufficient channels to compete with terrestrial cable television systems. In addition, the home receivers were too large and proved to be difficult and expensive to install (Farr 2008).

To the viewers at home, however, these technical improvements were not significant. It was the range and quality of the programming made available by satellite – plus the low cost of the receivers and service – which created a new mass market. The DirecTV satellite system launched by Hughes and now spun off as a separate corporation, together with Echostar, has today established their leadership in the direct broadcast television service in the USA. Meanwhile, Rupert Murdoch's Sky broadcasting ventures in Europe, Australia, and Asia began to develop new television services which competed seriously with the traditional terrestrial broadcasters by acquiring rights for major sporting events and building up significant audience numbers for their pay-per-view or subscription services. In Europe, Sky together with other operators utilized capacity of the highly competitive direct-to-home television services provided by SES Astra out of Luxembourg that offers a "quasi-DBS service" via very high-powered FSS satellites. Further, Eutelsat now operates a series of Hotbird direct distribution television services across the countries of Europe.

The result is that satellite TV is now received directly in over 25% of TV homes in Europe, but with variations from country to country. In smaller countries such as Belgium and Holland with widespread cable networks, the figure is below 10%.

In Japan, BSS services have been quite successful. The first experimental broadcasting satellite, called BSE or Yuri, was launched in 1978. The major national broadcaster, NHK, started experimental broadcasting using the BS-2a satellite in May 1984. This provided just three channels of programming to 40–60 cm (13 to 20 in.) home antennas. However, two of the three transponders failed within a few months and regular transmissions did not begin until the launch of BS-2b in 1989. Another Japanese company, JSB, started broadcasting via the BS3 satellite in 1991, and by 1996 the total number of households receiving satellite TV exceeded 10 million. Today, in addition to the NHK service, there are in Japan direct broadcast satellite systems operated by BSAT, JCSAT, WOWOW Broadcasting and SKY PerfecTV.

In Australia, satellite television has proved to be more feasible than cable TV due to the long distances between cities. Foxtel operates both cable and satellite services to all major cities and its main competitor is Optus Vision. Meanwhile, rural areas are served by Austar.

Another country with a widespread area to cover is Russia, where satellite TV began in the days of the Soviet Union using the Moskva system via the Gorizont and Express satellites. Today, satellite broadcasting is based on the more powerful Gals, Express, and Yamal satellites, while Eutelsat also provides program channels to Russian homes.

The Middle East region also has a high penetration of homes receiving satellite TV. MBC broadcasts from Cairo via Arabsat and one of its competitors is One TV, based in Dubai and mainly serving the expatriate community with Western programming. The first digital DTH network was Orbit Satellite TV, transmitted from Rome, and there are now many other channels available to viewers in the region from the Arabsat, Asiasat, Eutelsat, and Panamsat systems. In Israel, satellite TV is distributed via the national Amos system.

In all regions of the world, as the technology developed for satellite television "contribution," "indirect distribution," and "direct broadcast" services, the various methods have been increasingly combined on a single satellite. Today, whether it is a significant news story, even an earthquake in a remote part of the world, or a major sporting occasion, the incoming program material is frequently simultaneously retransmitted to a broadcaster's viewing audience. Thus the distinction between "contribution" and "distribution" and "direct broadcast" has largely become invisible to the viewer at home.

The Special Role of CNN in Global Satellite Television Development

A major driver in the development of TV by satellite was the launch of CNN, the 24-h Cable News Network which began as a continuous channel distributed by satellite to cable-TV stations around the USA in 1980. The founder, Ted Turner, boasted that once he had switched on the first transmission, it would never stop! And so far, that has proved to be true.

From its Atlanta, Georgia, headquarters, CNN started to build a national and then international newsgathering network and although it was at first derided by the established US networks, it soon became serious competition. In time, the character of CNN began to change to include more in-depth reporting, news features, and interviews – including long-running series with commentators such as Larry King and Wolf Blitzer. However, in 1993, they launched a companion channel called CNN Headline News to maintain the continuous news style.

Major news events such as NASA's Challenger disaster, the Gulf War, the 9/11 terrorist attacks, and successive US election campaigns helped CNN to build growing viewer loyalty. Meanwhile, the Turner broadcasting group, TBS Inc. has grown at a phenomenal pace and now also includes the following networks and businesses:

TBS, Turner Network Television (TNT), Cartoon Network, Turner Classic Movies (TCM), truTV, Adult Swim, Boomerang, TNT Europe, Cartoon Network Europe, TNT Latin America, Cartoon Network Latin America, TNT & Cartoon Network/Asia Pacific, Cartoon Network Japan, Cable News Network (CNN), HLN,

CNN International, CNN en Español, CNN Airport Network, CNN en Español Radio, CNN.com, and CNN Newsource. Today, CNN claims that its services reach nearly one billion people around the globe but it also has to compete with other 24-h news channels transmitted by organizations such as the BBC, Rupert Murdoch's Sky, and the Qatar-based Arabic operator, Al Jazeera.

The international expansion of CNN is another story – and one of the early targets was China. This involved an agreement where CNN would give free satellite terminals to the Chinese leaders so that they could see the world's news. In return, China would ask only $50,000 from CNN for the broadcasting rights for the first year. However, CNN soon found that the Intelsat satellite tariffs to deliver this service would be in excess of $20 million a year when per-minute charges were computed.

At about the same time, another TV mogul, Kerry Packer in Australia was urging his national member on the Intelsat Board of Governors to lease a full-time television circuit to spread his Channel-7 programs around the Pacific region. This led to spirited discussions among Intelsat's international members and after a few months, a new rate emerged enabling the annual lease of a TV channel on a full-time basis – subject to the availability of suitable capacity. It took even longer to agree a formula which imposed no restrictions on the number of downlinks within the same coverage area.

Daily News by Satellite

The daily flow of news by satellite outside of CNN and the USA took much longer to achieve. Throughout the 1970s and 1980s, most of the world's TV stations and networks, large and small, relied on companies who were geared to move television news and entertainment around the world efficiently. These organizations included Visnews, based in London, UPITN based in London and New York, and the syndication service of America's CBS network. These organizations had camera crews located in all corners of the globe and they relied on air freight to ship their news stories back to their headquarters to be edited, scripted, and copied, and then redistributed again by air shipments to hundreds of waiting TV stations around the world. Originally, this was 16 mm film – first black and white and then color – until the first technology shift in this business came with the arrival of video cameras and videotape.

For television news editors and their viewers in every part of the world, these air shipments of film and then videotape provided the only source of foreign material for inclusion in daily news bulletins. This material included not only the top news events of the day, but also sports events, fashion shows, and even those humorous items such as "the skateboarding dog" so often used at the end of news programs.

The news agencies provided an increasingly efficient service and as the world TV market expanded, they became large and established businesses. But their product inevitably reached the viewing public 2 or even 3 days after the event in many countries and news commentary writers had to evolve creative ways to

prepare scripts which incorporated days-old video into their bulletins without loss of immediacy. But there was a solution waiting – in outer space.

It was at the London-based agency, Visnews, where in the late 1970s it was recognized that satellites could provide a whole new future for news coverage and syndication. By then, it was already technically possible to collect video news pictures from almost anywhere in the world by satellite, and then to redistribute it from a single uplink to any TV station within the coverage footprint of a satellite. And yet, as CNN was discovering in China at about the same time, the costs of video transmissions via Intelsat and its members made this uneconomic. The charges for video on a per-minute basis was far beyond the news budgets of even the larger TV stations around the world.

News is a 24/7 business and the first objective was to find a way to distribute the daily packages of news stories to overseas TV stations, 7 days a week and 365 days a year. Executives at Visnews began to explore the possibilities for changing the structure of Intelsat tariffs charged by their national members. But it was not until 1982 that a breakthrough was finally achieved. It came in Australia, where the five competing national TV networks (all of them customers of Visnews) agreed that they would share the costs of a daily transmission from London. Protracted negotiations took place in London between Visnews and BT (British Telecom), and also in Sydney between Visnews, the Australian broadcasters, and OTC (their national telecommunications carrier). Eventually, BT and OTC agreed a basis for securing a block of satellite capacity from Intelsat on the Indian Ocean satellite for a regular daily 10-min transmission from the UK to Australia.

This pioneering breakthrough effectively ended the constraints of regulation and tariffs which had blocked the evolution of video transmission by satellite for many years – and Australian viewers entered a new era of same-day pictures on their evening news shows.

The next step was to exploit the fact that the same daily transmissions from London via the Indian Ocean satellite could also be received in other Asian countries and negotiations began with the TV networks in Japan and their national telecommunications carrier KDD. This was the biggest market in the region and as in Australia, the competing TV stations agreed to share the cost of a downlink from the daily news feed from Visnews in London. And so, by 1982 the international distribution of TV news by satellite began to take off.

Transatlantic services were a bigger challenge. Because of the size of the competing broadcasters in the USA and Europe, they were unlikely to reach a cooperative agreement and Visnews took a different approach by forming a new company to lease a full-time transponder from Intelsat and then sell spot capacity to the broadcasters. This company was named Brightstar, a joint venture between Visnews of London and Western Union in the USA. (Note: A few years later, Visnews was acquired by the Reuters news agency.) Then as digital technology began to replace analog, the American news agency, Associated Press entered the market aggressively with the launch of its AP-TV service.

Meanwhile, other companies such as IDB and Wold International in Los Angeles and Bonneville in Salt Lake City were operating satellite delivery services

to broadcasters in the US domestic market. In 1989 Wold and Bonneville merged to create Keystone Communications which went on to expand into international markets. Next Keystone acquired the video distribution business of IDB and in 1996, the company was acquired by France Telecom and became Globecast, which remains the largest supplier of transmission service to the global broadcast market and leases over 100 satellite television channels around the world.

How Television via Satellites Influenced Global Politics

One of the most momentous events of recent years was the end of the Cold War and in particular the fall of the Berlin Wall in November 1989 (Marshall and Wold 2004). These major political events owed much to how the people of Eastern Europe, through TV and radio became aware of the freedoms enjoyed in the West.

A West German commentator put it this way: "Totalitarianism could not survive in the East when the people's antennas were pointing to the West" (Shane 1994). The East German leader had recognized the danger when he said: "The enemy stands on the rooftops" and ordered the communist youth brigade to clamber on to houses and remove the offending antennas. But there were just too many antennas by then and they gave up the unequal task of trying to stop people watching West German TV. This was the period of Mikhail Gorbachev's *glasnost* which brought an era of increased political freedom in the Soviet Union. According to author Scott Shane in his 1994 book *Dismantling Utopia: How Information Ended the Soviet Union*, the KGB's information blockade turned out to be "more like a tennis net than an iron curtain" (Shane 1994).

Meanwhile, in Moscow the headquarters of the Central Committee set up an antenna to receive news of world events from CNN's satellite transmissions and when this became known, many of the general public began to receive the programs too through the use of illegal and home-made antennas. Radio also played its part in the downfall of communism and the Polish leader Lech Walesa wrote later: "If it were not for independent broadcasting, the world would look quite different today. Without Western broadcasting, totalitarian regimes would have survived much longer" (Nelson and Walesa 1997).

As satellite broadcasting resources expanded, "instant news" became a factor in shaping public opinion. In the later stages of the Vietnam war, Americans came to trust the coverage shown by Walter Cronkite of CBS News and became less willing to accept the official statements from the US military. At that time, the film coverage from Vietnam had to be flown to Hong Kong or Tokyo for satellite transmission across the Pacific but it still made a graphic impact on the TV audiences and this was an important factor in the change of US policy and the eventual withdrawal.

Another early example of how instant television coverage could make a large political impact came in 1984. This was the case of the Shatila Refugee Camp massacre in Lebanon. Several Western TV cameramen arrived on the scene at

almost the same time as the main Israeli forces. This was just a day after an advance Israeli force had descended on the camp and had essentially engaged in a massacre. The Israeli government was still considering how to explain and minimize the political impact of this horrible event when the first video images were being fed by satellite to Visnews in London. The undeniable images of mass death were widely shared by broadcasters in Europe and the USA. This resulted in a global outcry that took the Israeli government by surprise before top leaders were even aware of what had transpired.

The Chinese government was certainly not prepared in 1989 for the live global video coverage of the events in Tiananmen Square. By coincidence (or perhaps not), international TV crews and satellite links were already in Beijing for a visit by the Soviet leader, Mikhail Gorbachev, and there was little the authorities could do to restrict their activities. The use of tanks against unarmed students and the symbolic picture of a lone student facing down the advancing military units had immediate impact around the world.

By the first Gulf War in 1991, the technology had moved on and the images of invasion, and the bombs and missiles over Kuwait and Baghdad were also seen live via satellite. So too were the military incursions into Grenada and Somalia.

More recently, on September 11, 2001, the terrorist attacks on the World Trade Center and the Pentagon were seen "live" by satellite around the world, thereby influencing public opinion on a global scale. What then followed, first in Afghanistan, then in the United Nations and in the invasion of Iraq in March 2003, were examples of ways in which both public diplomacy and warfare are now carried out in the full glare of instant media coverage.

Sports Programming on Communication Satellites

Sports coverage on television – and especially live sport – invariably works to attract big audiences. It is therefore to be expected that some of the biggest audiences in the world have come from satellite coverage of the Olympic Games and World Cup soccer tournaments.

It was the experimental Relay-1 satellite in 1964 that provided limited television coverage of the Winter Olympics at Innsbruck in Austria. The European Broadcasting Union's production people were able to send limited coverage to the USA. Although the EBU's programming only lasted 20 min at a time due to the satellite's low Earth orbit, it still gave audiences a taste of what was to come.

A few months later, in October 1964 came the Tokyo Olympic Games described earlier. Since 1964 the global audience for the Olympics has grown to staggering levels. According to Nielsen Media Research, 4.7 billion viewers worldwide tuned in to at least some of the television coverage from the Beijing Olympics in 2008 – one-fifth greater than the 3.9 billion who had watched the previous 2004 Olympic Games in Athens.

In the intervening years, the IOC (International Olympics Committee) has been able to exploit the growing global audience and today it has become more than a billion-dollar enterprise. More than half of the IOC revenues come from the TV rights which are the subject of fierce negotiations among the major broadcasters.

As satellite systems expanded the market, viewership increased exponentially and the escalation of broadcasting rights has rocketed. For example, the rights to the Winter Games in Innsbruck in 1964 were sold for $936,000. For the Atlanta Summer Olympics in 1996, NBC paid $456 million and by the Sydney Olympics in 2000, the figure had nearly doubled to $707 million.

Then for the Winter Olympics in 2006 and the Summer Olympics in 2008, NBC was willing to pay a combined total of $1.5 billion. And this has been followed by a staggering sum of $2.2 billion for the 2010 Winter games in Vancouver together with the 2012 Summer games in London.[1]

Estimates of the worldwide viewership were 600 million for the Mexico City Games in 1968; then 900 million for Los Angeles in 1984. But the arrival of global satellite distribution swelled the audience to an estimated 3.5 billion for the 1992 Games in Barcelona. Every 4 years the number of viewers continued to rise and in 2008, the Beijing Olympic Games attracted the largest global TV audience ever. The Nielson Company estimated that between August 8 and 24, some 70% of the world's population tuned in to watch the TV transmissions.[2]

Over the same period, the worldwide enthusiasm for soccer produced comparable viewing figures for each successive World Cup tournament. This event was first televised in 1954 and now competes with the Olympics as the most widely viewed and followed sporting event in the world. The cumulative audience for all of the matches played in the 2006 tournament in Germany was estimated to be over 26 billion spread across 214 countries and territories. For the Cup Final, it was estimated that 715 million individuals watched the match (a ninth of the entire population of the planet).[3]

The Growth of "Direct-to-Home" Satellite Television

With "national" satellite systems in place, it was not surprising that a number of scientists, engineers, and businessmen around the world (and especially in North America and Europe) began to ask why not send the satellite signal directly to the home and bypass the terrestrial networks. There were regulatory as well as technical obstacles to be overcome, but many consumers, particularly those in rural and remote areas without access to cable television or over-the-air broadcast, began buying back yard television receive only (TVRO) dishes to receive satellite TV intended for cable stations. In many cases they also bought "descramblers" and began to watch so-called premium channels such as HBO and Cinemax. After a few years there were literally millions of these "pirate" backyard dishes in the USA, Caribbean, and elsewhere and it became obvious that so-called direct broadcast satellite television services would be a viable business.

Indeed today direct broadcasting represents the largest commercial satellite market, as measured in the size of its global revenues. Because of this history, whereby the initial service evolved from fixed satellite service (FSS) satellites providing service to backyard dishes before the arrival of new types of more powerful direct broadcast satellites designed for direct-to-home services (DTH), there is a confusion as to where one service ends and the other begins. Today, with a more pragmatic approach, the designation DTH is generally used to cover both types of services, whether for backyard dishes served by FSS satellites that operate in several downlink frequency bands and broadcast satellite services (BSS) spacecraft that operate in another downlink band.

However, BSS remains the formal terminology used by the International Telecommunication Union (ITU) for a direct-to-the-consumer television service. This BSS offering is considered by the ITU to be separate from the FSS satellites and it is different in that higher transmit powers are authorized, different frequencies are allocated to the service, and the user terminals are typically much smaller and compact and cost less money than the backyard dishes. BSS terminals are typically 30 cm to 1 m in size (12–39 in.) while backyard dishes are 3–7 m (10–29.5 ft) in size.

The "true" BSS direct broadcast service is one in which many television channels are uplinked to a broadcast satellite in geosynchronous orbit and then downlinked at very high power via a highly concentrated beam to a country or region so that the signal can be directly received by quite small home or office-mounted satellite antennas. Under the ITU allocation of frequency bands for this type of service, different spectra are used in different parts of the world. These high-power downlink transmission bands are as follows: The spectrum 11.7–12.2 GHz is allocated in what is known as ITU Region 1 (This region includes Europe, the Middle East, Africa, and parts of Russia). The downlink spectrum of 12.2–12.7 GHz is allocated for ITU Region 2 (this region includes all of the Americas). Finally the spectrum band of 10.7 to 12.75 GHz is allocated for downlinking BSS services in Region 3 with includes Asia and Australasia.

Many countries, especially developing countries and those without satellite communications capacity, thought that the initial process by which frequencies were allocated for FSS services at special and extraordinary sessions of the World Administrative Radio Conferences in 1959 and 1963 were biased in favor of the most advanced countries. Thus when sessions were held to allocate frequencies for this new broadcast satellite service in the 1970s, a number of countries supported the idea that allocations of frequencies for this service should also include allotments for countries that did not yet have satellite capabilities against their future needs.

In the late 1970s and early 1980s the European Space Agency was interested in launching an experimental direct broadcast satellite that was given various names L-Sat, H-Sat, and eventually Olympus. This project was complicated by a procurement process in which the French and German governments decided that they would not sign on to fund their "voluntary allocations" associated with

this project and instead decided to proceed with their own joint project known as TV-Sat in Germany and the FR-3sat in France.

As these various "official projects" proceeded to develop direct broadcast satellites in accord with ITU allocations, the Luxembourg-based company now known as SES decided that it would use a high-powered fixed satellite service (FSS) satellite, operating in the FSS Ku-band to broadcast over Europe to provide what became known as direct-to-home (DTH) service. Consumers could have small antennas installed at their home and receive a wide range of television programs – far wider than that offered by national terrestrial broadcasters – and the consumers cared little about ITU allocations or service definitions. In short SES, via its Astra satellites, stole a march on these European official BSS alternatives. In the UK, the company known as British Satellite Broadcasting made expensive investments but when they ran into financial difficulties, they merged with Rupert Murdoch's Sky Television, operating on the Astra satellite system. This company became known as BSkyB and now transmits the multichannel Sky services targeting much of Europe and with nearly ten million subscribers in the UK alone.

Satellite Radio Broadcasting

Radio broadcasting is now an important additional aspect of the satellite industry. Commercial satellites from the earliest days were able to use broader band channels to send high-quality audio, music, and radio shows from one location to another. As the transition was made from analog to digital satellite transmission, the idea that satellite radio systems might be deployed came into much clearer focus. The motivations for this type of service came from many different perspectives. One motivation was from "official" national radio broadcasting systems that might be characterized as sending favorable "propaganda" on behalf of one country or another. During the Cold War years, very large amounts of money was spent to establish high-power terrestrial broadcast systems to send music, entertainment, and news to locations across "Cold War boundaries."

It was not surprising that many satellite planners thought that a global or regional satellite system, able to broadcast to small compact shortwave radio receivers, could be a technically more efficient and cost-effective method for sending this information to millions of listeners. From another perspective, broadcasters, educators, and news people in the developing world recognized that there were more radios than television sets in some of the poorest countries. They believed that a satellite radio broadcasting services might be an effective way to reach a new and broader audience in these parts of the world.

In the more economically advanced areas, entrepreneurs envisioned that a satellite radio broadcasting system might be a way to reach a broad new audience in their automobiles and even home listeners and office workers who wanted high-quality news, entertainment, and sports on a commercial-free basis by paying just a small monthly subscription fee. For these various reasons the ITU allocated

frequencies for a satellite broadcast service that is variously known as digital audio broadcast service (DABS), direct access radio service (DARS), and broadcast satellite service-radio (BSS-R).

Once a frequency allocation for satellite radio broadcasting was finally agreed a number of companies actively pursued this business. At the lead was a company called Worldspace, headed by a charismatic Ethiopian visionary, who had actually been a leading advocate of this new radio service and who had led the fight for new satellite frequency allocations within the ITU processes. Worldspace proceeded with the immediate design, manufacture, and launch of its Worldstar satellites in partnership with the French firm of Matra Marconi. The first of these launches put Worldstar 1 into geostationary orbit for new radio broadcasting services to cover all of Africa, the Middle East, and parts of Europe.

The business model for Worldspace was to offer a lease of one or more individual radio channels to broadcasters who would provide their own programming to be transmitted on the satellite broadcasting down links. These channels could be used not only for radio news, sports, and entertainment, but at nighttime (or even daytime in some places) they could be utilized to provide educational programming. Indeed these digital radio channels could even be used to download short video educational programs which required a period of hours to send a "slower and narrower signal," using the smaller "digital pipe." The cost of the radio sets (including the attendant national import tariffs that often doubled the actual cost of the satellite radio receivers) ranged from about $100 to $200 (US) and resulted in a largely unsuccessful business model. As a result Worldspace encountered major financial difficulties.

Meanwhile, in the USA, two domestic systems known as Sirius (using highly elliptical orbits) and XM Radio (using geosynchronous satellites) were licensed by the FCC and both companies managed to successfully deploy radio broadcast satellite systems. However, the very high cost of building and deploying very large aperture satellites for this type of service led to financial difficulties for both these satellite radio services. Their services were marketed largely to automobile owners on a subscription basis, offered on a 24/7 basis with a very wide range of radio programming, emergency communications, and antitheft services. XM was offered via General Motors automobiles and Sirius was offered via Chrysler and Ford. When the financial difficulties mounted these two systems eventually merged, with XM Radio essentially acquiring Sirius.

In Europe there are also plans for a radio broadcasting satellite service, but the financial problems with the other systems appear to have delayed the deployment of a European system.

High-Definition Television (HDTV) via Satellite

One of the major developments over the past 20 years has been the introduction of HDTV, which provides much greater video resolution than standard TV – i.e., one or two million pixels per frame. The early experiments in the USA, Japan, and

Europe used analog systems, but these were rapidly overtaken by the introduction of digital broadcasting.

The most successful analog experiment was carried out by NHK in Japan with their MUSE system, but bandwidth restrictions meant that only one channel could be carried by satellite. Although it provided public transmissions until 2007, it was superceded by a growing range of digital HDTV channels from 2000 onward.

In Europe, there were also analog experiments in the 1990s with the 1,250-line HD-MAC system, but this did not lead to a public broadcast system and it was abandoned when digital systems were developed.

In the USA, the Federal Communications Commission (FCC) brought together a group of TV companies, together with the Massachusetts Institute of Technology (MIT) and created the Digital HDTV Grand Alliance in the 1993. They began field testing in the following year and the first public broadcast took place from the WRAL station in Raleigh, North Carolina, in July 1996.[4]

This led to the formation of the American Advanced Television Systems Committee (ATSC) which organized the live coverage of astronaut John Glenn's return space mission in October 1998, which was transmitted in HDTV to specially equipped theaters and science centers across the country.

Following that inaugural transmission, the TV networks gradually introduced HDTV technology and as more and more channels became available, the sale of home receivers expanded. The first major sporting event to be broadcast in HDTV was the Superbowl XXXIV in January 2000.

Satellite TV companies, such as DirecTV and the DISH network, started to carry HDTV programming in 2002 and today, dozens of channels are available on all the major networks as well as via satellite, cable, and Internet systems. A survey in November 2009 by Home Media magazine found that between 33% and 50% of Americans have at least one HDTV receiver in their home (Home Media Magazine 2009).

Meanwhile, the European Broadcasting Union (EBU) coordinated the development work among major broadcasters and the first HDTV broadcast in Europe was the traditional New Year's Concert from Vienna on January 1, 2004. The Belgium-based company Alfacam was the first to begin broadcasting its HD1 channel on SES-Astra's 1H satellite with 4 or 5 h of programming each day. This helped to "jump start" the sale of HDTV receivers and over the following years, the number of channels gradually increased so that today there are 114 HD channels carried on Astra and Eutelsat satellites.

The number of European households receiving HDTV channels increased to 6 million by the end of 2009 and SES-Astra has forecast an increase to 24.7 million by 2013 (SES-Astra Presentation 2010).

Future Trends

Another technological development seeking to provide a new dimension for both terrestrial and satellite TV is 3D (three-dimensional television), but this is still in its

early and experimental stages. Following on from the success of "blockbuster" 3D Hollywood movies in the cinema, the TV industry is now trying to catch up.[5] Major broadcasters in many parts of the world have invested in the new equipment necessary to shoot and transmit in this format and have carried out a range of test and promotional transmissions.

However, despite a great deal of competitive activity, most content providers and manufacturers appear to be seeking standardization of 3D home electronics technology across the industry before moving too quickly into program and movie production. Disney, Dream Works, and other Hollywood studios, and technology developers such as Philips, have asked SMPTE (Society of Motion Picture and Television Engineers)[6] and other industry groups for the development of a 3DTV standard in order to avoid a battle of formats and to guarantee consumers that they will be able to view 3D content and to provide them with 3D home solutions for all pockets.[7]

With improvements in digital technology, 3D movies have become more practical to produce and display, putting competitive pressure behind the creation of 3D television standards. There are several techniques for Stereoscopic Video Coding, and stereoscopic distribution formatting including anaglyph, quincunx, and 2D plus Delta.

Most currently available receivers equipped to receive 3D signals are very expensive and require the viewers to wear special eyeglasses. Several major companies are planning to enter this market in 2010 or 2011 and one of the first is Panasonic – their Panasonic Viera TC-P50VT200 comes with glasses and has a retail price of approximately US$2,500. In June, 2010, they also announced a 152 in. (390 cm) 3D-capable TV (the largest so far) that will go on sale for 50 million yen (US$576,000). However, it is also reported that the Samsung UN46C7000 46-in. 3D TV can now be purchased for US$2,000 or less.[8]

Meanwhile, companies including Philips and Toshiba are reported to be developing 3D television sets using autostereoscopy technology which will obviate the need for special glasses to be worn by viewers. Also, the Chinese manufacturer TCL has worked on the same technology to develop a 42-in. (110 cm) LCD 3D TV which is currently available in China. This model uses a lenticular system and currently sells for approximately US$20,000.

TV program transmission of 3D services has been launched in the USA, Europe, and Asia. In fact, the first 3D programming was reportedly broadcast on the Japanese cable channel, BS11 in 2008 with four programs each day. However, the first complete channel is believed to be the Sky3D channel which began broadcasting in South Korea in January 2010. Then in the UK, the Sky Sports 3D channel was launched in April 2010, with coverage of the World Cup soccer in July and the company has announced that its Sky Movies 3D will follow later in the year.[9]

In the USA, the satellite broadcaster DirectTV provided a live demonstration of their 3D feed at the Consumer Electronics Show in Las Vegas in January 2010 and they became the first US service provider to offer a complete 3D channel package in July 2010.[10] Now with four 3D channels (DirecTV Cinema, N3D, N3D On Demand and ESPN 3D), it is ahead of most cable and satellite providers.

Meanwhile, Cablevision launched a 3D version for subscribers to its MSG channel in March 2010[11] and ESPN transmitted 25 matches in 3D from the World Cup soccer tournament in South Africa – the first major sporting event to be broadcast in 3D (Satnews April 2010).

ESPN is planning to start transmitting a new channel dedicated to sports and with up to 85 live events a year in 3D (Guardian January 6, 2010).

In Australia, Nine Network and SBS also used a major sporting event, this time a Rugby tournament, for a joint 3DTV trial from May to July, 2010. Using a system developed by the Harris Corporation, the events were broadcast in 3D in Australia's five major metropolitan centers – Sydney, Melbourne, Brisbane, Adelaide, and Perth – as well as in Wollongong and Newcastle (Satnews June 21, 2010).

Conclusion

Satellite television and audio distribution and broadcasting have dramatically changed our world over the past half century. Today people expect – and receive – instantaneous coverage of global news and sporting events in what seems almost like live coverage. Satellite technology has evolved rapidly to allow more immediate origination of news programming in the field. Reporters with mobile units are able to uplink programming instantly from even the most remote parts of the world. Satellite technology has also become more sophisticated in distributing radio and television programming. Satellites are essential to global television and radio programming since it is satellites that is the most common way to deliver programming either to the head-end of cable television networks or alternatively to broadcast programming directly to consumers in homes, offices, or even automobiles, auto buses, trains, and airplanes. As the technology has matured and the cost of satellite programming has dropped over time, satellite video, and audio has become more and more pervasive around the world and the number of satellite television channels has soared to a number perhaps greater than 20,000. The advent of digital satellite television has particularly allowed the cost of satellite television to drop and the number of channels available to expand rapidly. Today the latest frontier is the development of live broadcast of 3D high-definition television directly to consumer-receiving units and digital audio broadcast services directly to consumers on the move.

Notes

1. http://en.wikipedia.org/wiki/Olympic_Games
2. http://blog.nielsen.com/nielsenwire/media_entertainment/beijing-olympics-draw-largest-ever-global-tv-audience
3. http://www.FIFA.com
4. History of WRAL Digital www.**wral**.com/**wral**-tv/story/1069461/
5. http://www.bbc.co.uk/news/10446419
6. www.televisionbroadcast.com/article/93370
7. http://www.eetindia.co.in/ART_8800569756_1800010_NT_d9538c56.HTM
8. http://en.wikipedia.org/wiki/3D_television

9. http://www.3dtv-prices.co.uk/sky.html
10. http://news.cnet.com/8301-17938_105-20009692-1.html#ixzz15XRmc6xS
11. http://www.msg.com/3d

References

Arthur C. Clarke, "Peacetime Uses for V2" *Wireless World*, (London, UK, February 1945)
Arthur C. Clarke, "The Space Station," *Wireless World*, (London, UK, October 1945)
K. Farr, "*Satellite Television: How Did It Start*?" Background Notes for History of Telecommunications Course (2008), kfarr.com/2008/01/26/satellite-tv-how-did-it-start/
Guardian.co.uk – 6 January 2010
Home Media Magazine November 25, 2009
P. Marshall, R. Wold, The World of satellite TV: news, the Olympics and global entertainment, in *Communications Satellites: Global Change Agents*, ed. by J.N. Pelton, R.J. Oslund, P. Marshall (Lawrence Erlbaum, Mahwah, NJ, 2004), p. 256
M. Nelson, Lech Walesa, in *War of the Black Heavens* (Brassys, UK, 1997)
New York Times, 10 October 1964, p. B1
J.N. Pelton, *Global Communications Satellite Policy: Intelsat, Politics and Functionalism* (Lomond Press, Mt. Airy, MD, 1974), pp. 48–53
Satnews April 6, 2010 – www.satnews.com/cgi-bin/story.cgi?number=447140674
Satnews June 21, 2010 http://www.satnews.com/cgi-bin/story.cgi:number=90185488
SES-Astra Presentation March 17, 2010 (unpublished)
S. Shane, *Dismantling Utopia: How Information Ended the Soviet Union* (Ivan R. Dee, Chicago, 1994)

Mobile Satellite Communications Markets: Dynamics and Trends

8

Ramesh Gupta and Dan Swearingen

Contents

Introduction 164
First Commercial Mobile Satellite System 165
Early Studies and Programs 167
First Commercial Mobile Satellite System – MARISAT 167
Aeronautical Mobile Satellite System Introduction 172
Portable Mobile Earth Stations for Telephony and Data 174
Inmarsat-4 Satellites and Broadband Global Area Network (BGAN) 174
National Mobile Satellite Systems 176
Low Earth Orbit (LEO) Satellites 177
Geosynchronous Regional Satellite Systems 178
Terrestrial Cellular Service Convergence with MSS 180
Hybrid Transparent MSS Networks 180
Mobile Satellite Systems with ATC 181
FSS Convergence with MSS 184
Summary 185
Cross-References 185
Notes 185
References 186

Abstract

The first commercial mobile satellite service (MSS) system was implemented to meet the urgent needs of the maritime community for improved communications. As enhancements occurred in satellite technology and circuit integration

R. Gupta (✉)
LightSquared, 10802 Parkridge Boulevard, Reston, VA, 20191, USA
e-mail: Ramesh.Gupta@lightsquared.com

D. Swearingen
120 N. Galveston St, Arlington, VA, 22203, USA
e-mail: d.swearingen@verizon.net

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_7, © Springer Science+Business Media New York 2013

resulted in availability of low-cost digital signal processors, MSS systems were deployed to support smaller earth terminals for aeronautical and land mobile applications. The growth of the Internet and improved wireless access has led existing MSS system operators to introduce new capabilities that include improved data and multimedia access. This transformation in the MSS markets has offered several technology challenges, particularly with higher power spot beam satellite deployment designed to operate with low-cost personal user terminals.

This chapter addresses the history and evolution of the technology starting with the Marisat system in the 1970s. This historical review covers both the space technology and user terminals, as well as the market characteristics of the various types of MSS systems. Some systems have not been successful in the market and have not done well financially for reasons associated with market demand, cost, and reliability of service and technology. Currently, there is a convergence of terrestrial wireless services with mobile satellite services. One approach has been to offer dual frequency band handsets that switch from one band to the other depending upon whether the user is within the terrestrial system's coverage. The other approach is for the MSS system operator to support both satellite and terrestrial coverage in the MSS frequency bands to achieve seamless operation for the user. The demand for very high data rates for better Internet access had also led to a convergence of MSS and FSS system capabilities, where many of the capabilities of an MSS are being offered by FSS system operators and vice versa. Fortunately, continuing innovations are assuring mobile users the availability of reliable communications from anywhere in the world.

Keywords

Aeronautical Mobile Satellite Service (AMSS) • Ancillary terrestrial component (ATC) • Asia Cellular Satellite System (ACeS) • Complementary ground component (CGC) • Geosynchronous earth orbit (GEO) • Globalstar Satellite System • Hybrid Satellite Systems • ICO Satellite System • Inmarsat Satellite System • International Telecommunication Union (ITU) • Iridium Satellite System • Land mobile satellite service (LMSS) • LightSquared • Low earth orbit (LEO) • Marisat • Maritime Mobile Satellite Service (MMSS) • Medium earth orbit (MEO) • Mobile satellite services (MSS) • MSATs • SkyTerra-1 • Thuraya Satellite Systems • Very small aperture terminals (VSATs) • World Radio Conference

Introduction

Mobile satellite service (MSS) refers to a radio-communication service between mobile earth terminals and one or more satellites. In most MSS systems, the mobile to satellite links are cross connected to feeder link frequencies which operate with fixed gateway earth stations. MSS systems provide voice, data, and Internet services to a wide variety of mobile users including those based on ships, aircraft, and land. One can think of an MSS system as a cellular system with a repeater in the sky

servicing multiple cells (coverage beams) which are much wider than terrestrial cells. The first MSS systems evolved shortly after the successful initiation of fixed satellite services (FSS) which initially used fixed earth stations with large antennas. Ultimately, more powerful FSS satellite technologies evolved which enabled services between very small aperture terminals (VSATs) with 1–2 m diameter antennas.

Mobile satellite communications systems, on the other hand, have enabled travelers to the remote corners of the world to communicate with the rest of the world with small portable or handheld terminals. Those systems, initially developed for maritime communications, were deployed on land for mobile users because large areas of the world were not served by terrestrial systems. While the coverage of terrestrial wireless systems rapidly increased around the world, mobile satellite systems have still been needed to fill in coverage gaps and to satisfy the growing demand for a wide variety of services. These include global messaging, aircraft position reporting, remote area connectivity, disaster relief communications, search and rescue communications, and Internet access for portable and mobile terminals (see Fig. 8.1).

Today a traveler will most likely have a choice of more than one system to choose from, since the coverage areas of numerous systems tend to overlap. There are systems which provide global coverage like Inmarsat (Gallagher 1989), Globalstar, and Iridium, whose coverage overlaps the regional coverage of systems like Thuraya and AceS as well the regional/national coverage of systems like Terrestar and LightSquared (Whalen and Churan 1992) (USA and Canada), Solidaridad (Mexico), Optus (Australia), INSAT (India) and JCSAT (Japan). Although there are systems optimized only for low data rate communications like Orbcomm, most systems were developed to support telephony and various data communications capabilities (Reinhart and Taylor 1992). To understand the capabilities of these various systems and the likelihood of their continued availability, it is helpful to review how these systems evolved along with a brief description of their performance characteristics.

This chapter addresses the evolution of MSS systems over the past four decades, during which satellite technologies have evolved toward higher power satellites with corresponding reduction in the size of mobile terminals.

First Commercial Mobile Satellite System

Mobile satellite communications services were first provided to the maritime community which needed better radio communications resulting in significantly improved safety services on the high seas. Although the fixed satellite services to link the continents were firmly established in the 1960s, it was not until the 1970s before a civilian satellite system for ship to shore communications was established by Communications Satellite Corporation (COMSAT). This system, called MARISAT, was the predecessor to the INMARSAT (International Maritime Satellite) System that began its operations in early 1982.

MARISAT and its successor, INMARSAT, made use of the frequency bands allocated to the Maritime Mobile Satellite Service (MMSS) by the 1971 World Radio Conference. These were a pair of 7.5 MHz wide bands at L-Band near

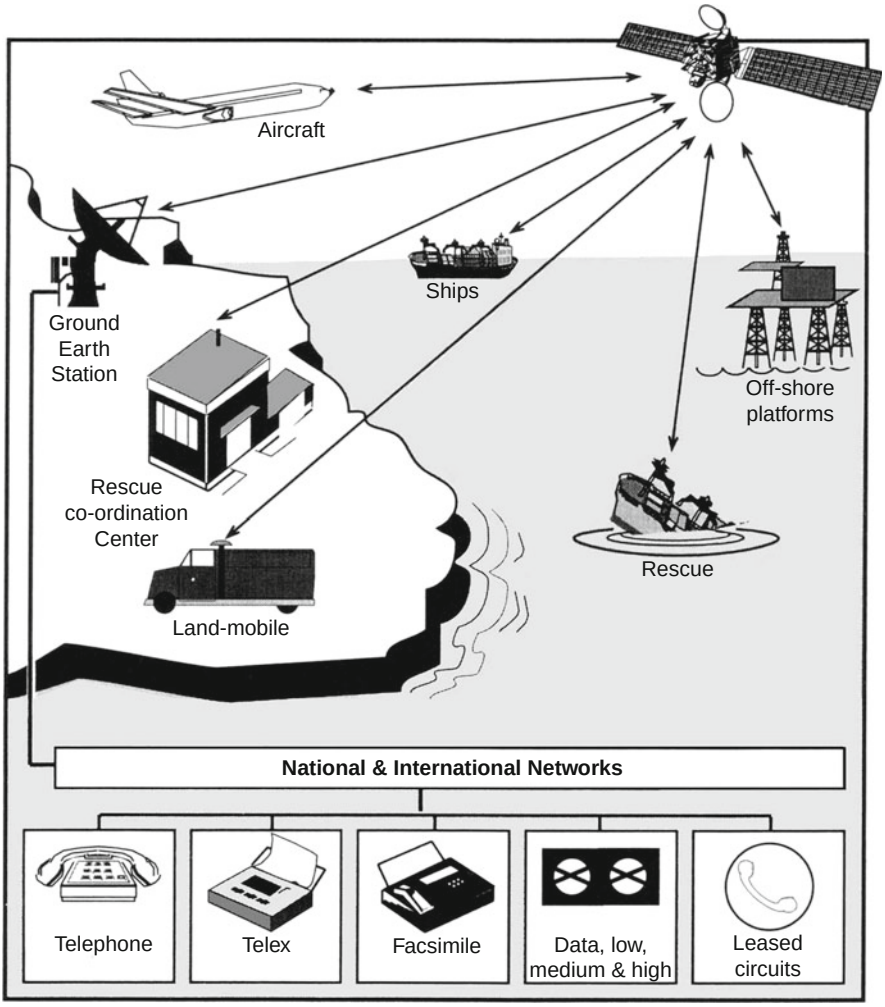


Fig. 8.1 Traditional mobile satellite services (MSS)

1.5 and 1.6 GHz for the MMSS (one band for ship to shore transmissions and the other for shore to ship transmissions). Those allocations were conserved by cross strapping with shore to satellite feeder links in the Fixed Satellite Service bands at 6/4 GHz. It should be noted that the 1971 World Radio Conference had also allocated a pair of 15 MHz wide bands to the aeronautical mobile satellite service (AMSS) which were close to the MMSS allocations. At that time the only L-Band allocations for a generic mobile satellite service were a pair of 1 MHz wide bands between the MMSS and AMSSS allocations that were restricted to distress and safety communications usage. In effect, because development of a land mobile satellite service (LMSS) was not a high priority in the early years, there were no L-Band allocations for LMSS until some years later.

Early Studies and Programs

Prior to MARISAT, high seas maritime communications depended upon terrestrial radio transmissions in the medium and high frequency bands (MF and HF), which suffered from extreme propagation variability, congestion, and lack of automated station facilities. For this reason, various groups around the world conducted communications satellite technology studies and experiments for aeronautical and maritime applications. The US Navy also sponsored studies and deployed the LES (Lincoln Experimental Satellite) series of experimental UHF band satellites for ship and aircraft use in the late 1960s. The NAVY then contracted with TRW Corporation for the construction of five operational ultrahigh frequency (UHF) mobile satellites (i.e., the FleetSatCom series).

The commercial deployment of operational systems for the commercial maritime and aeronautical communities faced serious business start-up challenges. There needed to be enough users equipped with mobile earth stations to recover the satellite investment costs within the 5–7-year lifetimes of the satellites. Furthermore, the maritime and aeronautical communities were somewhat cautious regarding investments in new types of radio equipment.

The maritime community had also begun studies of satellites to improve radio communications for ships under the auspices of the Inter-Governmental Maritime Consultative Organization (IMCO). These studies led to consideration of a possible international maritime satellite system (INMARSAT) to be created through intergovernmental agreement and with an organization similar to that of INTELSAT. The INMARSAT Convention and Operating Agreements were agreed in 1976, but the new organization did not come into existence until the treaty was ratified in 1979.

In the meantime, the European Space Agency began an experimental satellite program called MAROTS to support an L-Band maritime communications payload. Within INTELSAT, consideration was given to the viability of adding a small L-Band maritime payload to its planned INTELSAT-V series of satellites and pricing its usage on an incremental basis. As promising as the studies were, INTELSAT initially decided (in 1972) not to complicate its INTELSAT-V program with an additional payload.

Another opportunity for an add-on commercial maritime payload occurred in 1972, when the US Navy sought a pair of "Gapfiller" UHF satellites to service its ships in the Atlantic and Pacific for a couple of years before the Navy's new FleetSatCom series could be deployed. COMSAT undertook to explore development of a hybrid satellite that could support both a small L-Band maritime payload for commercial maritime services and a UHF payload to satisfy some, if not all, of the Navy's stated capacity requirements.

First Commercial Mobile Satellite System – MARISAT

In 1973, COMSAT enlisted the help of Hughes (Now Boeing) to develop a small multi-payload satellite design concept with both UHF for the US Navy and L-Band

capabilities for civilian users. A fortunate combination of similar coverage requirements (Atlantic and Pacific) and complimentary capacity requirement schedules enabled MARISAT to offer cost advantages to both the anchor user (US Navy) and the civilian users (Martin and Keane 1974).

At its launch in 1976, the MARISAT configuration provided only a small amount of power for the L-Band transponder in addition to the power needed to support the full UHF capacity for the Navy. Later, as the Navy's capacity requirement was reduced or disappeared, more power would be switched to the L-Band payload to satisfy growing civilian traffic demands. A third MARISAT spacecraft was built to be an on-the-ground spare, but after the first two spacecraft were successfully launched in early 1976, the Navy requested that the third be launched to cover the Indian Ocean region. Consequently MARISAT spacecraft were in position to enable global maritime service by the end of 1976.

The MARISAT goal was to improve the reliability of telegraphy and telephony services available to ships in the Atlantic and Pacific satellite coverage areas. Because the digital technology of the mid-1970s was still at the early stage of development and relatively expensive (the first microprocessors were just being introduced), the MARISAT system architecture was relatively simple (Lipke and Swearingen 1974).

Since the 1971 WARC had allocated only 7.5 MHz of L-Band spectrum for each direction of the ship-satellite link, those allocations were conserved by cross strapping with shore to satellite feeder links in the Fixed Satellite Service bands at 6/4 GHz. The high receive sensitivity of the 13 m diameter coast earth station antenna enabled the satellite to shore links to be supported with relatively little satellite power. The bulk of the satellite power could, therefore, be dedicated to supporting the satellite to ship links at L-Band.

The shore to ship satellite repeater translated the 6 GHz carriers uplinked by the coast earth station to 1.5 GHz for reception by the ship earth station. The ship to shore repeater translated the 1.6 GHz carriers transmitted by ship earth stations to 4 GHz for reception by the coast earth stations. In addition to a 13 m diameter antenna system, each coast earth station included telephone and telex channelization systems, an access control subsystem, and switching systems to interface with the terrestrial telephone and telex networks. In order to enable operator assistance to maritime customers, operator positions were also included at the MARISAT coast earth stations.

MARISAT Channelization

A single channel per RF carrier (SCPC) design was selected for telephony, while a time division multiplexed (TDM) digital RF carrier was used for the telex (50 baud data) service in the shore to ship direction. SCPC was also used in the ship to shore direction, while a time division multiple access (TDMA) was used for telex. The telephone channels used frequency modulation (FM) to ensure a transparent audio path for voiceband data. To lower the perception of background thermal noise during voice conversations, 2-to-1 companding was included in the channel path (similar to the choice made a few years later when the first cellular systems were introduced by the Bell System in the USA).

In the early 1970s, telex had become a nearly ubiquitous mode of automatic telegraphy and was well standardized and widely used for international maritime business data communications. For that reason, telex channelization became the MARISAT low speed data channel standard. Each ship terminal included a 50 baud telex machine to support automatic interworking in both the shore to ship and ship to shore directions. Shore to ship telex channels were transmitted on a 1,200 bps BPSK modulated RF carrier using a Time Division Multiple channel (TDM) format that supported 22 channels per carrier. In the ship to shore direction a burst mode Time Division Multiple Access (TDMA) format was selected with up to 22 ships sharing the same frequency by bursting once every 1.8 s with a carrier BPSK modulated at 4,800 bps and containing 12 characters.

MARISAT Access Control

Channel assignment and call alert messages were piggy-backed onto the shore to ship TDM carrier. Channel requests from the ships were carried in short digital RF bursts on a shared random access frequency called the Request Channel. This basic access control architecture (TDM for outbound signaling and Random Access Request Burst Channel for inbound signaling) turned out to be the essential template for later mobile satellite systems, starting with INMARSAT. The access control design objective was to enable calls to be set up quickly in either direction, in a manner as close as possible to terrestrial telephone and telex calls. For prompt ship response to calls from shoreside parties, the ship terminal was to always be listening to a call announcement channel for possible call alert with its address when not already engaged in a call. In addition to a ship terminal identification code, each assignment would contain codes for the type of call (e.g., telephone or telex) and terminal instructions (e.g., frequency/time slot tuning for the call).

In order to provide rapid ship to shore call setup, historical maritime polling protocols were abandoned in favor of a random access burst signaling channel design. The design provided for a separate ship to shore frequency on which the ship could send a short request message when it wished to place a new call. The message was sent via a single short digitally modulated burst at 4,800 bps. The use of the "Request Channel" frequency would be on a random basis by all ships accessing the satellite (similar to the "Aloha" protocol). The ship to shore request burst contained only minimal information needed by the coast earth station. It included the ship terminal identity code, the type of channel requested, and priority. The coast earth station would validate the ship terminal's eligibility for service and assign a channel frequency (and time slot if telex) via the shore to ship TDM Assignment Channel.

MARISAT Ship Earth Stations

The ship earth station was comprised of an above decks antenna subsystem and a below decks terminal subsystem. The antenna subsystem was enclosed in a protective radome and included a circularly polarized transmit/receive antenna, a frequency diplexer, a transmit power amplifier, a low noise receiver, and an antenna-pointing/stabilization system. Connected to the above decks unit by signal

cables was the below decks equipment. It included the RF to baseband transceivers, access control interface, telephone termination unit, and telex machine.

Practical considerations (e.g., need to locate antenna high on ship for clear sky view) limited the ship terminal antenna size to something on the order of a meter diameter, even though a larger antenna would have required less power from the satellite. For MARISAT, the compromise was an antenna size of 1.2 m. Since the ship antenna was directional and would require some sort of automatic pointing, each MARISAT could fly in a slightly inclined geosynchronous orbit. This in turn allowed the station-keeping fuel budget to be greatly reduced and more of the satellite mass budget to be allocated to payload. By initially launching into an inclined orbit with a phasing such that the inclination would decrease with time before increasing again, the absolute value of inclination was assured of remaining within 3° over the 5-year-specified design lifetime of the satellite.

MARISAT Multi-coast Earth Station Interworking

The availability of the third MARISAT satellite offered the possibility of commercial L-Band service in the Indian Ocean region. In early 1977, the Japanese company, KDD, decided to undertake the construction and operation of a coast earth station and contracted with COMSAT for the use of the Indian Ocean C/L-Band capacity. In order to insure compatibility, minimum performance requirements for the new station were developed by COMSAT. These were akin to the technical requirements developed earlier for the ship terminals. Acceptance test procedures were also agreed and tests were included to demonstrate compatibility with existing models of ship terminals.

Toward the end of the 1970s, other maritime nations considered implementing a coast earth station (CES) to operate initially via MARISAT before Inmarsat capacity became available. Anticipating this possibility, the MARISAT designers had included a 4-bit "Coast Earth Station ID" code field within the "Request Message" burst that ship terminals transmitted when requesting a channel. Other provisions that had been made to facilitate possible evolution to multi-CES operation included spare codes within the assignment/alerting partition in the shore to ship TDM carriers (Swearingen and Lipke 1976).

To coordinate the sharing of satellite telephone channel frequencies among coast earth stations, one of the stations was designated as the "Network Coordination Station" (NCS). This station transmitted a "Common TDM" carrier that all ships tuned to when idle. For shore to ship calls, the NCS relayed the call alert from the gateway CES via the Common TDM. If the call was a telephone call, the NCS also added the channel frequency pair to be used for the new call in the alert message.

A ship-originated telex channel request was responded to by the desired CES, which then sent a call assignment message on its own TDM. If the ship sent a telephone channel request, the desired CES would send an incomplete assignment message (without a specific frequency pair) in the signaling partition of its TDM carrier. The NCS would then select a specific frequency pair for the call, complete the assignment message, and relay it via the Common TDM.

The NCS also maintained a list of terminals engaged in calls so that a busy signal could be returned to a shoreside party attempting a call when the ship terminal was already busy in a call. In order to insure that telephone channel frequencies were clear after calls ended, the NCS monitored the frequency slots of recently ended calls to confirm that the RF carriers were gone and sent special clearing signals if a slot still had an RF carrier present.

INMARSAT Start-up

In 1982, INMARSAT assumed the role of system manager for the existing commercial maritime system, hitherto known as the MARISAT System. The 1982, transition was managed smoothly with INMARSAT leasing the L-Band capacity on the three existing MARISAT satellites. Using COMSAT's multiple CES interworking design as a basis, INMARSAT adopted its own compatibility standards that enabled existing ship earth stations to continue operation under INMARSAT system management.

Several other satellites with L-Band capacity were nearing completion by the time INMARSAT became operational including ESA's two MARECS satellites and INTELSAT's maritime L-Band payloads on its INTELSAT-V series. Within a few years, all of the planned new payloads were deployed and brought into INMARSAT service. The locations of the ultimate first generation INMARSAT satellite payloads provided coverage from three nominal locations with two payloads near each, one in an operational role and one or more payloads in a standby backup role.

INMARSAT Satellite Capacity Evolution

Because of traffic growth, INMARSAT contracted in 1986 for a new series of Inmarsat-2 satellites with higher power and more bandwidth to be built. The new satellites were successfully deployed in the period 1990–1992 and included bandwidth in Aeronautical Mobile Satellite Service (AMSS) bands as well as additional Maritime Mobile Satellite Service (MMSS) bandwidth. The L-to-C transponder specifications provided for higher signal gain to support new types of lower power ship earth stations. The four INMARSAT-2 satellites offered the possibility of introducing a new western Atlantic operating location near 54° W to complement the eastern location near 15° W.

Higher capacity Inmarsat-3 satellites were launched in 1996–1998 which provide spot beam coverage and incorporate the wider frequency band allocations provided to the three mobile satellite services by the 1987 Mobile World Radio Conference. In order to allow for changes in bandwidth requirements, the transponder passband in each direction was segmented in a variety of widths between 450 and 2,300 kHz. The segments were made switchable between beams and capable of being joined to form contiguous channel passbands significantly wider than the widest segment. In order to accommodate diurnal variations in RF power requirements among the various antenna beams, a matrix power amplifier design was used. Each of the five INMARSAT-3 satellites also includes an add-on navigation transponder that serves to relay a GPS overlay signal and to support the GPS Wide Area Augmentation Service for aeronautical applications.

INMARSAT Ship Earth Station Evolution

In order to reduce power and bandwidth consumption per telephone channel and to take advantage of the wider bandwidth and greater sensitivity of the new INMARSAT satellites, new ship earth stations were needed. The new stations needed to be "future-proofed" so as to be compatible with future satellite spot beam configurations and frequency plan changes.

The first new type of ship earth station was the full capability digital *Standard B*. In order to reduce the required RF power, a 16 kbps Adaptive Predictive Coding speech encoding was chosen. The channel design included a rate ¾ forward error correcting code and offset quadrature phase shift keying (O-QPSK) modulation. The feature of the new Inmarsat Standard B design which "future-proofed" the standard was the "System Bulletin Board" included in the broadcast channel. The System Bulletin Board contained essential information regarding the current spot beam configuration and access control frequencies. Changes are notified via a new page of the System Bulletin Board. The "System Bulletin Board" scheme was subsequently adopted and is used as the technique for ensuring ship terminal compatibility with different generations of satellites and various spot beam configurations. In this scheme, the ship terminal reviews the bulletin board page number to determine whether it has already received the current information or needs to be updated. The information in the bulletin includes the frequency of at least one shore to ship signaling channel for each of the spot beams, and the ship to shore frequencies for each beam to be used for requesting service and for acknowledging call announcements. A portable version of Standard B terminal was introduced for land mobile communications.

The next new type of ship earth station introduced in the 1980s was the *Standard C*. This type fills the needs that require only a minimal messaging capability, but a more compact and less costly terminal (e.g., fishing ships, and yachts). The Standard C incorporates a nondirectional antenna and a low bit rate messaging channel to support basic store and forward messaging services, as well as distress alerting, data reporting, and shores to ship group calling. The Standard C was well received by the maritime community and adopted by the International Maritime Organization as a minimum communications capability component of the new Global Maritime Distress and Safety System (GMDSS) standards for the high seas. (See Fig. 8.2 to see a variety of MSS antennas that have decreased in size over time.)

Several aspects of the Standard C access control design are similar to the Standard B access control design. Shore to ship TDM carriers were used to announce message transmissions to the ships and to deliver messages. Initial ship originated channel requests were transmitted via random access slotted-Aloha channels. Subsequent ship message transmissions were normally transmitted via reserved time slots in a TDMA burst mode similar to the telex transmission modes used for Standards A and B.

Aeronautical Mobile Satellite System Introduction

In the late 1980s with the INMARSAT-2 satellites under construction, a cost-effective Aeronautical Mobile Satellite Service became a realistic possibility,

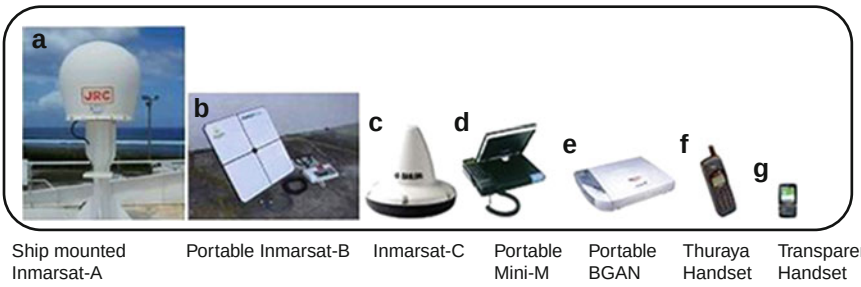


Fig. 8.2 Evolution of MSS terminals (Graphic courtesy of R. Gupta)

provided the aeronautical community was prepared to make use of the INMARSAT-2 satellite capacity. One of the prerequisites was that the aeronautical system design should satisfy requirements established by the Airline Electronics Engineering Committee (AEEC). Consequently, the INMARSAT technical staff participated actively in the working groups of the AEEC to ensure a suitable set of aeronautical system standards.

There was early agreement within the AEEC that data communications would be a "core capability" for air traffic control (ATC) and that an early introduction of such capabilities was important. Furthermore, there was a need for the "core capability" data channels to be compatible with a low gain aircraft antenna. In the forward direction toward the aircraft, the data channels and signaling were to be carried via low data rate (600 bps) time division multiple access (TDM) carriers. In the return direction from the aircraft, data traffic was to be sent on low data rate TDMA channels. Service request signaling message were to be transmitted via slotted-Aloha burst channels. Before the total aeronautical system specifications were completed, an early prototype of the core service was deployed in 1991 for flights over the Pacific. This enabled the airlines to experience the benefits of reliable automatic position reporting for entire flights.

INMARSAT continued to complete the full set of system requirements to support telephony and higher rate data services. These requirements included a requirement for a directive aircraft antenna with auto-pointing capability. The receive sensitivity (G/T = 10 log [antenna gain/noise temperature]) for the high gain aircraft antenna was to be ≥ -13 dB/K, as compared to the -26 dB/K requirement for the low gain "core capability" aeronautical earth station. With the higher gain, telephony service could be supported in a 21 kbps carrier with a 9.6 kbps speech encoding plus framing and rate ½ error correction coding. Higher rate data channels were also to be supported via 10.5 kbps carriers.

By the mid-1990s, aeronautical system capabilities were added to several Inmarsat coast earth stations to enable aeronautical service to be offered in all ocean regions. At the same time, satellite communications avionics for various aircraft were developed and were adopted by several airlines. Since the aeronautical system's introduction, it has been adopted not only by airlines, but also for general aviation by various government agencies.

Portable Mobile Earth Stations for Telephony and Data

In 1988, INMARSAT decided it should develop a new standard that could operate with a smaller antenna and lower power than the maritime Standard B. Key to its introduction was the selection of a more efficient speech encoder. A competition was held that invited developers from academia and industry to submit their encoding/decoding algorithms for testing against agreed test criteria. The competitive test program was completed in 1990 and resulted in the selection of a very robust 6.4 kbps algorithm known as Improved Multi-band Excitation (IMBE).

The new Standard M combined the 6.4 kbps encoded speech output with some signaling overhead and some additional coding to form a 8 kbps channel, which was then modulated with Offset QPSK. With such a low channel rate, the terminal effective radiated power was about 19 dBW under typical conditions and the required receive sensitivity was also greatly reduced so that a much smaller antenna could be used. The typical Standard M terminal was about the size of a standard hard back briefcase, and opened up in a similar manner with the lid containing a flat plat multi-patch antenna.

The introduction of the Inmarsat-3 satellites with their improved sensitivity in spot beam coverage area offered an opportunity to make the portable terminals even smaller than the Standard M terminals. In the early 1990s, INMARSAT revisited the choice of speech encoding design and selected Advance Multi-band Excitation (AMBE) which reduced the data rate to 4.8 kbps. Operating through the more sensitive Inmarsat-3 satellite spot beams, the new Mini-M terminal (see Fig. 8.2) power and size could be significantly smaller than the Standard M. The typical Mini-M terminal is much like a notebook computer with the patch antenna in the raised lid.

By the late 1990s, INMARSAT decided to bring out a service to provide high-speed Internet access via a modified version of its mini-M terminals. The new GAN terminal standard (Inmarsat M-4) incorporated powerful Turbo forward error correction coding and more bandwidth efficient modulation to enable packet data links to the Internet at speeds up to 64 kbps. This service took advantage of the higher sensitivity of the Inmarsat-3 receive spot beams and the higher effective power levels of the satellite.

The benefits of the improved capabilities of the Inmarsat-3 satellite and the signal processing technologies used for the GAN terminals were put to use in INMARSAT's development of new types of maritime and aeronautical communications standards. The new maritime standards (Fleet 77, Fleet 55, and Fleet 33) offered improved data and voice capabilities via three alternative antenna sizes. The new aeronautical service, Swift 64, offered a compact service optimized for data communications.

Inmarsat-4 Satellites and Broadband Global Area Network (BGAN)

Although the Inmarsat-3 spot beams covered all the high traffic areas for the maritime and aeronautical services, they did not provide totally global coverage.



Fig. 8.3 Inmarsat-4 satellite with multiple spot beams (Graphic courtesy of EADS Astrium)

In order to improve and expand the availability of its broadband services, INMARSAT decided to procure three Inmarsat-4 satellites for deployment starting in 2005 (see Fig. 8.3).

Each Inmarsat-4 satellite provides 19 regional spot beams covering the entire visible earth disk. In addition, each Inmarsat-4 supports 200 narrow spot beams. With the greater sensitivity of the Inmarsat-4 narrow spot beams, an improved Internet access service, BGAN, could be introduced with data speeds up to 492 kbps and more compact mobile terminals.

The much higher RF power of the Inmarsat-4 satellites and greater sensitivity of the satellite receivers also have enabled INMARSAT to introduce a handheld telephony service. The handheld phone introduced in 2010, the IsatPhone Pro, is

similar to a GSM cellular phone. The progression of mobile satellite terminals toward smaller sizes (shown in Fig. 8.2) illustrates just how much smaller the terminals can be as a result of introducing more powerful satellites (Gupta 2004).

National Mobile Satellite Systems

USSR

Shortly after MARISAT was deployed, the Soviet Union undertook to deploy a similar system called VOLNA to serve its maritime communications needs. Like MARISAT, it made use of L-Band transponders on geosynchronous satellites stationed at around the earth to provide global service. They included global coverage L-Band transmit and receive antenna beams and used L-Band frequencies below the bands used by MARISAT. Also like MARISAT, the ship earth stations used antennas approximately one meter in diameter and similar system parameters. Over years, the VOLNA program was expanded to include repeaters that also operated in the aeronautical frequency allocations, although there was very little descriptive material about the associated avionics made available to the western public. The Soviet Union's maritime fleets also made significant use of the INMARSAT system, but little is known about how heavily the VOLNA system was used.

Australia

Australia, with its vast under-developed interior land mass, had a serious need to improve radio communications for its land mobile community. To meet that need, the Optus mobile satellite system was developed and its first satellite launched in 1985 with antenna coverage of Australia and New Zealand. The system provided for small transportable terminals similar to the Inmarsat M terminals as well as small vehicular terminals for use while in motion. The Optus system developers made use of speech encoding chosen after a competitive set of laboratory trials and field trials conducted in collaboration with INMARSAT.

Canada and USA

The earliest commercial use of satellites for land mobile communications in the USA made use of existing domestic satellites operating in the Ku fixed satellite service frequency bands. The OmniTRACS system, introduced in 1988, was designed to support the trucking industry in its fleet management. By exploiting the low power spectral density of low data rate code division multiple access (CDMA), the system developer, QUALCOMM, was able to obtain US regulatory approval for the use of fixed satellite frequency bands for mobile service.

In the late 1980s, a company was established to develop a dedicated mobile satellite system to serve Canada called Telesat Mobile Inc (TMI) with mobile telephony and data services. Shortly thereafter in 1988, another company called American Mobile Satellite Corporation (AMSC) was formed to provide mobile satellite services to the USA. The two companies collaborated and procured similar geosynchronous satellites (MSAT-1 and MSAT-2) to enable service restoration in

case of a failure in one of the satellites. Both systems were designed with wide enough frequency bands to enable not only land mobile, but maritime and aeronautical mobile services as well. Unlike the Inmarsat satellites, the MSAT satellites used L-Band frequencies cross-strapped to Ku-Band. The two satellites were successfully deployed in 1995 and 1996 and have been extensively used to support a wide variety of services (e.g., truck fleet management and coastal maritime communications).

Mexico

In 1993, the second generation domestic Mexican satellite, Solidaridad-1, was launched into a geostationary orbit. This satellite was a multipurpose satellite that included a pair of L-Band repeaters to support mobile services for Mexico. The mobile terminal technology used with the Solidaridad system was similar to that used for the Optus, TMI, and AMSC systems.

Japan

In 1996, MSS services were started in Japan with NSTAR satellites at S-Band (2.6/2.5 GHz) operated by NTT-DoCoMo. The feeder link frequencies were at 6/4 GHz bands. Portable land mobile, maritime, and aeronautical services were offered with 5.6 kbps voice and 4.8 kbps data. Subsequently, JCSAT-9, a hybrid C-, Ku- and S-band satellite, was launched in 2006 to augment MSS services at S-band over a wider maritime coverage in Japan.

Low Earth Orbit (LEO) Satellites

Iridium

In mid-1990, Motorola announced its plans to develop a satellite system using 77 satellites in a low earth orbit to enable handheld terminals to communicate from anywhere in the world. The atomic number for the element Iridium being 77, the new system was given that name. However, the system design was modified so as to use 66 operational satellites, but the system name was retained. The Iridium system called for the 66 satellites deployed in six orbit planes at an altitude of 781 km with an inclination of 86.4°. To minimize the number of required gateway earth stations, inter-satellite links are provided using Ka-Band frequencies to enable mobile stations outside the coverage area of the satellites visible to a gateway to have its signals relayed via one or more inter-satellite links. Unlike the other mobile satellites of the time, the Iridium satellites provide for onboard digital packet processing and packet switching to enable the packets to be relayed via the inter-satellite links when needed. The satellites also use a band in the lower L-band segment (1,616–1626.5 MHz) in a burst mode. Technically, the Iridium satellite system was a success with a successful series of multi-satellite launches in 1997–1998. However, the original company with its approximate $5 billion investment was not a success financially and went into bankruptcy. Fortunately for users, the assets were bought by a group of private investors (for approximately $25 million), who provided for continued operation and development of the system.

Globalstar

A different low earth orbit system with 48 satellites called Globalstar was deployed by Loral Corporation in 1998. Unlike Iridium, Globalstar uses simple bent-pipe repeaters on spacecraft flying at higher orbits (1,400 km versus Iridium's 781 km) and a lower inclination (52°). The system design employs code division multiple access (CDMA), which allows seamless satellite to satellite handoffs. Access to Globalstar services is dependent upon proximity to a gateway earth station, since both the mobile terminal and a gateway must be within the field of view of the same satellites for access to work. Globalstar uses L-Band frequencies (in the 1,610–1,621.5 MHz band near those used by Iridium) for mobile to satellite links. For satellite to mobile transmissions it uses a segment at S-band (2,484–2495.5 MHz). This system, like Iridium, also experienced financial difficulties during the early years of its deployment.

ICO

When Motorola and Loral announced plans for the Iridium and Globalstar systems in the early 1990s, INMARSAT carried out a study program of alternative orbits and technologies to support handheld mobile telephony. The study concluded that an intermediate orbit at an altitude of 10,390 km with 10 satellites at an inclination of 45° would provide superior satellite visibility for mobile users. However, the INMARSAT Council (governing board) would not approve the necessary investment by INMARSAT, so a separately funded corporation was created in 1995 named ICO. ICO was not able to launch its satellites before going into bankruptcy in 1999. The assets were bought by a private investor and a new company was created called New ICO. Only one of the ten satellites was successfully deployed before New ICO decided to abandon its plans for a non-geosynchronous constellation in favor of a geosynchronous satellite, which was launched in 2008.

Little LEO Satellite Systems

Some applications such as position reporting and status monitoring require only low data rates and can be supported by small low power satellites. In 1993, Orbital Sciences announced the ORBCOMM system which was designed to address those needs with a constellation of small low earth orbiting satellites. The satellites were to enable communications with very low power VHF transceivers using store and forward onboard message processing by the satellites for delivery to ORBCOMM gateway earth stations.

Geosynchronous Regional Satellite Systems

By the late 1990s, it was technically possible to design a geosynchronous satellite with a very large aperture antenna system capable of supporting handheld mobile satellite communications. Furthermore, because cellular radio coverage continued to have large geographical coverage gaps, there was a need for dual-mode cell phones that could also operate via satellite. Two major projects were initiated, the

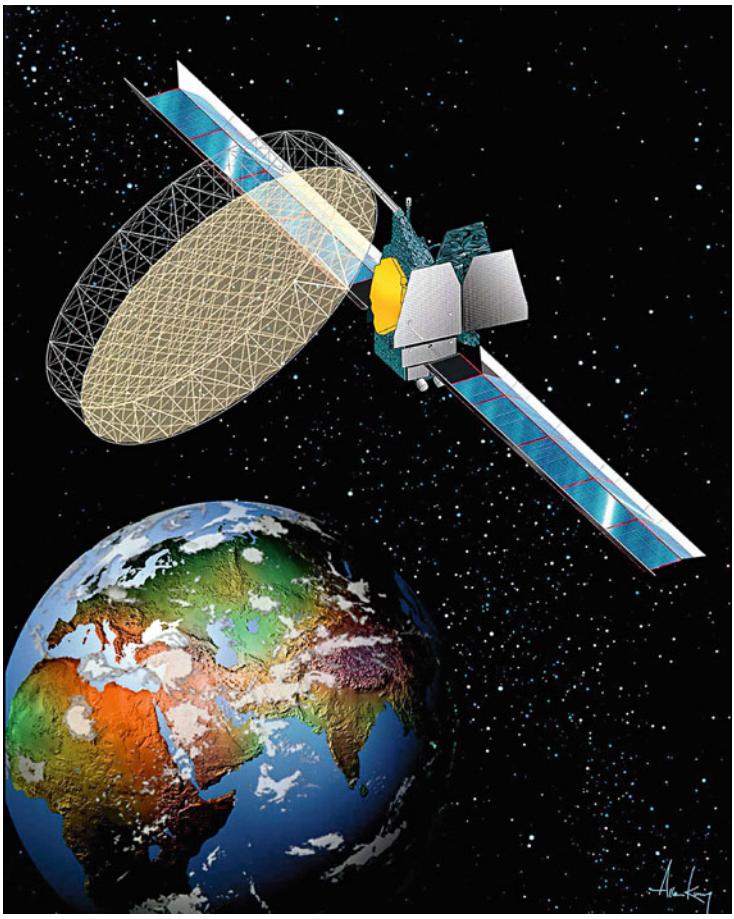


Fig. 8.4 Thuraya mobile satellite (Graphic courtesy of Boeing Corporation)

ASEAN Cellular Satellite System (ACeS) in 1995 and the Thuraya System in 1997. Both projects had a similar goal, to enable mobile users with GSM roaming subscriptions to have access to coverage via satellite when outside terrestrial GSM network coverage. In other words, the satellite system is to serve as an ancillary satellite component to the terrestrial networks.

ACeS was designed to cover Indonesia, Malaysia, Thailand, Philippines, Sri Lanka, Vietnam, China, and part of India. The satellite was launched in 2000 and used two 11.9 m L-Band reflectors to enable 140 spot beams to fill the coverage area from an orbit position of 123° E longitude.

Thuraya (Fig. 8.4) was designed to cover most of Europe, the Middle East, North, Central, and East Africa, Asia, and Australia. The first of the Thuraya satellites was launched in 2000 and used a single 12 × 16 m reflector to enable over 200 spot beams to fill the coverage area from an orbit position of 44° E longitude.

Terrestrial Cellular Service Convergence with MSS

Wireless communications services have emerged as one of the largest growth engines in the telecommunications industry. Second and third generation (2G and 3G) cellular wireless services have been leading growth over past decades, with annual rates exceeding 30%. According to International Telecommunications Union (ITU) statistics released in 2010, there are estimated five billion mobile cellular subscriptions worldwide, with approximately 20% of subscriptions to 3G services.[1] There is a rapid transition to 3G platforms with 143 countries offering 3G services in 2010 as compared to 95 in 2007. As growth rates in Europe and USA have leveled off, Asia and Africa have become very attractive growth markets, with India and China leading the charge. To complement the terrestrial mobile service, a number of satellite-based global personal communications systems (PCS) were deployed. These included Iridium and Globalstar (low earth orbit (LEO) systems) and ICO global (medium earth orbit (MEO) systems and Asia Cellular Satellite Systems (ACeS) and Thuraya (geostationary or GEO) systems). All of these systems provided narrowband voice and data (narrowband) services to handheld terminals, although many of these handheld terminals required special antennas. Some of these systems have encountered major financial problems because of large system start-up costs and slower than required market growth due to a variety of challenges. These have included regulatory barriers and the more rapid than anticipated deployment of terrestrial cellular systems in both the developed world and the developing world.

Recent decades have witnessed a rapid growth in Internet users and growth in broadband Internet traffic driven by demand for data, video, and multimedia services. According to ITU, the number of Internet users in 2010 surpassed two billion with 60% of these being in developing countries. The largest growth contributors have been the mobile Internet subscribers who are rapidly taking over the fixed users (Lee, et al. 2006). These trends have resulted in deployment of new integrated satellite and terrestrial networks using standard devices with form factors similar to current PCS/Cellular devices. Figure 8.5 shows the convergence between emerging wireless and mobile satellite services. Examples include deployment of S-band and L-band integrated MSS networks in the USA by ICO Global Communications (DSDB recently acquired by Dish Network), Terrestar, and LightSquared.

Hybrid Transparent MSS Networks

Until recently, the MSS frequency bands were separate from the bands used for terrestrial cellular, so that the mobile user either needed a dual frequency band handset or two separate handsets. However, in the past decade, several system planners have proposed that segments of the MSS frequency bands be used for both terrestrial cellular and satellite communications, so that the handsets might be simplified and the user's service is always through the same service provider (Karabinis and Dutta). The terrestrial cellular network to support this mode of operation is called the ancillary terrestrial component (ATC). Although this will put additional burdens on the existing

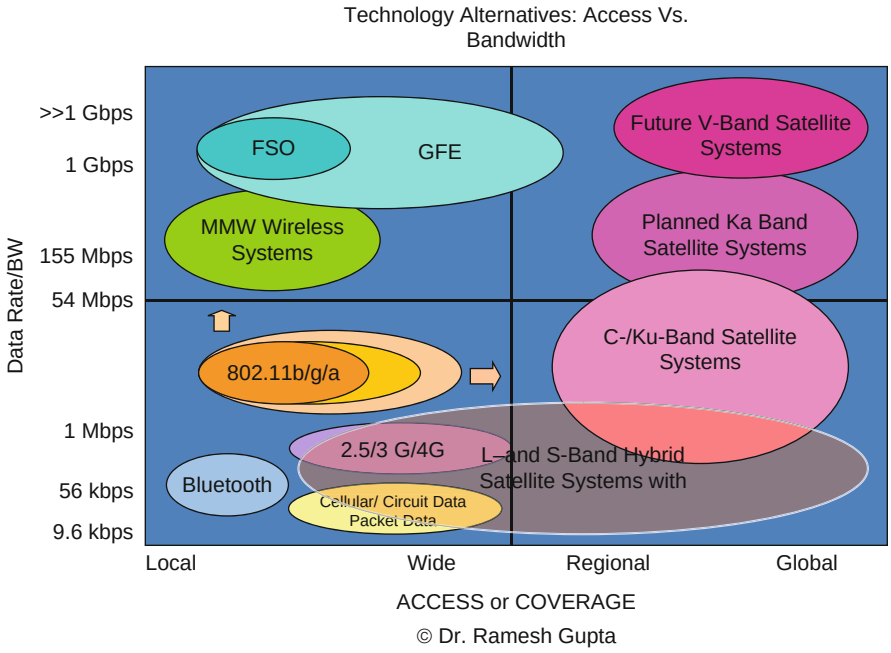


Fig. 8.5 Satellite and wireless service segments and role of hybrid mobile satellite systems with ancillary terrestrial segment (ATC) [MMW- millimeter wave; FSO- Free Space Optics; GFE – Gigabit Fiber Ethernet] (Graphic courtesy of R. Gupta)

frequency allocations and require special precautions to protect GPS operations in adjacent bands, conditional approvals for concept have been already obtained in the USA.

Following the release of US regulatory framework guiding the MSS services with ATC, the European Communications Committee (ECC) decided to review the European regulatory framework for MSS in 1,980–2,010 MHz and 2,170–2,200 MHz bands. In the ECC decision of December 2006, the ground-based infrastructure operating in these bands is named complementary ground component (CGC).[2] Both FCC and ECC required that the MSS operators ensure that interference between these transparency ATC and CGC systems will be considered during intersystem coordination. The regulatory agencies believed that these MSS services will be beneficial to satellite operators and manufacturers, will enable innovative services including provision of emergency services, and will thus serve the public interest.

Mobile Satellite Systems with ATC

In the USA, the satellites for all three licensed systems have already been launched between 2008 and 2010. ICO Global Communications (ICO-G/DSDB) satellite built by Space System Loral (SS/Loral) was launched on April 14, 2008, and placed at



Fig. 8.6 LightSquared SkyTerra-1 mobile satellite (Graphic courtesy of LightSquared)

92.85° W longitude. Terrestar-1, another S-Band satellite built by SS/Loral was launched on July 1, 2009, and placed at 111° W longitude. LightSquared launched its first L-Band satellite (SKYT-1) built by Boeing on November 14, 2010, and placed at 101.3° W longitude. A photograph of the SKYT-1 is shown in Fig. 8.6. The terrestrial segment development and deployment for all three systems is continuing with development of reference chipset designs and firmware for satellite-adapted versions of the chipset for mobile terminals. Terminal equipment for transparent integrated networks (consisting of MSS and ATC) targets a large consumer market, driving economies of scale for chipset as well as device manufacturing.

As an illustrative example, LightSquared next generation system with MSS and ATC is discussed in this section, which consists of two integrated networks: a Space-Based Network (SBN) consisting of satellites and four gateways and an Ancillary Terrestrial Network (ATN) consisting of several ATCs in high-density population centers. The network is designed so that a single handheld device provides seamless two-way data services and voice through terrestrial as well as space segment. The space segment has sufficient antenna gain and EIRP with narrow beams to establish communication with the user devices. For example, SKYT-1 was designed with a 22 m deployable L-band reflector which enables

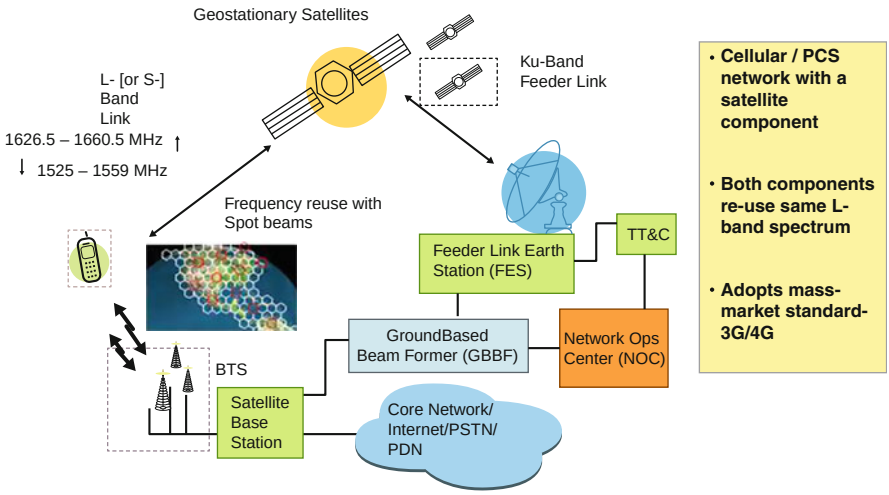


Fig. 8.7 Simplified block diagram of hybrid MSS with ATC (Graphic courtesy of R. Gupta)

the feed element array to form hundreds of spot beams with increased frequency reuse. Network design functions permit seamless transition of device communication from satellite cells (beams) to terrestrial cells, thus achieving transparency to the end user. Satellite provides wide area coverage and is suitable for low-density population areas, where terrestrial infrastructure may not be cost effective. The terrestrial component on the other hand ensures availability of high-speed broadband services in major population centers at affordable cost. This hybrid system offers the availability advantage for seamless communications in situations like earthquakes or hurricanes when the terrestrial infrastructure may be disabled (e.g., hurricane Katrina in 2005 and earthquake followed by Tsunami in Japan in 2011). The satellite capacity and resources (Bandwidth and Power) can be reallocated in case terrestrial infrastructure is disabled by natural or man-made disasters (Editor's Note: The recent decision by the U.S. FCC in February 2012 to disallow LightSquared's ability to operate as a high-speed wireless terrestrial network in the U.S. in the requested frequency band that they had intended to operate–due to interference with the G.P.S. satellite navigation service–will clearly impact this system and its implementation). In fact Light Squared has declared bankruptcy, triggered in part by the frequency interference problem, while ICO/DBSD North America and Terrestar have been bought out of bankruptcy by DirectTV.

A simplified illustration of a next generation L-Band MSS/ATC hybrid network is illustrated in Fig. 8.7. The space segment depicted consists of two L-Band satellites in geostationary orbit, each one with an aggregate EIRP of 79 dBW and G/T of 21 dB/K over a minimum coverage area of 95% so that handheld cellular/PCS devices can close the forward and return links with some margin to spare. The second satellite acts as an in-orbit spare and also provides additional gain for faded mobile channels (because of multipath effects and shadows) through

diversity combining. Four Satellite gateways at Napa, Dallas (USA) and Ottawa, Saskatoon (Canada) are connected to satellites via Ku-Band feeder links, and also provide connectivity to the terrestrial networks. These gateways serve as communications nodes and are connected to the same core public switched telephone network and public data network (PSTN/PDN), thus enabling seamless communications between the satellite and the terrestrial networks.

A new technology is the Ground-Based Beam Former (GBBF), which digitizes the beam signals and divides them into component beams so that appropriate beam weights can be applied digitally. The amplitude and phase beam weights are used together with the onboard satellite feed array, to form beams with maximum gain over the desired coverage and also to suppress side-lobes for minimum co-channel interference as well as to suppress ATC signal-induced interference. The beam forming is also used to minimize interference to and from other systems using same or adjacent L-Band frequencies. The ability and flexibility offered by the GBBF for formation of multiple spot beams with interference suppression is a key design feature of the MSS/ATC networks that enables an increase the spectrum usage efficiency. The network operations center (NOC) monitors the satellite and terrestrial traffic and allocates bandwidth, frequency, and bandwidth resources to minimize interference.

The satellite network is designed to be largely independent of the air interface standards. Therefore, the air interface can be selected based on terrestrial offerings including 4G LTE air interface. The satellite adaptation of the air interface requires the satellite gateway to compensate for satellite delay, and Doppler shifts (because of satellite motion), making the MSS/ATC network independent of air interfaces.

The MSS/ATC integrated network approach is quite different from existing Thuraya, Iridium, and Globalstar MSS networks in that these systems use different frequency bands to provide interworking between terrestrial and satellite networks. A handset for use in an MSS/ATC transparent network is designed to use the same frequency and air interface, so the handset is less complex and therefore likely to be less expensive.

FSS Convergence with MSS

VSAT systems operating in the FSS frequency bands were used for offshore maritime applications, especially in the oil exploration industry. With the development of low-cost antenna stabilization system, VSAT systems were introduced onto cruise ships operating within the coverage areas of domestic and regional FSS satellites. Subsequently, a VSAT became a serious alternative to an MSS portable terminal when the user needed a high data rate link, since bandwidths available in the FSS bands are much wider than in the L-Band MSS allocations.

Wide proliferation of relatively inexpensive and easy to install VSAT terminals with data rates exceeding 2 Mb/s resulted in integration of multimedia services, Internet and video distribution, IP video conferencing, video streaming, and IP multicasting services in the FSS satellite networks. These developments and trends have had a significant impact on the architecture of both FSS and MSS

communications networks, capable of providing "global services," wide area coverage, and emergency rapid response services at relatively low cost (Evans 1994). For example, communications satellites played a major role for disaster recovery operations during tsunami in Asia (2004), aftermath of hurricane Katrina in USA (2005), and also during major earthquake and tsunami in Japan (2011). Regulatory agencies have recognized the critical role of satellites in the global telecommunications infrastructure and have approved new architectures and systems with convergence between the various satellite and terrestrial radio services.

A critical factor in the evolution of radio-communication services for mobile users will be the availability of bandwidth to support higher data rates. Even Inmarsat, with its history rooted in the L-Band allocations, is recently ordered a fifth generation of satellites that is to operate in the 20–30 GHz bands to enable much higher data rates to be supported not only for portable terminals, but mobile terminals as well, such as those on aircraft. However, it should be noted that Inmarsat is not abandoning its L-Band user community and is continuing to also develop follow-on L-Band MSS satellite capacity.

Summary

MSS systems implemented over the past four decades have significantly improved the communications capabilities for the mobile user communities. Vital maritime, aeronautical, and remote area services will continue to be provided by reliable global MSS system operators and several ongoing regional MSS operators. For land mobile services, 3G wireless services continue to be made available to areas outside terrestrial cells with roaming onto MSS satellite systems. With the deployment of ATC, users will have an option when to obtain seamless satellite and terrestrial access from a single device and service provider. For users with very high data rate requirements, more system choices are likely to be introduced with more portable and mobile terminals.

Cross-References

- ► An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ► Fixed Satellite Communications: Market Dynamics and Trends
- ► Satellite Communications Video Markets: Dynamics and Trends

Notes

1. http://www.itu.int/ITU-D/ict/material/FactsFigures2010.pdf
2. ECC Decision of 1 December 2006 on the designation of the bands 1980-2010 MHz and 2170-2200 MHz for use by systems in the Mobile-Satellite Service including those supplemented by a Complementary Ground Component (CGC), (ECC/DEC/(06)09) (2007/98/EC) amended 5 September 2007

References

J. Evans, Satellite and personal communications, in *15th AIAA International Communications Satellite Systems Conference*, San Diego, February/March 1994, pp. 1013–1024

B. Gallagher (ed), *Never Beyond Reach – The World of Mobile Satellite Communications* (International Maritime Satellite Organization, 1989)

R.K. Gupta, Low cost satellite user terminals: lessons from the wireless industry, in *22nd AIAA Conference on Satellite Communications Systems*, Monterey, June 2004

P. Karabinis, S. Dutta, Recent Advances that may revitalize Mobile Satellite Systems. LightSquared, www.lightsquared.com

S. Lee, et al., The wireless broadband system for broadband wireless Internet services. IEEE Commun. Mag. 106–113 (July 2006)

D.W. Lipke, D.W. Swearingen, Communications system planning for MARISAT, in *International Conference on Communications*, Minneapolis, June 1974, IEEE Cat. No. CHO859-9-CSCB

E. Reinhart, R. Taylor Mobile communications and space communications. IEEE Spect. 27–29 (1992)

E.J. Martin, L.M. Keane, MARISAT-Gapfiller for navy satellite communications. Signal Mag. (Nov./Dec. 1974)

D.W. Swearingen, D.W. Lipke, MARISAT Multiple Access Capabilities, in *International Conference on Communications*, Philadelphia, June 1976, IEEE Cat. No. 76CH1086-0CSCB

D. Whalen, G. Churan, The American Mobile Satellite Corporation space segment, in *14th AIAA Conference on Satellite Communications*, Washington, DC, March 1992, pp. 394–404

An Examination of the Governmental Use of Military and Commercial Satellite Communications

9

Andrew Stanniland and Denis Curtin

Contents

Introduction 188
Nationally Critical Satellite Communications 193
The Dedicated Satellites Approach 195
Hosted Payloads or Hybrid Satellites 200
Intergovernmental Agreements 204
Guaranteed, Long-Term Leases and Ad Hoc Leases of Capacity 206
Commercial Satellite Communications Augmentation 209
The Future 211
Cross-References 217
Notes 217
References 219

Abstract

This chapter provides information concerning the use by governments of military and commercial satellites systems for strategic and defense purposes. It discusses dedicated communications satellite systems designed for particular uses and the so-called dual use of commercial systems to support military and strategic purposes. It explains various pathways that can be followed by governments to obtain communications satellite services to support military uses. These paths include: (1) dedicated satellites, (2) hybrid satellites (both military and commercial payloads on a single satellite), (3) shared satellite facilities via

A. Stanniland (✉)
Paradigm Secure Communications Ltd, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2AS, UK
e-mail: andrew.stanniland@paradigmsecure.com

D. Curtin
Eagle Group Partners, 14105 Arctic Avenue, Rockville, MD, 20853, USA
e-mail: dns.crtn@gmail.com

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_8, © Springer Science+Business Media New York 2013

intergovernmental agreements, (4) guaranteed long-term leases, (5) ad hoc leases of capacity on demand, and (6) a long-term partnership between a government and a commercial partner as is the case with the Skynet 5 program in the United Kingdom.

In this chapter the authors will also examine how various countries obtain their national satcom, how and why commercial capacity has become, and will continue to be, a significant part of national satcom capabilities. It will examine the present and future contracting approaches and procedures used in various countries, but primarily in the United States and other NATO countries. Finally there will be a discussion of the issues involved when nations decide between purchasing nationally owned satellites and leasing capacity commercially.

Keywords

Ad hoc leases • Advanced Extremely High Frequency (AEHF) Satellite • ANIK Satellite of Canada • Athena-Fidus Joint French and Italian Satellite • CBERS satellites of China • COMSATBw of Germany • Dedicated Satellites • Defense Information Systems Network Satellite Transmission Services-Global (DSTS-G) contract • Defense Satellite Communications System (DSCS) • Dual Use of Commercial Satellites • End-to-end Services • European Defence Agency (EDA) Satellite Communications Procurement Cell (ESCPC) Global Broadcast Service (GBS) • Haiyang Satellite of China • Hispasat • Hosted Payloads • Hybrid Satellites • Joint Tactical Radio System (JTRS) • LEASAT • Marisat • Milstar • Ministries of Defense (MODs) • Mobile User Objective System (MUOS) • NATO • NATO NSP2K • Paradigm • SICRAL Satellite of Italy • Skynet • Syracuse Satellite System of France • Tranformational Satellite (TSAT) System • Transformational Communications Architecture • TURKSAT • Turn-Key Services • UHF Follow on (UFO) • Unmanned Aerial Vehicles (UAVs) • Wideband Global System • X-band • XTAR • Yahsat • Yaogan of China • Zhongxing Satellite of China

Introduction

Communications have been critical to military organizations for centuries. With the advent of the use of space and the increased sophistication of military forces it was natural that the military would look to space to help supply their communications needs. After the launch of Sputnik in 1957, it was clear that satellites offered opportunities for all types of applications including communications. Almost immediately military organizations around the world began to look at ways to use satellites to provide communications to their forces. Over the last 50 years, a variety of satellites, deployed into different orbits, have been developed by many nations to help meet their government's communications needs.

Governments tended to focus on the geostationary orbit because of its specific beneficial characteristics for supporting communications. A single satellite

deployed into geostationary orbit provides visibility of almost a third of the earth's surface. In addition, the most distinct characteristic of the geostationary orbit is that satellites placed there have the same orbital period as the earth and appear to be fixed above the same point on Earth and therefore satellite terminals on the ground do not have to track the satellite's movement. This means that the terminal will be lower in cost than a more sophisticated terminal with the ability to track a moving satellite, and can also be much easier to use. Ease of use and minimizing training requirements are key factors that are considered when implementing infrastructure for military forces.

The United States, Soviet Union, and some NATO countries initially set about procuring their own, unique, national satellite systems in the 1960s when the field of satellite communications was just being pioneered. These same nations also played a key role in the start-up of the commercial satcom industry. The United States was, and still is, the leading pioneer in the military Satellite Communications (satcom) arena. The United States, alone now has multiple constellations of satellites in geostationary orbit, with the total number estimated at 25 spacecraft.

A number of different contract approaches have been used to obtain defense-related communications satellite capacity and a whole industry has been developed around providing communications requirements for military purposes. In some cases, industry has provided the military user with full end-to-end (sometimes known as turn-key) services including the capacity leased from a commercial provider, terminals, other hardware and operation of the full service to supplement the military capacity.

This chapter will discuss a number of the satellite solutions that have been developed and deployed by different countries and how some of these countries have procured their military communications. It will also discuss the continuing emergence of new and diverse communications requirements, the issues raised by these new demands and the approaches utilized to satisfy them. It will also look at what the future might bring for both the military and the commercial industry.

A good summary of the history of both commercial and defense-related satellites is discussed in an earlier chapter of this book entitled ▶ Chap. 3, "History of Satellite Communications." The authors refer the reader to that chapter rather than repeating the history here.

From the mid 1960s until the mid 1990s, commercial satellites and military satellites were on different development paths. Military systems were initially designed to meet the demands of the Cold War. Starting in the 1980s, there were only a few dedicated satellite services designed to support tactical level forces via a small number of satellite ground terminals. Such systems began to be fielded in larger numbers, predominantly by the United States, to tactical level forces during the middle to late 1980s.

The generation of military communications satellites available for use during the 1990s and 2000s were, for the most part, defined, designed, developed, and produced beginning in the late 1980s. This planning was largely carried out before the widespread advent of cell-phones, the ubiquity of personal computers, the evolution of the Internet to a public global utility and the subsequent innovation

of what became known as the World Wide Web. Technology innovations developed specifically for commercial systems could be, and often were, adopted by military systems and vice versa.

This result was unsurprising since the firms that built the satellites built them for both military and commercial customers. Commercial systems being developed during the late 1980s were evolving faster than their military counterparts to meet the needs of international telephony and the broadcast industry, especially with the advent of using communications satellites to broadcast television channels direct to people's homes.

During the Cold War period, the vast majority of the United States and NATO defense-related communications traffic was carried by a country's own national satellite(s) with some minor military traffic being carried by commercial satellites. This began to change in the 1990s as more international conflicts started to occur simultaneously in different geographical locations, for example, the Bosnia, Kosovo, and Gulf War I conflicts between 1990 and 1999. This evolution in the operational context meant that the existing military satcom systems were becoming outdated and struggling to meet the new demands as flexible coverage and increasing throughput to support a wider range of applications became the order of the day.

By the beginning of the second Iraq war in 2003, the bandwidth requirements had increased to the point where the available capacity of the defense satellites was no longer able to meet the demand. This resulted in the US DoD and other NATO Ministries of Defense (MODs) turning to commercial satellite capacity to meet the shortfall. This was a fundamental and far reaching change, which has continued and indeed greatly increased since 2003.

Due in no small part to budget constraints and delays in satellite procurement programs, as well as the greater demand for capacity, the amount of on-orbit defense satellite capacity in the United States and elsewhere has not been able to meet the demand. In 2004, the US DoD began acknowledging that commercial satellite capacity was providing over 80% of US satcom bandwidth that the US military used. This continued to evolve in the 2006–2007 time period and during this period over 95% of the Satcom used in the US Central Command (CENTCOM) area of responsibility was provided by commercial systems (DISA conference proceedings 2009).

This trend has been advanced by the growing sophistication of the communications devices employed by the military, including communications on the move terminals, man-packs and the increasing use of Unmanned Aerial Vehicles (UAVs) with sophisticated sensors and cameras necessitating a large amount of capacity to transmit the data to both tactical and garrison facilities where it can be stored, processed, and analyzed. It is predicted that these trends will continue and that even with the launch of additional DoD and other countries' government satellites, the vast majority of the defense-related bandwidth requirements will be met by commercial satellites for the foreseeable future.[1]

The US DoD employs commercial capacity more frequently and in greater volume than any other country or alliance, including NATO. The implication of this fact for the US DoD is that commercially procured satellite communications can no longer be considered just an adjunct to military communications but must be

regarded as an integral part of the warfighters' communications inventory and must therefore be treated as a critical part of a nation's infrastructure. The Transformational Communications Architecture (National communications system fiscal year 2007) and the Joint Space Communications Layer (JSCL) Initial Capabilities Document (Satellite 2001) both state that commercial satcom is now an integral part of DoD's overall satcom capability portfolio.

This fundamental shift means that many commercial satellites currently in-orbit, as well as those being developed, will often be dual-use satellites with a significant portion of their capacity employed for noncommercial and often very sensitive communications. In some cases (for example, the XTAR X-band satellites), the satellites are developed specifically for the provision of government services but under commercial terms. In another example, a portion of the UK military Skynet 4 and 5 satellites, owned by Paradigm Secure Communications Limited, has been specifically designated under the UK MOD agreement with Paradigm to providing services to both the US DoD and other MODs under commercial terms.

The widespread use of commercial capacity by the DoD and others has led to some difficult questions for defense users and policy makers as well as for commercial owner-operators and commercial service providers. Likewise it also poses significant issues for defense contractors that manufacture and develop unique technologies for military satcom satellites. Such defense contractors have made the development of dedicated military satellites one of the cornerstones of their order book for over 40 years. Questions that now arise include the following: Can defense budgeting continue to fund extensive capital procurement programs? Can capital expenditure budgets be re-focused to fund a leased solution? What type of traffic can be sent over commercial capacity for the longer term? Are commercial systems reliable enough for sending sensitive traffic? Is commercial encryption sufficient? Can a commercial operator be trusted to guarantee communications at all times and in all places? Is the lack of nuclear, and other physical, hardening measures an issue given the high percentage of defense-related traffic on commercial systems? Where are the teleport facilities of the commercial operator? Will any military traffic be "landed" there? What control will the military operators have over the routing of their traffic, if any? Will supporting military operations have any knock-on effects for the day to day business of a commercial operator? Can a commercial operator separate out military and commercial traffic within its operations and its business processes? Can industry build military payloads that fit in with commercial operator business plans?

Solutions are already being found for many of the above questions. Clearly the nationally owned defense satellites will continue to be used for the most critical traffic by most nations but, with the growing demand from users, the nationally owned resources can no longer be relied on to have sufficient capacity and, therefore, it is inevitable that some sensitive traffic will have to be transmitted commercially. It is expected that as new commercial satellites are developed, the United States and other governments may ask that those satellites which could be

used partially for government purposes be fabricated with some additional attributes such as, for example, specific frequency bands, a limited jamming resistance or steerable antenna.

Various approaches have been used over the years to acquire the necessary capacity. In some countries with relatively small requirements, such as Spain with the original HISPASAT satellites, or Turkey with the TURKSAT satellites, the military has already added X-band or other military frequency transponders to commercial national satellites. Other countries like the United Kingdom and France have launched national multi-mission defense satellites having a combination of UHF, X-band, and EHF transponders and are following this up with sophisticated outsourcing programs. In the United States, the government programs have largely employed satellites dedicated to a single mission and frequency band although, beginning in the 1970s and 1980s, the US government also leased capacity on satellites such as the MARISAT and LEASAT satellites where commercial operators had added specific capability to commercial satellites for government use. This then paved the way for the future use of commercial satcom assets for military requirements.

In addition to leasing commercial capacity from a commercial operator, another approach to providing this needed capacity is through the use of Hosted Payloads where a payload for a specific government purpose is designed, built, and installed on a commercial satellite planned or under construction. The hosting of payloads is not a new concept: the US Government has hosted payloads on other Government satellites for decades, as stated above with the LEASAT program dating from the late 1970s, but increasingly hosting of military payloads has found its way back into the military satcom policy makers' thinking over the last decade. More recently, there have been a number of meetings between industry and the US DoD discussing the technical issues surrounding Hosted Payloads, as well as different ways to overcome any contractual, coordination or other issues that might delay the implementation of Hosted Payloads.

The commercial satellite industry is constantly evaluating what capacity will be needed in the future, with what frequencies and with what spot or zonal beams that could be available at different orbital locations. The industry's planning would be more accurate if it knew when and where defense-related capacity would be required and for what period of time the requirements would endure. This is a difficult question for any government to answer as operational requirements are not only often unpredictable but are also likely to be classified. How does a defense ministry contract for capacity servicing a specific geographical location for several years if there is a likelihood that the military situation at that location might change and the capacity would no longer be needed, or might be needed elsewhere? How does a commercial satellite operator manage the transient nature of the military requirements with the fact that the commercial mission is more than likely to be focused on a specific population area, which prevents the satellite being moved to a different orbital location? In the United States, the commercial satellite industry and US Government officials regularly meet to discuss the best way to frame and work through these types of issues.

Nationally Critical Satellite Communications

The justification behind procuring highly survivable, dedicated military systems, which often represent much higher cost facilities, versus the leasing of commercial satellite capacity, has become a highly contentious issue for procurement agencies worldwide. More nations are looking to implement dedicated military satcom systems than ever before. While some are satisfied with an initial reduced capability system in order to obtain a limited military satcom capability before upgrading to more capable assets in subsequent phases of infrastructure rollout, many nations are choosing to implement state of the art resources at the first attempt in the understanding that their requirements will expand to fill the available capability (e.g., UAE and Norway).

At the beginning of 2011, the main players in the field of designing, building, and procuring dedicated military satellites, or dedicated military payloads that are owned and operated by the military, are China, France, Germany, Italy, Russia, Spain, the United States, and the United Kingdom (although as will be explained later, the UK capability has now been outsourced as part of the Skynet 5 program).

Several other countries have developed a dedicated military satcom capability, typically by adding a military frequency payload to a nationally owned and operated commercial satellite. Among other countries, Australia, Brazil, Japan, South Korea, and Turkey have done this. Usually one or more small military satcom payloads on national satellites are sufficient to meet the national defense needs of the particular country rather than paying for dedicated, and sophisticated, military satcom satellites as are required by other countries with greater and more demanding national interests, associated military responsibilities and resulting communications requirements. This approach has been used successfully for the last 20 years and it is expected that additional countries who believe that they have a need for a specific, but limited, military satcom capability will either add military frequency capacity to a national commercial satellite or possibly arrange for a hosted payload on an international commercial satellite under development (for example, Australia's successful negotiation to have a commercially built UHF Hosted Payload included as part of the Intelsat 22 satellite).

Arguably the most critical attribute of military satcom systems is the confidence that the military communicator has to have that their communications equipment will work – always and on immediate demand. These systems are used by nations to support both peacekeeping and hostile deployments and, therefore, there is always a threat of hostile (or accidental) interference to the radio signals and a physical threat to either the ground or the space-based equipment, coupled with the political threat that the nation providing any commercial communications may not agree with the military activity being pursued and seek to deny access to the communications.

Nations with military satcom capability have analyzed their individual security and survivability requirements and typically followed one of five paths to secure access to guaranteed communications for their militaries:

- *Dedicated satellites*: These are satellite that are specifically designed, procured, and launched by the government itself. The whole satellite is solely for military or governmental purposes and both the payload and the platform can be designed to satisfy the demands for security and survivability.
- *Hybrid satellites*: This type of satellite has a payload that is designed and procured by the government but launched as a co-payload on a commercial satellite. The payload can satisfy the security needs, but the platform is normally built to commercial standards to keep cost to a minimum. Marisat, LEASAT, and Telecom are early examples. These are the forerunners of today's Hosted Payloads.
- *Intergovernmental agreements*: Under this type of arrangement, countries who are natural allies enter into agreements to provide each other with dedicated (or backup) communications capability as an alternative to the procurement of stand-alone capacity or infrastructure.
- *Guaranteed, long-term leases*: This type of lease provides for assured access to communications that are fully or partially guaranteed by agreeing to a long-term reservation or usage contracts with commercial operators. Usually the protection of the communications capacity itself cannot be guaranteed and so this approach is often favored by those nations with either a low threat assessment or who can rely on allied military satcom for requirements with a higher threat assessment.
- *Ad hoc leases of capacity*: This type of arrangement provides excellent value for money since a nation only pays for what it uses but there is no guarantee that the capacity will be there, or what it will cost, unless it uses a national commercial provider as its conduit. There will also be no guarantee of any information assurance features on the capacity procured.

Since the advent of the Skynet 5 program in the United Kingdom, a sixth option has emerged – a formal, contractual *partnership between government and industry* to provide a service-based approach for both commercial and military communications. In this case the total need is satisfied by a commercial company but this is accomplished via a mix of guaranteed access to protected military communications capacity that is owned and operated by the commercial service provider and long-term commercial leases managed and secured by the service provider.

The lead taken by the United Kingdom in this arena is indicative of the trend in Europe where decreasing defense budgets are forcing governments to explore financial and commercial innovation with just as much rigor as technical innovation. For example, the military satcom systems of France, Germany, Norway, Italy, Spain, and the United Kingdom have all been procured using different commercial and financial methodologies by the respective national defense departments. The next generation (beyond 2020) European systems are expected to further explore methods of increasing defense budget utility with increasing international consolidation but today each nation is ensuring that its diverse geographical and interoperability requirements can be met by a nationally procured solution using a multifrequency payload. In Japan the government is currently exploring how it can best provide future military communications capacity as well.

The Dedicated Satellites Approach

This section examines some key examples of dedicated defense-related satellite systems that have been implemented.

NATO

NATO has been a user of military satcom since 1970 and has owned and operated the NATO 1, 2, 3, and 4 series of satellites. The NATO 4 (sometimes designated NATO IV) satellites, which were launched in 1991 and 1993, were built to the same design as the UK Skynet 4 series. NATO changed its approach to the procurement of military satcom in 2004, as will be discussed later.

United States

The United States currently possesses the largest number of military satcom satellites and it has developed into a nation with multiple constellations of satellites. These constellations tend to be frequency specific. This often results in an additional, alternate payload of lower capability being use to provide cross compatibility with the frequencies employed on other constellations. The United States has divided its communications into four elements:

- Narrowband, unprotected communications using UHF
- Wideband communications with limited protection features on X-band and, more recently, Ka-band frequencies
- Protected communications with full hardening and survivability features using EHF frequencies
- Leased commercial satellite communications using L-band, C-band, Ku-band, more recently X-band and UHF (though these commercial capabilities are not the subject of this part of the chapter)

Despite being an extremely limited service in terms of throughput, UHF has become an enduring technology for troops worldwide due to its utility for highly mobile, deployed forces. This utility is unlikely to change in the future despite the advent of handheld commercial satcom systems such as Iridium, Thuraya, and Globalstar.

UHF services are currently provided by the national *UHF Follow On (UFO)* system[2] operated by the US Navy, augmented by long-term leases on LEASAT 5 and Skynet 5C. The *Mobile User Objective System (MUOS)*[3] also operated by the US Navy, is a geostationary communications satellite network designed to progressively replace the UFO satellite constellation from 2012 onward. The MUOS satellites integrate third-generation (3G) commercial cellular technology, 3G Wideband Code Division Multiple Access (WCDMA) waveform, and Universal Mobile Telecommunications System (UMTS) infrastructure.

The MUOS constellation will provide UHF secure voice, data, video, and network-centric communications in real-time to US mobile warfighters through 2030 and will be fully interoperable with the Joint Tactical Radio System (JTRS) and current radio systems. The system will maintain compatibility with UFO system and legacy terminals by the inclusion of a UFO legacy payload on the

earliest MUOS satellites. Individual terminals (users) will be able to access up to a 64 kbps link.

The majority of the US military communications are supported using the *Defense Satellite Communications System (DSCS)*[4] constellation and its successor, the *Wideband Global System (WGS)*.[5] In contrast to UFO and MUOS, both of these constellations are operated by the US Air Force. The first DSCS III was launched in 1982 and the first WGS in 2007 and they provide global wideband coverage to all US forces. As the backbone of the US military's global satellite communications capabilities, the DSCS constellation provides nuclear-hardened, jam-resistant communications in support of US Strategic capabilities and, high data rate, long haul communications to users worldwide.

The last DSCS III satellite (DSCS B6) representing some 14 consecutive successful launches was placed in orbit in August 2003. In 2010, nine DSCS satellites were still operational and continue to be used to support military operations around the globe and research and development of ground-based support capabilities. The last four DSCS satellites received modifications to provide substantial capacity improvements through higher power amplifiers, more sensitive receivers, and additional antenna connectivity options.

The replacement to DSCS, WGS, is a satellite communications system which was originally conceived as an interim system to meet the military needs of the first decade of the twenty-first century and that provides flexible, high-capacity communications for US warfighters. WGS provides a quantum leap in communications bandwidth over DSCS. Although one key difference is that while the DSCS satellites included technical features to make the success of denial of service threats by enemies more difficult, the DoD decided to design the WGS satellites without such features, relying instead on the Milstar and future AEHF systems to support the highly survivable communications requirements. The WGS satellites are therefore nearly identical to commercial satellites from the early 2000s in terms of their ability to respond to denial of service threats.

The WGS satellite system, originally designed as a constellation of three satellites, has now been expanded to at least six satellites and provides service in both the X-band and military Ka-band frequency spectrums. As well as replacing the DSCS X-band communications and the *Global Broadcast Service (GBS)*[6] one-way Ka-band service, WGS provides a two-way Ka-band service for US DoD users for the first time. The first WGS satellite entered service in 2007 with WGS 2 and 3 following in 2009. The second batch of WGS satellites will have a modified Ka-band payload configuration specifically intended to improve throughput for Unmanned Aerial Vehicles.

WGS will support US DoD high data rate Intelligence, Surveillance, and Reconnaissance (ISR) requirements as well as tactical warfighting units, many of the ISR requirements have hitherto been supported using commercially available capacity due to the shortfalls in high data rate capable capacity.[7,8]

Finally, the highly survivable and protected national communications requirements are supported on the *Milstar*[9] and *Advanced Extremely High Frequency* (AEHF)[10] systems.

Milstar is a joint service satellite communications system comprising five satellites launched between 1994 and 2003 (an additional satellite was lost on launch). The system provides secure, jam resistant, worldwide communications to meet essential wartime requirements for high priority military users. An important difference between Milstar and AEHF and other military communications satellites is that each satellite processes the communications signal within the payload, restoring the signal to its original form, and serving as a smart "switchboard in space" which can direct traffic from terminal to terminal anywhere on the Earth. Inter-satellite links further reduce the requirement for ground controlled switching. Milstar can support individual user link data rates from 75 bps through 1.5 Mbps.

AEHF provides global, secure, protected, and anti-jam communications for high-priority military ground, sea, and air platforms. The system will consist of at least 4 satellites in geosynchronous earth orbit (GEO) and will provide 10–100 times the capacity of the Milstar satellites which they will eventually replace, with maximum data rates on individual user links up to 8 Mbps instead of the 1.5 Mbps possible from Milstar. Without question the AEHF System is the most complex satcom satellite now in service for assuring communications to US military forces and is, in terms of survivability and security capabilities, the most advanced military communications satellite in the world (Fig. 9.1).

The first AEHF satellite was launched in August 2010. However, the satellite failed to initially achieve geosynchronous orbit due to a malfunction in the liquid apogee engine and was left in a low earth orbit. Utilizing the other thrusters on board the satellite, AEHF Flight 1 (AEHF F1) was subject to a long duration orbit raising exercise to raise this satellite to its correct geostationary orbital slot at 68° West Longitude. It was found that the liquid apogee engine had a malfunction rather than a design flaw; consequently, preparations and plans continue toward the launch of AEHF F2.

The AEHF system will be used not only by the US DoD for its highly critical communications links but also by a multinational consortium of allies who have all invested in the satellites and the associated ground systems. These nations (Canada, Netherlands, and United Kingdom) will be granted access to a specific amount of capacity across the AEHF constellation and will purchase appropriate terminal and teleport equipment to be able to access the system via government to government agreements.

The above summarizes the US DoD's operational military satcom systems; the future of US military satcom beyond MUOS, WGS, and AEHF was intended to be satisfied by a program called *Transformational Satellite Communications System (TSAT)*. This program was a US DoD program to provide high data rate military satcom and Internet-like services. TSAT was planned as a five satellite constellation with a sixth satellite as an in-orbit spare, with the first launch in the 2019 time frame.

An extremely ambitious project, utilizing many state of the art space borne technologies, TSAT was intended to ultimately replace the DoD's current satellite system and supplement the constellation of AEHF satellites. It was designed to



Fig. 9.1 Advanced EHF Satellite (Image taken from US Air Force website; U.S. Air Force fact sheet advanced EHF system, http://www.afspc.af.mil/library/factsheets/factsheet_print.asp?fsID=7758&page=1)

support net-centric warfare and would have enabled high data rate connections to Space and Airborne Intelligence, Surveillance, and Reconnaissance (SISR, AISR) platforms. The total RF throughput projected for the TSAT Program was more than ten times that of the AEHF system.

In April 2009, after almost $2 billion (US) of R& D expenditure and 6 years of development, Secretary of Defense Robert M. Gates asked that the project be canceled in its entirety.[11] High cost, technological risk, and development delays were given as primary reasons.

As an interim replacement strategy, Secretary Gates recommended the procurement of two additional AEHF satellites, bringing the total constellation to four satellites. Although some industry analysts would say that the cancellation of the TSAT program was inevitable in the current US defense budget climate, it is clear that the decision is already having an effect on the future of dedicated military satcom programs around the world. The appetite for governments to fund the design and development of quantum leaps in technology and capability is decreasing, which in turn is forcing industry and the military alike to examine the potential for incremental capability increases along with more innovative use of existing technologies. It is possible that a more advanced designed for a dedicated US

defense satcom program may be restarted in future years, although this does not seem to be a near-term prospect.

United Kingdom

The United Kingdom has been a military satcom user since the late 1960s and the Skynet 2 satellites were actually the first communication satellites to be built outside either of the United States or the USSR. Because of the breadth of geographical coverage needed by the UK Armed Forces, the United Kingdom has always opted for a multi-satellite constellation – despite its relatively small size. As stated earlier, the Skynet 4 design was reused by Matra Marconi Space Ltd as the basis of the NATO IV series of satellites, thus guaranteeing interoperability between the United Kingdom and NATO satcom equipment. Skynet 4 was the last series of UK satellites to be wholly owned and operated by the UK MOD. The advent of the Skynet 5 PFI program has transitioned this responsibility into industry and will be discussed in more detail later in this chapter.

France

France has been a member of the military satcom community since 1980 with Syracuse 1, but, until the advent of the Syracuse 3 constellation,[12,13] had followed the hybrid satellite route, sharing satellites with the "Telecom" commercial payloads, which were owned and operated by France Telecom. Syracuse 3A was launched in October 2005 and Syracuse 3B followed in August 2006. The Syracuse 3 series is hardened and protected to NATO standards, similarly to the UK Skynet series, and unusually for European military satcom does not have a narrowband UHF payload on board, concentrating instead on military X-band and EHF frequencies to support the French military.

In 2007, a third spacecraft was expected to be ordered, but this was canceled in favor of including the Syracuse-3C payload on the Italian SICRAL 2 satellite, ushering in a new era of allied collaboration which will no doubt have far reaching impacts on the whole of the military satcom arena.

Germany

Germany has only recently entered the military satcom arena with its own dedicated assets, relying for many years on NATO capacity, intergovernmental agreements and commercial leases. In 2006, the German Bundeswehr awarded a contract for the construction of two satellites for narrowband and wideband communications, a comprehensive ground user terminal segment and the upgrade of the network management center to a special company set up to deliver the capability, MilSat Services GmbH, which was a joint venture between EADS SPACE Services and ND SatCom.[14]

COMSATBw 1 and 2 were launched in October 2009 and June 2010, respectively, and are now fully in operation and owned by the Bundeswehr. However, their operation is carried out by MilSat Services GmbH, who is also responsible for the ongoing maintenance of the ground network as well as the long-term leases of any required commercial satellite capacity.

Italy

Italy entered the military satcom arena in 2001 with the launch of Sicral 1A[15] into geostationary orbit, providing UHF and X-band capacity to Italian armed forces. Like France and the United Kingdom, Italy has been a mainstay of the delivery of X-band and UHF capability to NATO forces since the signing of the NSP2K Memorandum of Understanding in 2004 between the Ministries of Defense of Italy, France, the United Kingdom, and NATO. This agreement meant that the then planned Sicral 1B satellite (which was duly launched in April 2009) was now even more critical for Italy.

As mentioned earlier, Italy and France have now jointly embarked on the SICRAL 2[16] program, which is expected to enter service in 2013.

USSR (and now Russia)

The USSR was the first country to orbit a satellite in 1957. The Soviet Union, and now Russia, has been very active in using their space-borne capability. It is reported that between 1960 and 1990 the vast majority of Soviet satellites that were launched carried military payloads, even though until the last decade of the twentieth century there was no official acknowledgment of a military space program. During the first decade of the twenty-first century, Russia has continued its launch program and now identifies specific military satellites but with no specific information as to individual missions.[17]

China

China launched its first satellite in 1970. Since then its satellite activity has increased, particularly in the last decade of the twentieth century and the first decade of the twenty-first century (Annual report to congress 2010). China has a large program of both reconnaissance and communications satellites utilized for military purposes:

- *Reconnaissance Satellites*: China continues to deploy imagery, reconnaissance, and earth resource systems that can also be used for military purposes. For instance, the Yaogan 1 through 6 satellites, the Haiyang 1B, the CBERS-2B satellite and the eight planned Huanjing disaster/environmental satellites are capable of visible, infrared, multispectral, and synthetic radar imaging.
- *Communications Satellites*: China utilizes communications satellites for both regional and international telecommunications supporting both military and commercial users somewhat like a number of other countries. China also operates a single data-relay satellite, the TianLian-1, launched in 2008. Most recently, China launched the Zhongxing 20A dedicated military communications satellite into geosynchronous orbit from a Long March 3A launch vehicle in November 2010, making that launch the 14th successful Chinese space launch in that year (Long March launch of Chinese Military Satellite 2010).

Hosted Payloads or Hybrid Satellites

For nations who do not have the budget or the overall requirement for their own military communications satellite, then a more limited payload is often the best

solution. However, since the satellite bus or platform and launch costs are the dominant factors in procurement of satellite capability, countries often search for less expensive alternatives. One such idea is the *Hosted Payload* concept. The term has been developed to refer to the utilization of available space on a commercial satellite's platform to accommodate additional transponders, instruments, or other items needing to be orbited. The original *Hybrid Satellite* concept is now more generally taken to mean a satellite developed with both the commercial and military payloads in mind from the start. Where the commercial operator and the government originate from the same country, these are often referred to as hybrid satellites (for example, France, Australia, and Japan have followed this approach) and where the military payload is opportunistically launched on another entity's commercial satellite they are referred to as Hosted Payloads (for instance, the Intelsat 22 UHF mission for Australia[18]). Either a hosted payload or a hybrid satellite can be interpreted to mean that a specific satellite fulfills multiple missions for different customers. For this chapter, we will treat the terms interchangeably for simplicity.

By offering "piggyback rides" or "hitchhiking" opportunities on commercial spacecraft already scheduled for launch, satellite firms allow organizations such as government agencies to have sensors and other equipment launched into space on a timely and cost-effective basis. The Hosted Payloads concept is similar to the ridesharing or multiple manifesting launch concept, but instead of sharing a space launch vehicle, the partners share a satellite bus (Fig. 9.2).

Hosted Payloads allow the government to plan, develop, and implement predefined space missions on much shorter cycles compared to the time it takes to design, procure, and launch an entire government satellite – typically 24 months versus many years. This is especially important for agencies facing impending gaps in operational capability. The partnership with the commercial satellite firm gives the government an opportunity to leverage an already planned or existing satellite bus, launch vehicle, and satellite operations.

Placing a Hosted Payload on a commercial satellite costs a fraction of the amount of effort required for planning, building, launching, and operating an entire satellite. The commercial partner only charges for the integration of the payload with the spacecraft and the incremental costs associated with the use of spacecraft power and fuel, launch services, and other resources. This means that the main contributor to the government costs is the dedicated payload and, therefore, the total price is far below that of deploying an independent, government-owned satellite.

Countries Employing Hybrid Satellites or Hosted Payloads

Australia

Australia, being fairly remote from the rest of the military satcom innovators, has long had a history of being innovative with its use of satcom for the Australian Defence Forces (ADF) and has implemented a multiphase program to investigate, develop, and deploy a range of military satcom capabilities. This philosophy is reflected in its Defense White Paper which is updated regularly by the ADF.[19]

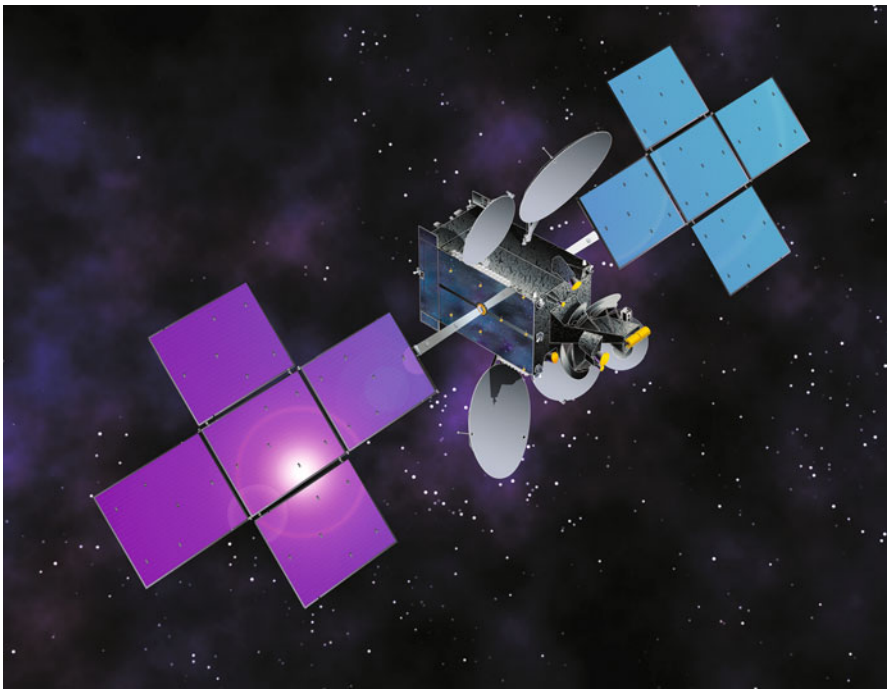


Fig. 9.2 Intelsat 14, host to the IP Router In Space (IRIS) Joint Capability Technology Demonstration

In 2003, Australia launched the Optus C1/D satellite[20] into a Pacific Ocean coverage area. The satellite was owned and operated by Optus, an Australian telecommunications company, and in addition to the primary commercial satellite communications payload contained a military payload funded by, and solely for the use of, the Australian Defence Force (ADF). This payload consisted of X-band, Ka-band broadcast, and UHF payloads and served to initially augment and eventually replace the heavy reliance that the ADF had up to that point on commercial leases of satcom capacity.

France

As stated previously, France's military satcom history is dominated by hybrid satellites, both Syracuse 1 and 2 employed defense payloads on board the Telecom series of satellites, which were owned by France Telecom. The Telecom 2 series of four satellites were launched between 1991 and 1996 and allowed the Direction Générale de l'Armement (DGA) access to military capability on board a national French satellite.

Japan

Japan has a long history in using satcom for its forces and has focused primarily on utilizing the frequency bands of the Superbird series of satellites. The Ku and

Ka-band payloads are used by commercial customers as well as by the Ministry of Defense and the Self Defense Forces but the X-band payload on each Superbird satellite is reserved for the exclusive use of the Ministry of Defense and provides military satcom for all three of the Self Defense Forces.[21] Currently Japan has been directed by the Japanese Diet to study the deployment of new satellite capabilities for surveillance and defense communications but no specific new programs have yet been launched.

Spain

The Spanish MOD[22] became a military satcom user with the launch of the Hispasat 1A satellite in September 1992. The satellite was placed at 30°W to provide transatlantic Ku-band commercial services between Europe, the United States, and South America. In addition to the commercial payload, Hispasat included an X-band payload for the sole use of the Spanish MOD. Hispasat 1A was followed in 1993 by Hispasat 1B, which had a similar payload configuration to 1A.

In 2001, the Spanish Ministry of Defense decided to move away from the policy of adding defense transponders to the commercial Hispasat satellites and explored the option of procuring their own military satellite. In July, 2001 Loral Space and Communications entered into a joint venture agreement with HISDESAT, a Spanish company owned by HISPASAT, INTA (a Spanish government organization) and a number of Spanish aerospace companies to found a Joint Venture company, XTAR LLC, to lease X-band communications satellite services to the US government and its allies (Fig. 9.3).

As a result of the Joint Venture, two satellites were launched. XTAR EUR[23] was launched in April 2005 into 29°EL and is wholly owned by the Joint Venture. SPAINSAT was launched in March 2006 into 30°WL and is wholly owned by the Spanish government with one of the on-board payloads serving as their dedicated military satcom resource and the other payload leased to the Joint Venture under the name of XTAR LANT.[24] XTAR's capacity on these two satellites is available to be leased to the US government and its allies.

UAE

In August 2007, Al Yah Satellite Communications Company (Yahsat) signed an agreement with a European consortium comprising of EADS Astrium/Thales Alenia Space to manufacture a state of the art dual satellite communications system[25] (Fig. 9.4).

The Yahsat program will result in two hybrid satellites, launched within months of one another in 2011, allowing Yahsat to provide civil and military customers with broadcast services, internet trunking via satellite and corporate data networks. The system is designed to accommodate the trends of emerging applications in the satellite industry like HDTV and other broadband satellite services using C-band, Ku-band, and Ka-band commercial frequencies as well as to provide a capability that provides for the move of military satellite communications into the military portion of the Ka-band.

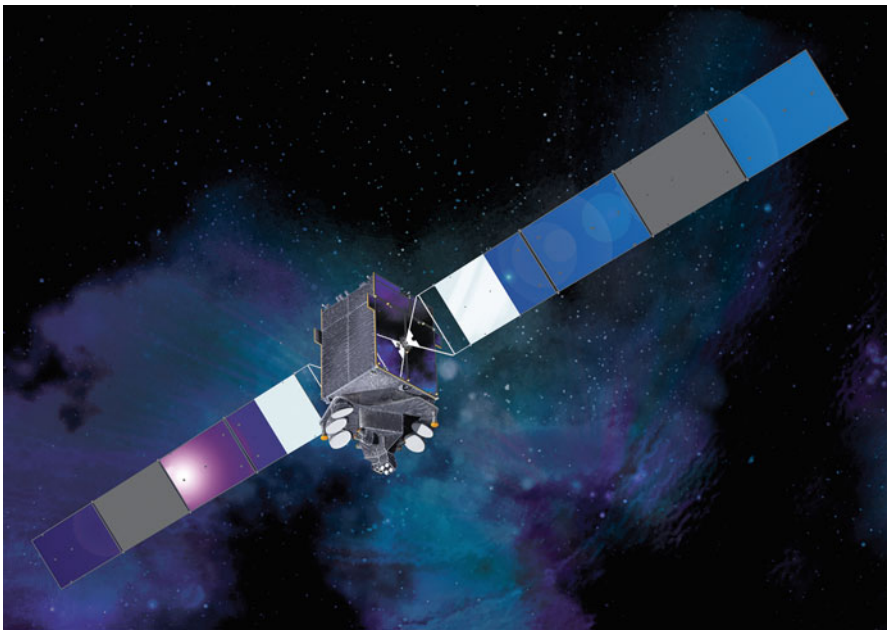


Fig. 9.3 XTAR-EUR satellite (Image provided by XTAR LLC)

Intergovernmental Agreements

Countries that do not have the resources for dedicated or hybrid systems often rely on intergovernmental agreements to obtain shared resources for defense-related communications.

European Nations

Most European nations, other than the ones mentioned above, have neither the budget nor the depth and breadth of requirements to justify investment in dedicated or hybrid satellite capability. These nations have typically used intergovernmental agreements with their allies to gain access to protected communications (Germany did this with France for many years prior to launching the SatcomBW program). When intergovernmental agreements are not possible, then long- or short-term lease contracts with commercial operators or service providers have often proved to be the vehicle of choice.

Nearly every nation has now leased one or more services from Inmarsat to include within its military portfolio for maritime or airborne communications and this has been augmented over the last 5–10 years with leases of Intelsat, SES, or Eutelsat capacity and more recently with commercial X-band communications leased from either Paradigm or XTAR. These nations include Belgium, Czech Republic, Denmark, The Netherlands, Poland, Portugal, and Slovenia.

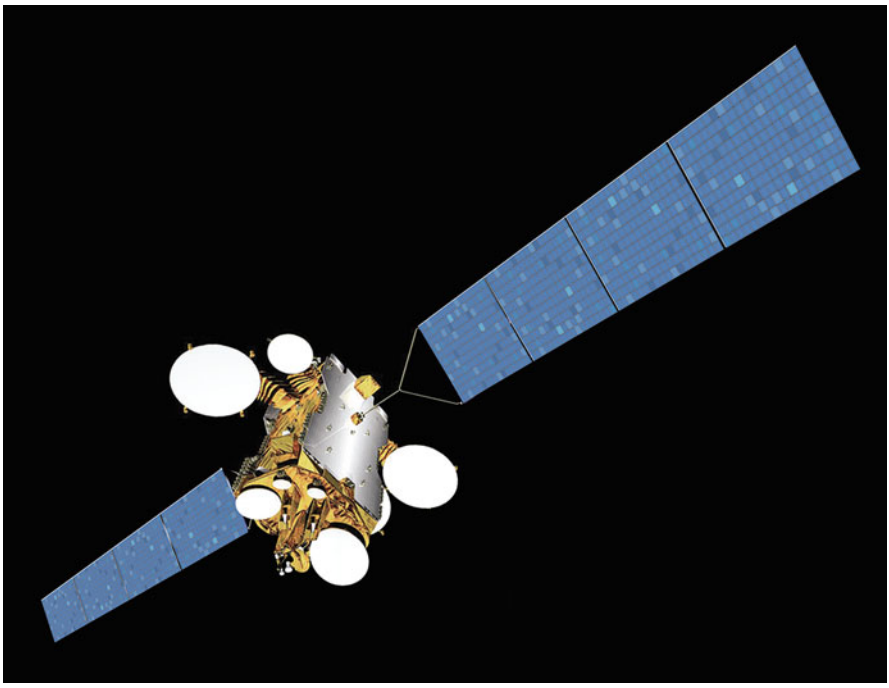


Fig. 9.4 Yahsat hybrid satellite for the United Arab Emirates (Image provided by Astrium Ltd)

NATO/France/Italy/United Kingdom

In May 2004, the NATO Consultation, Command and Control Agency (NC3A) decided to move away from owning and operating its own fleet of satellites and selected a multinational proposal to provide SHF and UHF communications through to 2020. This program, entitled the NATO Satcom Post-2000 (NSP2K) program[26,27], requires the French, Italian, and British governments to provide NATO with access to the military segment of their national satellite communications systems – Syracuse, SICRAL, and Skynet 4/5, respectively, – under a Memorandum of Understanding. This lease contract replaced the previous constellation of NATO IV satellites (discussed earlier) which were owned by NATO and operated by the UK MOD under an MOU with NATO.

NATO member nations are able to use the NSP2K capacity for their forces' communications needs whenever they are on a NATO exercise or operational deployment. The use of these satellites for national requirements, albeit on an ad hoc basis, has contributed in no small part to the perceived reticence of the nations with smaller requirements to procure long-term commercial satcom solutions for their national satcom needs. As NATO capacity requirements are increasing, spare capacity within NATO's allocation across the three fleets is often not available for individual member nations to use to satisfy their national requirements and

therefore it seems likely that NATO nations will increasingly seek alternative sources of military satcom capacity.

Australia/United States/United Kingdom

The Australian Defence Force (ADF) followed up the launch of Optus C1 with studies into the potential to enhance its capability by procuring capability in the Indian Ocean Region (IOR) and eventually procuring a more capable system for the Pacific Ocean Region (POR).

For the Indian Ocean Region, the ADF signed an MOU with the United Kingdom to allow access to the Skynet 4 and 5 capabilities and then followed this with an exclusive contract with Paradigm to secure UHF capacity on the Skynet 5B satellite.

For the POR and worldwide coverage, the ADF opted for access to the WGS constellation by signing an MOU with the US government in 2007.[28] This allows Australian forces access to the full constellation of five satellites and permitted the United States to expand the WGS constellation to six on-orbit spacecraft as, under the terms of the aforementioned MOU with the United States, the ADF agreed to provide sufficient funds to procure the sixth satellite. WGS1 was launched into the POR in late 2007 and became the ADF's primary satellite.

Italy/France

Athena-Fidus is a French-Italian geosynchronous military and governmental EHF/Ka-band wideband communications satellite capable of data transfer rates of up to 3 Gbits per second. Jointly procured by the French and Italian space agencies and defense procurement agencies, the system is intended to be used by the French, Belgian, and Italian armed forces as well as the civil protection services of France and Italy. Athena-Fidus will be launched in 2013 or early 2014.

Sicral 2, as described previously, is a joint Italian-French military satcom program, scheduled for launch in 2013, operating in the UHF and SHF bands. It will augment both the Sicral and Syracuse systems. The Sicral 2 Program will primarily support satellite communications for the two countries' armed forces, and is designed to meet the needs expected to develop in the near future. Like its predecessors, the new satellite and ground segment will provide strategic and tactical communications links for both domestic operations and foreign deployments. It supports all military, terrestrial, naval, and aerial platforms, operating in a single integrated network.

Guaranteed, Long-Term Leases and Ad Hoc Leases of Capacity

A lease contract is normally designed to deliver the twin objectives of managing both a nation's operational effectiveness and its defense budgets by ensuring that the nation has access to the required amount of guaranteed communications support without jeopardizing the ability to execute national deployments. Nations with only limited budgets or very small commercial augmentation requirements tend not to rely on a specific solution for their commercial satcom and simply procure what

they need, when they need it on an as available basis. Sometimes this is because their national defense system satisfies everything they might need and sometimes it is because it is simply not cost effective to preorder commercial capacity.

Obviously, in an environment where the supply of commercial capacity is limited, guaranteed access is not possible without placing pre-commitment contracts or reservations with the operator, so a necessary prerequisite of an ad hoc approach is that assured access is not a mandatory requirement for the users. This approach is often taken by nations first testing whether they need satellite communications before assigning specific budget lines for it. Any leases for commercial capacity that are subsequently entered into tend to be short-term because these governments normally cannot predict their future demand, and this prevents them from committing to lower-cost long-term contracts.

Several NATO nations with smaller military satcom requirements have opted to augment capacity provided under MOU with commercial leases to support their national operations. Often this approach splits out the procurement of military X-band capacity from commercial capacity but equally it can group all satcom requirements under a single commercial contract. Although Canada and Norway have entered into X-band lease agreements with Paradigm or XTAR as a precursor to a dedicated national solution in the future, more often a commercial satcom lease agreement is intended to be the long-term solution for a nation's military satcom needs.

Nations who can predict their future requirements with a little more certainty or who have more flexibility enter into long-term leases of capacity based on their best analysis of long-term requirements. Belgium, Denmark, The Netherlands, and Portugal procure their X-band military communications from either Paradigm or XTAR under fixed-term contracts with either a fixed amount of capacity assigned to them for their dedicated use or with the ability to call up capacity as they need it. The Netherlands have expanded on this approach for their C-band and Ku-band having entered into a multiyear, contract in 2005 with New Skies (now part of SES World Skies).

These types of contract result in assurances that the amount of contracted capacity will be available for use by the military *when it needs it* but does require some risk taking on the part of the customer that the capacity will be available *where it needs it* throughout the contract period as the requirements analysis is normally done before entering into the contractual arrangement.

Government-Industry Partnership

The sixth and newest option for governments, as highlighted above, has been implemented by the United Kingdom. The United Kingdom has had a military satcom program since Skynet 1 was launched in late 1969, making it the third country to launch its own national military satcom system. The United Kingdom has continued its involvement in military satcom up to the present day with the Skynet 4 and Skynet 5 series of satellites[29]; Skynet 4A being launched in January 1990 and followed by 5 further Skynet 4 satellites with three new Skynet 5 satellites being launched between March 2007 and June 2008. A fourth satellite, Skynet 5D,[30] is planned for launch in 2013. The United Kingdom uses the Skynet



Fig. 9.5 Skynet 5 series of satellites (Image supplied by Paradigm Secure Communications Ltd)

constellation for its protected communications and employs both UHF and X-band frequencies (Fig. 9.5).

Since 2003, the UK military satcom system has been owned and operated by a commercial company, Paradigm Secure Communications Ltd, which is solely responsible for providing the national critical communications to the UK MOD using the Skynet space and ground systems. This is done under a contract vehicle called the "Skynet 5 Private Finance Initiative (PFI) Program." Paradigm took ownership of the existing Skynet 4 military satcom system and then obtained loans from the worldwide finance community in order to design, build, and implement the next generation Skynet 5 system.

Not only does Paradigm provide the UK MOD with guaranteed access to the state of the art Skynet 5 satellite system for its highly protected, dedicated military satcom requirements, it also guarantees to supply all of the UK MOD beyond line of sight (BLOS) communications requirements. The UK MOD therefore specifies the service characteristics of any communications link that it needs and Paradigm defines the system solution that it is best able to supply from an operational and cost-effective perspective to meet that requirement. This may be using military or commercial satellite capacity or it may be ground based fiber or GSM technology – whichever is most appropriate and meets the UK MOD's requirements.

Paradigm has to maintain the extremely high availability of the military satellites that it owns and operates and ensure that the ground systems are fully operational at

all times. However, to execute its responsibility to supply all beyond line of sight communications, Paradigm also has to ensure that it has access to sufficient commercial satcom in a variety of frequency bands, teleport assets in remotely diverse locations and enough fiber leased lines to connect all locations and customer sites. This is done through a variety of long-term leases and Paradigm conducts frequent re-competitions to ensure cost-effectiveness. There is an agreement between Paradigm and the MOD on an incentive scheme to ensure that everything possible is done to provide the capacity needed.

This solution is only successful because the traditional roles of supplier customer are deliberately blurred in the PFI approach. While it is true that there is a comprehensive and detailed contract in place between the UK MOD and Paradigm, a large part of the relationship has to be based on trust. The Skynet 5 contract duration is for a minimum duration of 19 years and the MOD requirements must be met in 2022 just as they were in 2003. Because the very nature of communications requirements is that they are constantly and rapidly evolving and expanding, it is not enough to simply say "the capacity will be there when you want it." Paradigm and MOD therefore work very closely at both a working level and a management level to ensure that new developments in requirements are shared as soon as possible. In this way, Paradigm can make sensible investment decisions because it understands that the users will be there once the capacity is available and MOD can rely on the capacity being available for future platforms and applications because it worked closely with Paradigm to ensure those requirements have been taken into account within the joint planning process that they share.

To date, this symbiotic relationship between the UK MOD and its industry partner is unique but is being closely monitored and reviewed by many other nations, as can be observed by France's investigation into the outsourcing of Syracuse during 2009 and 2010 (as discussed later in this chapter).

Commercial Satellite Communications Augmentation

As stated above, in 2011, Europe's military and defense forces now procure an increasing percentage of their satellite communications capabilities from commercial sources, with some nations approaching 40% through commercial leasing. In contrast, the US DoD procures as much as 80% of its total satcom capability commercially through long- and short-term leases and has at times even exceeded that level. While originally this capacity was procured as ad hoc leasing of commercial capacity for urgent requirements and to cover shortfalls and "gaps," there is a growing tendency among all nations to look toward a more centralized procurement model. An overarching contract vehicle goes some way to alleviating some of the problems associated with an ad hoc commercial satcom requirement. Terms and conditions are pre-agreed with one or more suppliers, ensuring that if capacity is actually available when needed, there is no delay in activating the capacity because of protracted contract negotiations. There is also more likelihood that the contract will be flexible enough to grow and change with the customer's requirements.

In 2001, to enable capacity from the commercial satellite operators to be procured to augment the increasing military requirements, the US DoD's Defense Information and Systems Agency (DISA), issued the Defense Information Systems Network Satellite Transmission Services-Global (DSTS-G) contract. It was an Indefinite Delivery Indefinite Quantity (IDIQ) contract, allowing the military communications users to procure as much or as little commercial services and capacity for as long or as short a time as it wants under an overarching set of contractual terms.

Although some commercial capacity and services are still leased via other means by some DoD elements, the majority from 2001 to 2011 were procured through this contract vehicle. The DSTS-G contract has been replaced with a new program jointly administered by the Government Services Administration (GSA) and DISA. This program, entitled the Future COMSATCOM Services Acquisition Program (FCSA) commenced in early 2011 and is discussed later in this chapter.

The US model of using one overarching contract vehicle and then procuring each element of commercial capacity underneath this "umbrella" has proved to be extremely cost-effective, and while not necessarily being focused on delivering value for money or operational effectiveness, is becoming more popular with allies. In Europe, the procurement of both military and commercial satellite communications is characterized by smaller procurement budgets and, historically, a mistrust of national consolidation. Therefore any method of reducing procurement costs is welcomed and embraced. European nations are increasingly looking at more innovative ways of satisfying their commercial satcom needs.

The French ASTEL-S contract (awarded in 2005) aims to go one step further than DSTS-G by providing fixed tariff sheets for commercial capacity over a wide coverage area and for a fixed period of time. Capacity is still provided on an "as available" basis by the contractor (Astrium Services Ltd) but the French Navy can plan its budgets in advance due to the surety of the pricing and the contract terms. Astrium Services takes the risk of providing the capacity for the price specified in its fixed-term contract.

The European Defense Agency (EDA) is currently setting up the EDA Satellite Communications Procurement Cell (ESCPC) to fulfill a similar function to DSTS-G for European nations. However, the ESCPC is designed to not only provide a contract vehicle for nations to buy commercial capacity under but also to pool the demand for all European nations through a central procurement body. This allows a lower-cost per Megahertz to be negotiated by the procurement cell and for those savings to be passed on to the member nations. Current estimates put European governments' total expenditure on commercial satellite capacity leases in the region of 50 million euros (which is about US$72 million) per year. This program is expected to save participating governments as much as 30–50% on spot market spending.

Finally, there is the concept of an end-to-end service contract whereby the military procurement agency estimates its long-term needs and then contracts with an industry partner to guarantee this capability throughout the contract lifetime without the customer needing to specify the technical solution to be used. This is

precisely the situation with the UK MOD and its contract with Paradigm, discussed earlier in this chapter.

A fundamental remaining question is at what point does "augmentation" capacity become "core" capacity, critical to the warfighter's capability? In the case of the US DoD, as previously stated, over 80% of the required military satcom capacity is now procured commercially rather than using dedicated US satellites. It is therefore difficult not to believe that the commercial capacity is as much "core" capacity as the dedicated capacity – a situation that would have been impossible to imagine even 10 years ago. This condition is strongly shaping the future of military satellite communications procurement and policy.

The Future

There are historically two major components to military communications traffic: strategic and tactical. Both have an impact in shaping the way the future looks for commercial and military satellite communications:

- *Strategic* traffic tends to be high data rate, fixed location to fixed location and relatively easy to predict for a significant period of time. This enables solutions to be deployed using fixed coverage beams and for capacity to be committed over a longer period of time.
- *Tactical* traffic is characterized by the use of smaller ground terminals in dispersed locations. While data rates for tactical traffic can still be high (and are growing all the time), mobility and flexibility are of paramount importance. The solution calls for rapid redeployment and reconfiguration of assets both in space and on the ground. It is often very difficult to make long range predictions about the precise location of the deployments and the total capacity needed.

In the future, military planning units will continue to see an increase in theaters of conflict being engaged on multiple fronts in disparate locations. This will lead to a shift toward an increase in tactical traffic and, potentially, a decrease in strategic traffic. It is also foreseen as more likely that strategic communications will switch more to other forms of communications and be less dependent on satellites. Since tactical traffic is by its nature harder to predict, this will put greater emphasis on more flexible and capable communications solutions able to respond to an ever changing military environment.

The military satellites currently in production for launch within the 2012–2015 time frame are already starting to incorporate more and more transponder power to support the increased throughput requirements and more flexibility in the shaping of the spot beams in order to satisfy these more intensive "tactical communications" needs. However, it is apparent that some of these needs are overstretching the industry with the quantum leaps in capability and the pressure being put on design and implementation schedules. The US DoD has decided to split the WGS program into two phases to allow Phase 2 to be modified for

requirements which were not apparent when Phase 1 was completed. In Europe, Syracuse 3C was canceled in favor of investigating an outsourcing approach coupled with the joint approach with Italy on Sicral 2. NATO chose to procure its satcom through MOU rather than to replace the NATO IV satellites with a more capable NATO V series.

Therefore, one might assume that with a number of dedicated military satcom programs being merged, changed, or canceled, there is an opportunity for commercial satcom to become an integral part of the military warfighter's arsenal instead of always being referred to as an add-on, augmenting the critical national infrastructure. However, in conflict with the need to replace or augment military capacity is the US DoD's increasing need for flexibility in support of its current theaters of operations. The existing commercial satellites can only partially satisfy these types of requirements and this has been at least partly responsible for the world shortage in commercial satcom capacity, especially within the Middle East and Asia.

Commercial satellite operators have been unable to procure additional satellites with more flexibility and capacity optimized to defense-related needs in order to meet growing military requirements. This is because an operator has to present a viable business case to its shareholders showing that revenue will be recovered over the lifetime of the satellite to offset its investments. The US DoD (and other MoDs around the world) often have difficulty defining a *core* or *fixed* requirement in terms that will allow an operator to take a risk on the revenues that it will receive. Dialogue between US DoD and industry on this topic has been steadily increasing over the last few years, mirroring that which has been taking place in Europe over the last decade.

What therefore are the defense procurement agencies and satcom industry focusing their efforts on and what trends will be increasingly apparent over the next 3–5 years?

1. More outsourcing of critical and noncritical communications from the military operators into industry
2. New and improved contract vehicles will be introduced designed specifically to improve flexibility
3. Increased investment in Hosted Payloads by governments around the world
4. International partnerships between Allied Nations

Each of these four trends is described below using a specific example to illustrate the trend:

1. *Syracuse outsourcing*
   France has recently issued requests for proposals to outsource its Syracuse system to an industry partner and lease back communications services for the lifetime of the satellites. This contract will be similar to the German and UK programs, and will leverage the lessons learned by these nations, while retaining a French national independence. Interestingly, the plans look set to include the future Sicral 2 satellite, which means that the French government is planning from the outset to have an element of not only protection from the future growth in capacity requirements but also international collaboration to maintain value for money.

2. *Future COMSATCOM Services Acquisition (FCSA) Program*
   The DSTS-G contract for the procurement of commercial satellite communications services by DISA was due to expire in February 2011 but was extended until its replacement contract, FCSA, is fully in place. Although there remains a range of different contract vehicles for procuring capability in place across DoD, FCSA will be the main vehicle for DISA and DISA customers for the foreseeable future.

   The FCSA program consists of a set of acquisition parts that will replace three expiring DISA and GSA contracts, including DSTS-G, Inmarsat, and SATCOM II. Previously under the GSA schedule 70 and the DSTS-G contract there were a limited number of firms leasing capacity directly, and only three firms; ARTEL, CapRock, and DRS were permitted to sell satcom services directly to the US Government as of 2011. The implementation of the FCSA program will bring a major change to how satellite capacity and services are procured.

   The FCSA program has two new General Service Administration (GSA) Schedule Item Numbers (SINs) under the GSA IT Schedule 70. These new SINs, 132–54 *Transponded Capacity* and 132–55 *Subscription Services* are open to bids on a continual basis. They will provide specific satellite services requiring no development or systems integration activities. The FCSA program will also have two IDIQ contracts for providing end-to-end communications satellite solutions.

   In addition to allowing several more organizations to lease transponded capacity directly to the government, there will be several new service providers entering the market in addition to the three original DSTS-G providers. This will include the commercial satellite operators as well as those service providers without their own satellites who will be able to purchase satellite capacity from the satellite operators and resell it to the US Government.

   From the government's perspective, this maintains the current situation of allowing the warfighter to obtain the required capacity and services at lower costs through ongoing competition while enhancing the scope and nature of the marketplace. The government will be able to select services from a much wider range of competitors and technologies on an ongoing basis as the requirements evolve, while still ensuring that government assurance and protection requirements can be met. It is intended that this contract vehicle will be so all-encompassing that a communications procurer will be able to procure services from a few kilobits all the way up to a full payload capability for multiple users. It will take some time before this can become reality but it is destined to change both the way in which the procurement authority thinks about its requirements as well as the way in which industry sets itself up to address the evolving and ever more flexible requirements.
3. *Hosted Payloads*
   The Hosted Payloads concept has gained significant popularity within both government and industry. Satellite companies, recognizing the opportunity to further monetize their capital investments, have created new divisions focused specifically on Hosted Payloads. Government agencies, facing new budgetary

realities, have issued solicitations and held special invited "industry days" to investigate the cost and feasibility of various commercial solutions, including Hosted Payloads, as a way of fulfilling their mission requirements.

For government agencies, a key challenge to developing and launching a Hosted Payload is the ability to meet the rapid pace of commercial satellite development. Satellite operators have hard, fixed deadlines for launching their spacecraft in order to meet the huge commercial demand for communications. In many cases, the satellites being launched are replacing older ones that have degraded performance or are reaching the end of their useful life. Communications satellite companies cannot afford to delay replenishment satellites to accommodate developmental problems that can often occur with government payloads.

In August 2009, the Office of Space Commercialization, FAA's Office of Commercial Space Transportation, and Futron organized the first government-industry workshop on Hosted Payloads to share lessons learned and develop a common approach to facilitate governmental use of Hosted Payloads. Futron organized follow-on workshops in April and July of 2010 to develop approaches, recommendations and options for moving forward. It will be interesting to follow how the Hosted Payload concept evolves.

While nations like Australia have agreed to add operational payloads to commercial satellites as they are doing with Intelsat 22, to date the US Government has not followed suit and has only placed research and development payloads on commercial satellites. Interestingly, the US National Space Policy published in June 2010 contains specific language encouraging the US military to obtain space capabilities using more innovative approaches including Hosted Payloads. As stated the US government should "work jointly to acquire space launch services and Hosted Payload arrangements that are reliable, responsive to United States Government needs, and cost-effective."

In Canada, Telesat took the approach in 2010 of installing a three channel X-band payload, exclusively for government use, on board its new Anik G1 commercial C-band and Ku-band satellite.[31,32] This satellite, which will be located at 107.3°W when it goes into service in the second half of 2012 is a multi-mission spacecraft predominantly for direct-to-home (DTH) television broadcasting in Canada and broadband, voice, data and video services in South America. However, in a move viewed as daring by industry experts at the time, Telesat decided that there was sufficient latent need for government users in the Continental United States and Pacific regions that it would initiate a Hosted Payload program at its own risk. Within only a few months after Telesat's announcement, the full portion of the X-band capacity was purchased by Paradigm to augment the coverage provided by its Skynet 4 and Skynet 5 fleet and satisfy the needs of Paradigm's existing customers which are not served by Skynet today.

Intelsat is following on from the success of its involvement in the Internet Router in Space (IRIS) Hosted Payload and its Australian Defence Force UHF payload on board IS22 with its Intelsat 27 satellite.[33] This satellite will carry

a UHF payload identical to that on board Intelsat 22, but, similarly to Telesat when it announced the Anik G1 contract without a customer for its X-band payload, does not yet have a committed customer to take the UHF capacity. With the delays to the MUOS program and the lack of additional UHF satcom in orbit, Intelsat is confident that there will be customers for the capacity and that some of them may even sign up prior to the launch of the IS27 satellite.

Iridium has also announced plans for a Hosted Payload initiative.[34] The Iridium Next constellation of 66 satellites, for launch in 2015 to replace the original Iridium constellation has Hosted Payloads at the heart of its vision. Each of the satellites will have space for a 50 kg scientific or research payload for anybody to take advantage of.

4. *More national alliances*

As well as the alliances referenced above for the use of WGS by Australia and the partnership within AEHF between the United States, Canada, Netherlands and the United Kingdom, there is a growing interest in national alliances in Europe. This is led by France and Italy who are already collaborating on the Sicral 2 satellite and the Athena program as referenced above but there are also senior level discussions between the United Kingdom and France on a whole range of defense topics.

The Norwegian government has recently decided to launch a partnership with Spain for the purchase of a military communications satellite to contribute to the stability and effective monitoring of Norwegian interests, and to support the increasing Armed Forces participation in operations abroad where there is no necessary communications infrastructure. The project, which is called HisNorSat, is designed in cooperation with the company HISDESAT and will become operational around the 2014 time frame. The satellite will be partly owned by the Spanish MOD. The partnership will give Norway ownership of a defined part of a joint communications satellite with full control of the Norwegian-owned portion of the satellite which will operate in both X-band and Ka-band. The partnership will give the Norwegian MOD access to a capability far more sophisticated than if it were to procure a stand-alone satellite, or even a Hosted Payload on board a commercial satellite. It seems logical that more nations will opt for this approach in the future to exploit synergies in military communications requirements and allied operations in the same geographical regions while increasing cost-effectiveness.

Conclusions

The development and use of communications satellites followed shortly after the launch of Sputnik. Initially in the United States and the United Kingdom commercial and military satellites were on different paths with specific satellites used by each for their own missions with very little use by the military of commercial assets. In other countries, with fewer requirements and smaller budgets, different commercial and military payloads were placed on the same satellite usually with a small military payload on a commercial satellite. In the modern day, development of commercial and military satellite communications programs is not only

converging but these programs and satcom assets are increasingly critical as a joint solution to satisfy a nation's communications needs.

As has been repeatedly noted, some 80% of the US core military satellite communications requirements objectives are provided by commercial satellites and they can no longer be considered a supplement to dedicated military systems. This has, of course, raised a number of issues concerning the suitability of commercial satellites to carry sometimes quite sensitive traffic. The current pressing operational needs, and the cancellation of future military satcom programs, have forced US military organizations to utilize commercial satellites with no increased enhancements except satellite command encryption and encryption of the traffic being transmitted. Conversely, in Europe the majority of requirements for those nations with access to national infrastructure are still supported on the national military systems with commercial satcom playing a growing part in national infrastructure but still fulfilling an augmentation role. With defense budgets being increasingly constrained, European innovation has focused on increasing value for money without sacrificing operational effectiveness of the warfighter.

From all public predictions, the future over the next 5 years implies a growth in military satcom requirements. This implies the introduction of new platforms requiring more and more data transfer capabilities. This trend suggests that the number of dedicated satellites in orbit will only grow steadily. With the growth in number and sophistication of the UAVs being used worldwide, the need for large amounts of bandwidth to transmit the UAV data to processing facilities appears inevitable. However, there has been no indication on the part of the US or UK governments or others of a need for military Ka-band from commercial satellite firms. Consequently, the satellite companies are unlikely to plan for or launch military Ka-band capability without some indication of probable use. In addition, there is a dearth of military Ka-band terminals to support the reception of the data. While initial Ka-band users will be forced to use dedicated military satellites for their capacity, there will therefore be a continued reliance by the military on other commercial satellites to satisfy a significant portion of their extant and future communications requirements. With an ever increasing number of commercial companies entering the business of providing communications services to governments it is clear the landscape will be more uncertain and competitive.

The continuing question for the military and industry alike is how to best provide for the capacity and guarantee value for money without sacrificing military effectiveness.

Governments have already developed a variety of financing techniques to access the required capacity. These range from overarching lease contracts that encourage innovation and competition within industry through to solutions such as the UK's Private Finance Initiative (PFI) where the UK MOD sold its military satellite assets to a private company that operates the system for them and provides both dedicated and growth capacity.

The focus on hosted payloads provided by, among others, Intelsat, Telesat, and the Australian Defence Force provides a great opportunity for industry and military to work together to provide not only adjunct capacity but core capacity

in a timely fashion. Developing a business case for a hosted payload that not only meets the military's needs but also the schedules of the commercial operators is a fundamental challenge that has only just started to be investigated.

The business of providing commercial capacity to governments for military or other government uses has truly become an international business with many different players on both the provider and the customer side. National boundaries are becoming increasingly blurred as coalition forces are increasingly being deployed across the world and interoperability between those forces becomes a given rather than an option. Ensuring that national security requirements can continue to be satisfied in an ever increasing international environment will continue to be a challenge. France/Italy and Spain/Norway are leading the international collaboration developments on their future military communications satellites and will be working hard over the next decade to ensure that national requirements can continue to be met while sharing physical assets.

This chapter has allowed us to present what has happened in the past with the use military satellite communications and how this history has shaped the present day environment and the increasing usage of commercial communications by the military. The military satellite communications world has always been dynamic and innovative. The next 10 years will see great changes in the area of defense and strategic satellite communication systems. These changes will come not only in specific military and dual-use commercial technologies but also in the creation of yet more innovative business models that have never been seen in the industry before. The objective, however, will remain the same. This is to provide the military users with the communications they need, when and where they need them.

Acknowledgments The authors would like to thank the following for their contributions to this chapter: Lt Gen (Ret) William Donahue, Col. Patrick Rayermann, Philip Harlow, Robert Twining, Britt Lewis, Diana Goody, Dylan Browne, and Edward Beck.

Cross-References

▶ Fixed Satellite Communications: Market Dynamics and Trends
▶ Mobile Satellite Communications Markets: Dynamics and Trends
▶ Satellite Communications Video Markets: Dynamics and Trends

Notes

1. Defense systems article: commercial satellites plug bandwidth gap for military satcom, http://defensesystems.com/Articles/2011/02/28/Cover-Story-Commercial-Satellites-Evolve.aspx?Page=1

2. US Navy homepage for UFO system, http://www.public.navy.mil/spawar/PEOSpaceSystems/ProductsServices/Pages/UHFGraphics.aspx
3. Description of MUOS system, http://www.globalsecurity.org/space/systems/muos.htm
4. US Air Force fact sheet for DSCS system, http://www.losangeles.af.mil/library/factsheets/factsheet_print.asp?fsID=5322&page=1
5. US Air Force fact sheet for WGS system, http://www.afspc.af.mil/library/factsheets/factsheet_print.asp?fsID=5582&page=1
6. LA AFB fact sheet for the GBS system, http://www.losangeles.af.mil/library/factsheets/factsheet.asp?id=7853
7. Intelsat general UAV services, http://www.intelsatgeneral.com/services/applications/uav.aspx
8. Satellite markets and research: government/military demand for commercial satcom remains steady, http://www.satellitemarkets.com/node/769
9. U.S. Air Force fact sheet MILSTAR satellite communications system, http://www.af.mil/information/factsheets/factsheet_print.asp?fsID=118&page=1
10. U.S. Air Force fact sheet advanced EHF system, http://www.afspc.af.mil/library/factsheets/factsheet_print.asp?fsID=7758&page=1
11. Defense budget recommendation statement made by secretary of defense Robert M. Gates, 06 April 2009, http://www.globalsecurity.org/military/library/news/2009/04/dod-speech-090406.htm
12. Alcatel press announcement on Syracuse 3B, http://www.home.alcatel.com/vpr/vpr.nsf/DateKey/16012004uk
13. Description of Syracuse 3 system, http://www.deagel.com/C3ISTAR-Satellites/Syracuse-III_a000283001.aspx
14. Satcom BW overview, http://www.astrium.eads.net/en/programme/satcombw-comsatbw2.html
15. Sicral program overview, http://www.telespazio.it/pdf/Tes53_impg_3_4_09_ing_lowre-solution.pdf
16. Sicral 2 press release, http://www.thales-transportservices.com/Press_Releases/Markets/Space/2010/Thales_Alenia_Space_and_Telespazio_sign_contract_for_Sicral_2/
17. Overview of Russian space activities, www.russianspaceweb.com/spacecraft_military.html
18. Intelsat announcement of Intelsat 22 satellite procurement and procurement of UHF payload by Australian Defence Force, http://www.intelsat.com/press/news-releases/2009/20090427-2.asp
19. Australian white paper on defence, http://www.defence.gov.au/whitepaper/
20. Press announcement of Australian involvement in US WGS program, http://www.australiandefence.com.au/F4F2FBC0-F806-11DD-8DFE0050568C22C9
21. Basic guidelines for space development and use of space, www.mod.go.jp
22. Hispasat satellite fleet information, http://www.hispasat.com/Detail.aspx?SectionsId=67&lang=en
23. XTAR EUR satellite information, http://www.xtarllc.com/xtar-eur.html
24. XTAR LANT satellite information, http://www.xtarllc.com/xtar-lant.html
25. Yahsat program information, http://www.yahsat.ae/yahsecure.htm
26. DISA overview of NSP2K program, csse.usc.edu/gsaw/gsaw2005/s9f/stoops.pdf
27. NATO overview of NSP2K program, http://www.nato.int/issues/satcom/index.html
28. Space daily report on ADF entering the WGS program, http://www.spacedaily.com/reports/Australia_To_Join_With_United_States_In_Defence_Global_Satellite_Communications_Capability_999.html
29. Overview of Skynet 5 program, http://www.army-technology.com/projects/skynet/, http://www.astrium.eads.net/en/programme/skynet-5-.html
30. Announcement of Paradigm's fourth Skynet 5 satellite, http://www.astrium.eads.net/en/press_centre/paradigm-agrees-deal-with-uk-ministry-of-defence-mod-for-fourth-skynet-5.html

31. Announcement of Anik G1 satellite by Telesat and Loral, http://www.spaceref.com/news/viewpr.html?pid=30941
32. Announcement of Paradigm's leasing of Anik G-1 X-band capacity, http://www.spacenews.com/satellite_telecom/101013-paradigm-xband-anik.html
33. Announcement by Intelsat of its intent to launch Intelsat 27 with a UHF hosted payload, http://satellite.tmcnet.com/topics/satellite/articles/95425-intelsat-is-27-satellite-launch-2012.htm
34. Iridium NEXT program will include opportunities for hosted payloads, http://www.iridium.com/about/IridiumNEXT/HostedPayloads.aspx

References

Annual report to congress: military and security developments involving the people's Republic of China 2010, http://www.defense.gov/pubs/pdfs/2010_CMPR_Final.pdf

DISA conference proceedings 2009 for the commercial satcom session, http://www.disa.mil/conferences/2009/briefings/satcom/Commercial_SATCOM_DISA_Conference_2009.ppt (slide 33)

Long March launch of Chinese Military Satellite: November 2010, http://www.space.com/9606-chinese-military-communications-satellite-reaches-orbit.html

National communications system fiscal year 2007 report, http://www.ncs.gov/library/reports/ncs_fy2007.pdf, p. 26

Satellite 2001 daily news: military bandwidth migration path leads to Ka-, X-band satellite offerings, http://www.satellitetoday.com/eletters/satellite2011_daily/2011-03-11/36343.html

Economics and Financing of Communications Satellites

10

Henry R. Hertzfeld

Contents

Introduction 222
Satellite Telecommunications Services 223
The Business of Satellite Communications 224
Trends in Access to Space and in Manufacturing Satellites 228
Access to Space 228
Comparisons of Productivity in Manufacturing Satellites 231
Cross-References 237
References 238
Further Reading 238

Abstract

The economics and financing of satellite communications is a very large and complex topic. It ranges from normal business planning, analysis, and investment financing, to issues of government policy, dual-use technologies, and national security and defense. Commercial satellite systems represent a special case of economic analysis since such systems are heavily dependent on a government market that is focused on political considerations of budgeting and regulation. Today, satellite telecommunications systems are critical to almost all nations of the world, and they are especially important in the approximately 60 nations that have domestic launch and/or satellite operations capabilities. This chapter will specifically focus on four topics: (1) a summary of the economic characteristics of the industry and a review of major trends in the industry, (2) a summary of the elements of a business plan for satellite telecommunications, (3) an analysis of issues in the manufacturing productivity

H.R. Hertzfeld
Space Policy Institute, Elliott School of International Affairs, The George Washington University, 1957 E Street, NW, Washington, 20052, DC, USA
e-mail: hrh@gwu.edu

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_9, © Springer Science+Business Media New York 2013

for satellites and an analysis of commercial satellite manufacturing compared to government satellites, and (4) a brief discussion of future cost considerations including the increasing risk of space sustainability, insurance, and rules concerning disposal of satellites after their useful lifetime.

Keywords

Auction of spectrum • Commercial satellite systems • "Dual use" of satellite networks • Economics • Insurance • Investment financing • Launch costs • Manufacturing • Market sectors • Operating and capital costs of satellite networks • Satellite services • Satellites • Size of markets • Telecommunications • Video services

Introduction

In 1876, Alexander Graham Bell invented the telephone. The invention spread rapidly and became the standard mode of remote voice communications during the first half of the twentieth century. Copper wires were strung and these became the major mode for the transmission of voice communications.

In the 1940s, Arthur C. Clarke suggested the possibility of using geostationary satellites to beam telecommunications signals from a point on Earth back to multiple points. However, during the mid-twentieth century, telephone signals were being relayed on land with copper cables and microwave towers and across the oceans with similar cables of limited capability compared to those that are in use today. During the 1960s, the first telecommunications satellites were launched successfully to low Earth orbit.

Although communications satellites had greater capacities than ground systems for overseas transmissions when first deployed in the 1960s and 1970s, this was no longer true when faster terrestrial communications using fiber-optic cables were developed in the 1980s. Cable transmissions of any type, it should be noted, are best designed for point-to-point communications while satellites are best for point-to-multipoint uses.

Local commercial television broadcasts over the air were inaugurated in the late 1940s, and grew rapidly. The larger selection of stations enabled by subscription cable delivery of television to households gradually became a standard form of TV delivery in most countries of the OECD. By the 1990s national and international cable TV distribution was widespread and direct broadcast satellite TV to consumers was also beginning to grow, enabled by smaller terrestrial receiving antennas and more powerful satellites (Satellite Industries Association, http://www.sia.org/satellites.html.) Satellite delivery is particularly advantageous in remote areas not served by cable but is also a strong competitor to cable in urban areas. As mentioned above, it is also the only system that can effectively deliver point-to-multipoint signals in nations that are not well wired, such as in developing countries.

There has been a very rapid and dramatic change in developed economies over the past two decades in terms of telecommunications delivery media. This has been a shift from wireless TV and wired telephone to wired (cable) TV and wireless cell

phone systems – although direct broadcast satellite services now represent a significant delivery mode in many economically advanced countries as well. This shift is also beginning to occur in developing countries as well, but some of these patterns are less pronounced and some satellite networks (such as O3b) are seeking to provide broadband Internet services and video services directly to consumers in Africa and other parts of the world and thus seeking to bypass terrestrial cable networks.

These shifts are particularly significant for the satellite and space industry. Space capabilities are actually central to all systems today by being a part of the overall telecommunications services delivery chain. Cable TV, although wire-based to connect to homes, is dependent on network uploads via satellites. Cell phones and systems are also linked and coordinated by backhaul precision timing through GPS satellites. And, of course, satellite TV and radio are now ordinary consumer services in a very competitive market, sold and distributed by many companies in many advanced economies such as the United States, Canada, Japan, Korea, Australia, and Europe to name only some of the countries where competitive satellite services are sold.

Satellite Telecommunications Services

Satellite communications networks are separated into several types as noted in earlier chapters and particularly ▶ Chap. 4, "Space Telecommunications Services and Applications." Fixed satellite services (FSS) typically transmit between a GEO satellite and one or more fixed locations terrestrially. The various types of telecommunications services that FSS networks provide include television distribution, broadband data, and voice communications. Some higher-powered FSS networks provide direct to the home video and audio services.

So-called broadcast satellite services (BSS), as defined by the International Telecommunication Union (ITU), provide direct transmission to the home via very small dishes. There are also what are called direct audio broadcast services (DABS) or satellite radio services. Although different frequencies are allocated by the ITU for FSS, BSS, and DABS services, the distinctions between these services are not always very clear to consumers. This is because higher-powered FSS satellites can provide services that look and act like direct broadcast satellites services for either television or radio services. Overall revenues from these various types of satellite networks are heavily weighted toward video services. Television distribution and direct television satellite services generate over 75% of the revenues, with data transmission representing about 20% and voice services representing only some 5% – at least for satellites serving the most economically developed countries.

Mobile satellite services (MSS) transmissions are conducted from satellites to receivers that are not fixed in any one point terrestrially. They include maritime, aeronautical, land mobile services, and other transportation-related uses, including emergency and police communications. Many of these networks are private.

In addition, there are many support services included in the satellite industry. As also described below, they include the manufacturing of satellites, launching services to get the satellites into proper orbits, ground receiving equipment, and many financial, insurance, and other business services.

All components of the industry have been growing, even during the recent economic recession beginning in 2008. Worldwide, the overall telecommunications satellite industry now generates the largest amount of revenues – by far – of any commercial space industry. The largest component of the communications satellite industry, in terms of revenues, is for provision of telecommunications services; this represented 58% of total industry revenues in 2010. This was followed by revenues from the sales of ground equipment (31%), manufacturing (8%), and launches (3%). (Satellite Industries Association, http://www.sia.org/satellites.html.)

There are many satellite applications that provide opportunities for very useful and profitable businesses. Figure 1.1 in ▶ Chap. 1, "Satellite Applications Handbook: The Complete Guide to Satellite Communications, Remote Sensing, Navigation, and Meteorology" of this handbook provides a good overview of the many types of satellite applications that affect our daily lives. Over time, it is clear that many of these applications have become a critical part of the economic infrastructure. Of the various market sectors of the commercial satellite world, communications satellites predominate in terms of total revenues, number of users around the world, and direct impact on people's lives. Despite this predominance of satellite communications, the other services, such as remote sensing and space navigation, are still greatly important. Satellite meteorology is typically not a commercial service, but it is nevertheless vital to public safety.

The Business of Satellite Communications

Satellite telecommunication, in light of its huge revenue stream and the billions of consumers that use this service, is clearly the most mature of space applications. In fact, some economists would say that the 50-year-old communications satellite industry virtually represents the only example of a mature commercial use of space.

Planning, financing, building, launching, and operating a communications satellite or satellite system has become routine business. It is a long-term investment, and fits a standard business model. Satellite systems require a high up-front capital expenditure, a reasonably long manufacturing and start-up period (over 2 years), and face a number of high investment risk factors. In spite of the obstacles, satellite telecommunications has proven to be a space application that can generate a long-term multibillion dollar (US) revenue stream and profitable returns.

These systems are in some ways very different from most industries and in other ways identical. The differences are centered on the large government presence in the technological developments as well as the role of governments as a purchaser of these services. The similarities are like those with any other regulated infrastructure or utility that requires complex and expensive investments in equipment and/or

distribution systems and that provide essential services to a large number of people. Because of these similarities to any private investment, this chapter will focus on specific topics unique to satellites and space businesses and will not attempt to describe normal business and economic issues that can easily be found in any basic management or economics textbook.

Early space telecommunication systems were not standard business ventures. They were built from a combination of public and private Research and Development investments and required access to space. This vital launch service, in the first decades of communications satellite service, could only be provided through a government launch vehicle. Until the late 1970s, the only vehicles capable of performing the launch services were either in the United States or the USSR. And, the USSR was not in the commercial launch business and did not launch private satellite payloads. The US government's involvement in technological development and regulation, government purchase and use of the services (both military and civilian), and government policy were integral to any corporate telecommunications business plan.

The US government's role has dramatically changed over time. But it is still very important today. Although the US Department of Defense (DoD) has its own satellite telecommunications system, it also purchases a large amount of commercial capacity to fulfill its total communications needs.

In addition, in recent years the DoD has dedicated transponders and instruments on commercial satellite platforms. The use of these "hosted payloads" (both in the case of the United States and Europe) is currently growing and is projected to grow even more. This combination, which creates a new and profitable business opportunity for private satellite operators, also potentially enables defense agencies to save money by requiring fewer dedicated expensive satellites within its own fleet. But it adds an interesting dimension to the relationship between government and industry both in the United States and Europe and raises numerous questions about the role of private business with security-related space assets. This subject of dual use of communications satellites was addressed in ▶ Chap. 9, "An Examination of the Governmental Use of Military and Commercial Satellite Communications".

The government is also a regulator. In the United States, the Federal Communications Commission allocates the available spectrum and is the US interface with the International Telecommunications Union. Most other countries have a governmental agency or ministry that oversees the use of radio frequencies including those for satellite communications. Often this is the entity that participates in the International Telecommunication Union (ITU) processes and international conferences. (As noted in ▶ Chap. 4, "Space Telecommunications Services and Applications" the ITU oversees international spectrum issues, defines different types of satellite services and the associated frequencies for that service. The ITU also oversees the assignment of valuable locations in the geostationary orbit, which is where the largest telecommunications satellites are placed. Constellations that operate in Low Earth Orbit and Medium Earth Orbit are also under the purview of the ITU in terms of international regulatory processes.)

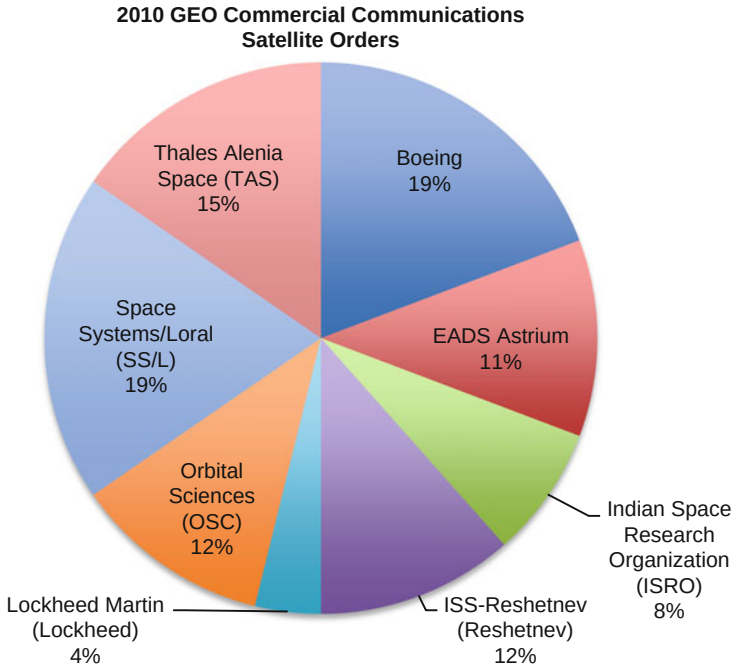


Fig. 10.1 Commercial GEO satellite orders in 2010 (Futron Corporation: 2010 Year-End Summary)

Governments fund and perform R&D that supports the technological development of the industry. It is also the province of governments to issue licenses for launching payloads into space. (In the United States, the Department of Transportation, Federal Aviation Administration, is responsible for licensing launches. In most other countries this is a ministry that addresses space, but in Europe, the European Aviation Safety Agency (EASA) is assuming authority for some suborbital flights for the emerging industry known as "space tourism" or "space adventures.") These licenses require companies to demonstrate a set level of financial responsibility for their space activities and mandate that they follow detailed safety procedures and take a number of steps to avoid the creation of space debris. (All FAA regulations can be accessed at: http://www.faa.gov/about/office_org/headquarters_offices/ast/regulations/.)

What has changed over time is that any company can now purchase a launch and obtain access to space for a legitimate business purpose. There is competition for these services and they are not limited to the United States. Over 10 nations now have launch capabilities and publically sell these launch services. Furthermore, many nations also now have the ability to manufacture satellites, and strong international competition now exists in the satellite manufacturing arena.

Figure 10.1 illustrates how widespread these satellite manufacturing capabilities are as demonstrated through orders for new commercial satellites. This data as compiled by

the Futron Corporation on behalf of the Satellite Industry association for 2010 shows for this year United States companies had 54% of the market share, European companies 27%, Russia 12%, and India 8%. Manufacturing capability also exists in China and Japan, and these could expand in international importance in future years.

Initially international telecommunications via satellite were the preserve of a consortium that was set up as an international organization with open-ended membership of the different governments involved. Financing for this entity called Intelsat was provided through the many governments that participated in the Consortium. It began with just a dozen members but expanded to over 100 countries. Beginning in the 1980s, this international monopoly approach began to be questioned and there were efforts to create competitive international systems. As the communications satellite sector developed to a more mature, reliable, and developed stage, private companies emerged and increasingly sought to offer GEO-based telecommunications services competitive with Intelsat. In the late 1990s, there were also plans for private LEO telecommunications satellites. (These included the proposed Teledesic system for FSS-type offerings and the Globalstar, and Iridium for mobile satellite services, for example.)

Although the satellite telecommunications industry has matured and produced many profitable and long-lived systems, it was not without risks and business failures. With the technology sector collapse in 2000, the Teledesic system was never completed, Globalstar had to file for bankruptcy protection; Iridium (owned by Motorola) also went into bankruptcy.

By the mid-1980s the privatization of the satellite communications industry had begun and the satellite telecommunications industry was also threatened by competition with fiber-optic cables. By 2001, not only had Intelsat been privatized, so had Imarsat (in 1999) as well as Eutelsat (also in 2001). This was a time of enormous change in the business and economic structure of the satellite communications industry. Along with the privatization was the influx of new companies, the creation of substantial debt equity, the purchase of Comsat by Lockheed, the end of the technology "bubble," and the "dot com" collapse (Mechanick 2011).

This radical and relatively fast change in the way the satellite communications industry was organized and structured created economic benefits as well as added business risks. Bankruptcies (e.g., Iridium, Globalstar, Teledesic, and ICO), new mergers, and international telecommunications conglomerates all became part of a constant flux in companies and market positions. These mergers and acquisitions and the transfer of ownership from governments to equity finance institutions has led to dominance in the international FSS business by Intelsat (headquartered in Bermuda), SES (headquartered in Luxembourg), and Eutelsat (headquartered in Paris). International mobile satellite service is dominated by Inmarsat (headquartered in London) but the reconstituted Iridium and Globalstar still offer key services around the world. The largest industry in terms of revenues is the direct broadcast satellite television services and these are spread around the world. Many smaller satellite companies also continue to operate. (See ▶ Chaps. 7, "Satellite Communications Video Markets: Dynamics and Trends," ▶ 8, "Mobile Satellite Communications Markets: Dynamics and Trends," and ▶ 9, "An Examination of the Governmental

Use of Military and Commercial Satellite Communications" to find more information about satellite market history and current structure. The appendices at the end of the handbook provides information with regard to the many different commercial satellite communications entities providing domestic, regional, or international services.)

Trends in Access to Space and in Manufacturing Satellites

Access to Space

There has been a long-standing expectation in the space community that the cost of access to space (i.e., launch vehicles) will drop exponentially. A major breakthrough in launch technologies has been a goal of numerous unsuccessful R&D programs such as the NASA/Lockheed-Martin X-33 effort that was canceled in 2001 as well as other similar efforts such as the X-34, the X-37, the X-38, and the X-43 (Pelton and Marshall 2006). The corollary of a dramatic drop in cost of access (and prices for launching) is that a floodgate will open and that markets will suddenly develop for new uses of the space environment. Profitable private ventures will flourish and government agencies will be able to purchase inexpensive launches for research satellites and payloads.

This call for inexpensive access illustrates the dramatic way in which economic factors *could* influence the demand for space. To date, their influence can only be found in the negative hypothesis: that expensive access to space has capped the demand for space activities and created a barrier to entry that is virtually insurmountable for most activities. This is an example of a "technology push" where the emphasis is on the supply side – providing access cheaply. There really is no *economic* reason to develop the technology unless there is a sufficient market demand to do something of value in space. There may be other reasons – social, political, or security – to go to space often and cheaply which could provide a public-goods stimulus for additional investments in cheap access technologies.

The assumption that cheaper access to space is the key to the future growth of space activities should be subjected to a closer analysis. The results may be very mixed. That is, cheaper access to space will clearly benefit both suppliers of space products and services as well as consumers of those products. But there are current and future space activities that will exist whether or not the cost of access is significantly decreased. Telecommunications is one of these.

First, consider the types of space applications that are not very price sensitive to launch costs. Essentially they require very large up-front investments that are recovered and exceeded relatively quickly over time from project revenues. Typically the services are sold to end users and therefore have a large mass market where the stream of revenues is relatively easily foreseen. In telecommunications, the existing large demand for voice and other transmissions existed before communications satellites were developed. Once a space satellite or facility is launched and placed into the desired orbit, it can have an expected lifetime of 15 years. This is a life expectancy

double that of a telecommunications satellite built just a couple of decades ago, which will result in a noticeable decrease in the demand for future satellite manufacturing and launches. Of course other unpredictable factors such as the future demand for telecommunication services and the crowding of the most profitable geostationary orbits and spectrum bands will also affect future launch demand.

Numerous business cases have demonstrated that even the high cost of building the satellite system and the high cost of launching it are relatively small percentages of the total revenue over its operational lifetime. Cheaper access to space might mean higher profits for the owner or operator of the system, but today's profits are sufficiently large that expensive up-front costs have not deterred companies from making these investments. These costs have made it more difficult for satellites to compete with high-capacity and high-cost-efficiency fiber-optic networks.

Second, consider the types of activities have the best opportunity to grow if there is cheap access. The largest opportunity in this respect might be activities that require multiple and regular trips to space and return to Earth. This implies one of three things:

1. That there is something to do in space itself (e.g., manufacturing or transporting people to space and providing for their return)
2. That point-to-point Earth transportation through space at high speeds could be proven to be technologically feasible and safe
3. That a true market for space adventurism or tourism exists and, as above, people will hopefully have something useful to do there

Third, consider the opportunities related to private Research and Development (R&D) involving space activities. Such activities are presently far too expensive for most companies or universities. (The availability of direct government subsidies and other incentives for space research has been the standard practice for many years. With the current budget deficit coupled with the increasing complexity and cost of research equipment, future government aid is not likely to match the demand for this type of research effort.) Private capital markets for high-risk R&D funds are often not large enough for a space project, since the cost of a launch is usually included in the cost of a corporate or university research program. In today's environment, an expensive launch can be the deterrent to proceed with the project.

Fourth, consider government programs or project activities that are subject to major budget pressures. There is, of course, a difference between an agency's budget and the project's budget. Many government project managers are advocating cheaper access in order to carry out their projects on a cost-effective basis since their individual funds are constrained. The agency-level huge capital requirements to fund a technology program that might lead to reducing launch costs are outside of the scope and capability of the project offices that are generally most concerned with current operating costs. Even though they are within the same government organizations, the role of a project manager is more similar to that of any final demand consumer.

Fifth, consider that there is a limit to how much launch costs can be lowered. Even if the cost of the launch vehicle is reduced dramatically, a number of other economic factors are not likely to change. Among them are:

- The high costs of launch facilities, payload integration, storage, testing, etc.).

- For the foreseeable future, only launches from a costal location or a very sparsely populated and remote area will be permitted because of safety considerations. This will make it necessary to transport, at considerable cost, payloads a significant distance from the point where the business is located or the product is manufactured to the launch site. The same delay will also exist at the delivery site. These costs will not be reflected in the launch price itself, but are real costs in time and transportation to the customer. (Sea launch operations or manufacture and launch from the state of California where many manufacturers are located, however, could possibly mitigate these considerations.)
- Delays in launches will frequently occur and add to launch costs.
  - Launch vehicles are complex machines and mechanical problems will occur with some frequency.
  - Weather will delay launches as it does today for both space launches and even normal airline traffic.
  - Security issues may cause delays.
  - Regulatory issues (safety, financial, environmental, etc.) will also likely continue to be complex and costly.

Payloads bound for space will need to carry very valuable commodities where the speed of delivery is of the highest priority. This means that a launch schedule has to have a high degree of reliability with little variance, otherwise alternatives will be financially more attractive. The time value of money, therefore, becomes a large expense, unrelated to the hardware costs of physically getting to space and returning. Export control issues will continue to dominate launches and will become particularly difficult if landings and relaunch occur in different nations. The demand for launches may never reach a level where economies of scale in manufacturing and launching will be realized. Insurance and liability issues will continue to be problems, particularly since the cost of insurance is related not only to the safety record of launch vehicles but also to the general level of claims payouts of all insurance and reinsurance policies. And, finally there is always the probability of an accident and the risks of suspended operations for a long period of time until the cause of the accident is determined and fixed.

Sixth, and last, one must consider the economics of the cost of developing a new and cheaper launch system. What will the government or private organization pay for the very expensive development of a new system? Who will bear the risks? Will the costs be amortized over the lifetime of the vehicles (and result in higher launch prices) or will a government underwrite the costs? Who benefits from such a system, and will taxpayers be willing to assume the burden of the cost?

The answers to these questions are not just an academic exercise. They go to the root of the linkages between economic and social motivations for future space activities, and how they are answered will shape much of future space development.

It is interesting to note that a 1975 study of the next 200 years in space made an assumption that access to space would be much cheaper by the year 2000 (Brown and Kahn 1977). The study analyzed many scenarios for the future using a variety of different assumptions. One of these assumptions stands out prominently. By extrapolating the rapid trend in technological improvements, most noticeably in

integrated chips and computers, and transferring that to launch vehicle improvements during the 1950s and 1960s, the report concluded that this trend of increased productivity and efficiency coupled with rapid decreases in prices would continue. Clearly, it has not. Space access is nearly as expensive today as it was in 1975.

A common thread of the literature on space commercialization is that cheap access is key to the future development of space. Given the above-mentioned parameters and the very difficult hurdles that will have to be overcome in many more areas than simply new launch technologies, this assumption comes into question and thus should be studied much more closely. It very well may be that some important activities will occur if launches are dramatically cheaper. But, history has already demonstrated that profitable space activities, particularly in telecommunications and related services that have large and mature terrestrial markets, are possible even with expensive launches. Likewise, it is possible that new launch systems using tethers, so-called space elevators, rail guns, or nuclear or electrical propulsion (as opposed to chemical propulsion) may be developed in future years, but such alternatives are not near-term prospects, and the implications of such alternative launch systems are not clear at this time.

Comparisons of Productivity in Manufacturing Satellites

Satellites have become more efficient for companies to manufacture by using state-of-the-art production techniques and by a steady demand enabling the realization of economies of scale. Satellites have also become larger, more powerful, and longer-lived. (An exception to this is the development of micro and nano-satellites for LEO applications. This discussion is primarily focused on the large GEO telecommunications satellites. In the future, it is possible that some telecommunications applications will be possible with very small satellites.)

As noted earlier, the expected lifetime of a new GEO satellite is now more than double what it was 20 years ago. Not all satellites are the same, and the following discussion documents the important differences between manufacturing a government satellite from those made for commercial purposes. Such satellites are different products: commercial satellites are produced relatively quickly and efficiently in response to for-profit pressures while the government satellites are often pushing the new technology edge and are also subject to government-mandated oversight and audits. Many military satellites have special requirements for radiation hardening, encryption capabilities, and redundancy or protective switches. Often the same companies produce both types of satellites. A comparison, discussed below, of manufacturing productivity and efficiency has documented that the commercial satellites are made faster and more efficiently (Coonce et al. 2010). But the study also highlights a number of important financial and economic characteristics of the manufacturing process and concludes that improvements in the efficiency of producing government satellites would also be possible without major systemic changes.

Three types of satellite systems are compared. First are commercial telecommunications satellites manufactured for private customers. Second are civilian

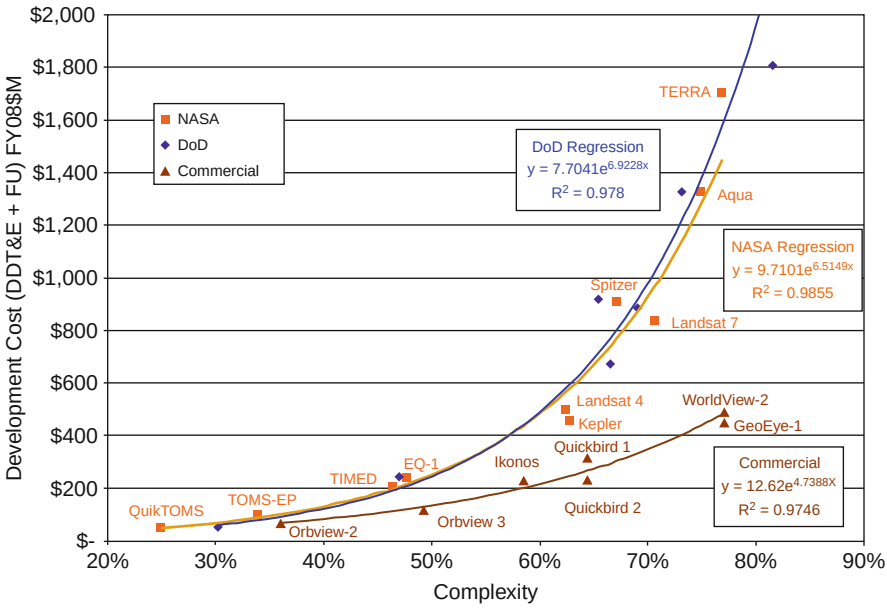


Fig. 10.2 Efficiencies in DoD and NASA production are similar but less than commercial

government telecommunications and research scientific satellites. Last are the military satellites for communications and Earth observations.

To compare systems of similar content or classes across agencies, a normalizing metric is necessary. Although some simple metrics, such as cost per kilogram, can be used to compare different systems, such a metric does not provide an assessment of the overall capability and complexity of a system. To assess the relative efficiency of different systems, the Complexity Based Risk Assessment (CoBRA) approach was chosen to assess a "dollar per unit complexity" metric.

The CoBRA complexity index is based on the order of 50 different system parameters, including mass, power, data rate, the number and type of instruments, solar array size, etc., and is used to determine the relative ranking of a system compared to over 120 other satellites. The measure is on a scale of 0–1.0, with the low values having the least capability relative to all of the spacecraft in the database, and a high value representing the most capable system (Bitten et al. 2005).

A regression of complexity versus cost for different customers reveals insights into relative efficiencies. Figure 10.2, shows the plot of a regression of complexity versus development cost for the DoD, NASA, and commercial imaging systems. Referring to Fig. 10.2, the DoD and NASA regressions, as shown at the top of the chart, indicated a substantially higher cost for a given level of complexity, relative to similar commercial systems.

A potential explanation for such a trend can be shown when looking at a similar regression for the same systems relative to the time schedule of production as

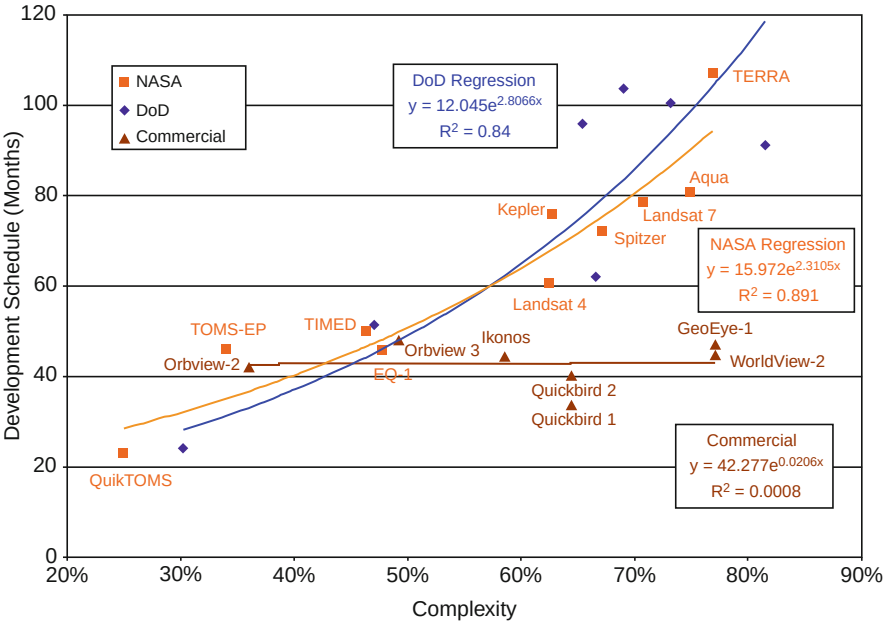


Fig. 10.3 Schedule increases with complexity for government systems

shown in Fig. 10.3. The regression for NASA and DoD missions are similar to the cost regression shown previously where schedule increases as the complexity increases. This makes intuitive sense as the development cost typically increases as schedule increases and both are greater with higher levels of complexity.

This trend, however, is not the same for commercial imaging systems. As shown, the regression for schedule relative to the increasing complexity for commercial systems is similar regardless of the level of complexity.

Commercial systems show cheaper costs and shorter manufacturing times because they rely on the same payload and spacecraft bus for each successive satellite. The commercial satellite manufacturers tend to develop "platforms" that can be used with a series of progressively larger satellites. This is in some ways comparable to the "platforms" that automobile manufacturers now use. Establishing a long-term commitment and partnership with industry providers enables this evolutionary approach where teams can build upon past experience to become more efficient for more complex future systems (An example of such a system is the QuickBird, WorldView 1 and WorldView 2 evolutionary approaches that migrated a 0.6 m imaging system on a standard bus [BCP 2000 for QuickBird] to a more capable 0.6 m imaging system on a larger, standard bus from the same provider [BCP 5000 for WorldView 1] to a 1.1 m imaging system [WorldView 2] on the same bus as used for WorldView 1. This evolutionary approach minimized risk and maximized team efficiencies.)

The evolutionary approach was developed because commercial satellites are able to take advantage of a number of options not usually applicable to government satellites, including:

- A fixed-cost tight time schedule
- A much less complex design and instrumentation than government research and development satellites
- An evolutionary approach to cutting-edge technology
- A ground system that is an integral component
- An unambiguous set of technical requirements which demand lower skill levels
- A payment schedule based on progress and in some cases incentives for efficient, timely, and reliable performance

The analysis in the study as well as comments from industry and government reviewers had the following conclusions:

- NASA and DoD spacecraft require about twice the systems engineering staff relative to what is done on commercial ventures.
- NASA and DoD spacecraft require about 1.5 times longer for assembly, integration, and testing than commercial ventures.
- On a cost per pound basis, DoD projects cost about 4 times as much as commercial projects while NASA projects cost about twice as much as commercial projects.
- On hours per drawing basis, DoD projects require more than 3 times as much time as commercial projects while NASA projects two times as much time.
- At least one company ranks productivity as follows (highest to lowest): commercial, performance-based government contracts, cost-plus award fee government contracts, DoD classified projects.
- Government-sponsored special communications satellites take more than twice as long to manufacture than comparable commercial communications satellites (over 5 years, compared to 2.5 years) and are even less efficient to manufacture because:
  - They have more reporting requirements that need formal approval
  - They have to undergo more testing
  - They have more on-site government and other personnel
  - They are designed to carry out more functions

Manufacturers of commercial satellites use proven component parts. They assemble and test them rather than invest in new technological development. They also noted the use of standardized processes and that customers do not change requirements once a contract is signed. This allows them to produce their satellites within a 24–36-month timeframe, which, as mentioned above, is approximately half the amount of time required for government projects. They also noted that commercial fixed-price contracts are easier to finance and administer than cost-plus government contracts since contract and payment schedules are negotiated between client and customer without restrictions imposed by complex government procurement regulations. Up-front and progress payments are scheduled according to milestones that are generated to encourage schedule compliance. Commercial entities also purchase risk insurance because they have incentives to deliver on

Table 10.1 Commercial satellite and government-sponsored projects compared

Category	Commercial	Government
Development trend	Evolutionary	Revolutionary
Production	Standardization and reuse of building blocks. Build multiple units	Unique designs. Build one of a kind
Requirement definition	Well understood before project start	Not well understood at project start
Requirement stability	Stable	Unstable
Stakeholders	Single customer/stakeholder	Many stakeholders
Performance specification	Specifies only performance requirements	Specifies performance requirements and methods
Design incentive	Profit driven	Science driven
Design approach	Satellite buses viewed as product line and are a known entity	Changes, especially after the start of the project, drives the design of the satellite bus
Cost and schedule	Based on known similar historical data (buy mode)	Cost and schedule estimates are optimistic (sell mode)
Funding stability	Stable	Potential annual changes (often a result of budget pressures)
Portfolio management	If a project gets into trouble, it typically gets canceled	Projects allowed to continue and usually cause collateral damage to the portfolio (a very inefficient outcome)
Procurement process	Streamlined	Long and complicated
Contract type	Incentives for early delivery and late delivery penalties	Cost-plus-type contracts
Oversight and reporting	Minimal oversight of subcontractors	Extensive oversight of primes and subcontractors
Test philosophy	Deletes non-value-added processes (profit driven)	Tends to avoid seeking waivers. Success valued on success of mission, not cost or schedule overruns

time. They noted that government programs impose far more technical reviews, changes in design, and oversight than commercial customers.

Another conclusion is that there is not that much difference in productivity among different government programs (although some of the data presented suggest that unclassified projects are more "productive" [in terms of cost efficiency] than the classified ones). Table 10.1, summarizes the full set of differences in producing satellites for the government compared to private customers. Although this study only examined the US experience, there is a reasonable expectation that similar results would be found if commercial, governmental, and military satellite projects were compared in other regions such as Europe.

In short, although both commercial satellites and government satellites share many technologies, commercial products are built and operated with profit as the objective. Government satellites, and in particular NASA satellites, are for research and development purposes and are often typically designed as first of a kind, using

cutting-edge technology and with knowledge or support of other government programs as the goal.

Because of the above reasons, there is no a priori reason to conclude that a comparison of commercial satellites and government satellites, even though they may be designed for similar purposes, should or will result in equivalent costs and performance.

Conclusion

The economics of communications satellite has evolved from a government-controlled, privately operated system to a heavily regulated, oligopolistic, somewhat competitive essential part of our economic infrastructure. From the early voice and data transmissions, there are now a wide variety of satellite services ranging from direct broadcast television to the rapid transmission of data and information for the global financial network.

The space system continues to be expensive and risky. Only relatively large companies can effectively compete for manufacturing satellites, launch services, and operations. The large number of mergers over the past 20 years is strong evidence of this, coupled with the emergence of only a few dominant firms. However, terrestrial services using the satellite-based relay and transmission are spread over many different sectors, many different companies, and many different end users. It is truly competitive and is the fastest growing part of the satellite communications business.

Also evidenced by the maturity of the industry is the international and global dimension of the industry. One of the main advantages of using satellites for communications centers on their global or at least broad regional coverage and their ability to broadcast information simultaneously from one point to many points on Earth. In the 1960s, the United States had developed the technology and had the ability to launch these satellites. This was matched only by the Soviet Union, mainly by their launch capabilities, not their advanced technology in telecommunications. Because of the strategic importance of this capability, the United States dominated the industry. Today, that has changed. The United States still has many capabilities in terms of advanced telecommunications satellite technology but very capable and competitive systems, particularly for civilian purposes, can be bought from commercial suppliers in many parts of the world and launched by many other countries as can be seen in Fig. 10.1 and in the appendices to this handbook.

From an economic perspective, there are a number of challenges facing the future of the industry. First, there are competing forms of transmission such as fiber-optic cables that did not exist when satellites were first deployed. Second, the available spectrum is limited and scarce.

Allocating spectrum for communications purposes is both an international and diplomatic exercise as well as an economic one. Nationally, it is handled differently in each nation, some using sophisticated economic means such as auctions and others using more political and less market-driven allocation schemes.

Space itself is more crowded with human-made objects. Some are controllable and working and others are older abandoned satellites or debris. These represent potential hazards to orbiting satellites. Furthermore, there are still no agreed-upon effective means of controlling the growth of the debris or ensuring that there will be a sustainable and secure future for satellite operations.

One of the more daunting and important issue facing the satellite communications industry is this increasing risk of serious damage and consequent service interruptions and the liability due to a collision due to debris or a derelict satellite. There are costs associated with developing better space sustainability. First, hardening a spacecraft when it is being built to minimize damage is expensive and adds weight to the launch payload, which entails additional expense. Second, while the spacecraft is in orbit, fuel must be reserved for additional maneuvers to avoid a collision with oncoming uncontrolled objects. Third, additional fuel must be reserved for end-of-life deorbiting or boosting to a graveyard orbit. (Or, in the future, servicing satellites may provide alternatives for end-of-life maneuvers. But at present, the cost of these still-to-be-developed services is undetermined. These types of in-space services will also face major regulatory issues and the combination of expense and administrative hurdles may not produce a viable economic business.) Fourth, additional personnel must be dedicated to minimizing debris during manufacture, operations, and possibly even when the satellite is no longer in use.

The manufacturing and operating firms would largely be the entities to incur these costs. In addition, governments now face monitoring, mitigation, regulatory costs associated with satellite communications and satellite applications. In the future, if technology permits, they may face cleanup costs. These can range from relatively trivial routine monitoring to very expensive in-orbit activities. The funds for these activities may come from a combination of governmental funds, insurance companies, and the owner/operator firm's themselves.

Although economics – the allocation of resources and the opportunity to make a profit from an investment in satellite communications businesses – will drive many aspects of this business, the involvement of governments will continue to add cost, risk, and political dimensions to any private sector activity in space. However, government's involvement has diminished somewhat as the industry has matured, and current trends indicate that the industry will continue to grow rapidly, and the degree of influence governments have over private satellite communications activities may thus also continue to diminish.

Cross-References

- ▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ▶ Fixed Satellite Communications: Market Dynamics and Trends
- ▶ History of Satellite Communications
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends

▶ Satellite Applications Handbook: The Complete Guide to Satellite Communications, Remote Sensing, Navigation, and Meteorology
▶ Satellite Communications Overview
▶ Satellite Communications Video Markets: Dynamics and Trends
▶ Satellite Orbits for Communications Satellites
▶ Space Telecommunications Services and Applications

References

R.E. Bitten, D.A. Bearden, D.L. Emmons, A quantitative assessment of complexity, cost, and schedule: achieving a balanced approach for program success. Sixth IAA international conference on low-cost planetary missions, Kyoto, 11–13 Oct 2005

W.M. Brown, H. Kahn, in *Long-Term Prospects for Developments in Space* (*A Scenario Approach*) (Hudson Institute, New York, 1977). NASW-2924, 30 Oct 1977

T. Coonce, J. Hamaker, H. Hertzfeld, R. Bitten, NASA productivity. J. Cost Anal. Parametrics **3**(1), 59–73 (2010). Society of Cost Estimating and Analysia – International Society of Parametric Analysis

Futron Corporation, 2010 Futron Corporation State of the Satellite Industry Report 2010, sponsored by the Satellite Industries Association, http://www.futron.com/resources.xml#tabs-4. Accessed 29 May 2011

Futron Corporation, 2010 Futron forecast of global satellite services demand – overview, http://www.futron.com/resources.xml#tabs-4. Accessed 29 May 2011

Futron Corporation, 2010 Telecommunications report, http://www.futron.com/resources.xml#tabs-4. Accessed 29 May 2011

M.J. Mechanick, *The Politics of the Establishment and the Eventual Privatization of the Three Major International Satellite Organizations*, White & Case, LLP, February, 2011, (unpublished manuscript)

J.N. Pelton, P. Marshall, *NASA's Unsuccessful X-Projects, Space Exploration and Astronaut Safety* (AIAA, Reston, 2006), pp. 149–178

Satellite Industries Association: Satellites 101, http://www.sia.org/satellites.html. Accessed 20 May 2011

Further Reading

R. Burkhart, in *Economics of Satellite Communications*. AIAA, Papers and Proceedings, 2000

D. Cavosa, Satellite Industries Association, COMSTAC presentation (2004), http://www.sia.org/present.html. Accessed 29 May 2011

European Satellite Operators Association, Economics of satellites, http://www.esoa.net/Economics_of_satellites.htm. Accessed June 2011

B. Henoch, Satellite technology basics, March (2007), http://www.sia.org/present.html. Accessed 29 May 2011

Z. Szajnfarber, M. Stringfellow, A. Weigel, The impact of customer–contractor interactions on spacecraft innovation: Insights from communication satellite history. Acta Astronom. 67, 1306–1317 (2010)

Satellite Communications and Space Telecommunications Frequencies

11

Michel Bousquet

Contents

Introduction 240
Radio Waves 240
Need for Radioregulations 242
Nomenclature of the Frequency and Wavelength Bands 243
Electromagnetic Waves 244
Radio Frequency Waves Characteristics and Maxwell's Equations 245
Propagation Mechanisms 250
Antenna Fundamentals 252
Path Loss in Wireless Communications 256
Tropospheric Effects on Satellite Communications 258
Introduction 258
Attenuation Effects 259
Scintillation Effects 262
Depolarization Effects 262
Conclusion 263
Ionospheric Effects on Satellite Navigation Links 263
Basic Concepts 264
Satellite-Based Navigation Technique and Influence of Ionosphere 264
Ionospheric Effects as a Function of Latitude 265
Mitigation Techniques 266
Ground-Based Ionosphere Monitoring 269
Cross-References 270
References 270

Abstract

Radio frequencies allow information to be transmitted over large distances by radio waves. The essential element to high-quality satellite communications is

M. Bousquet
Institut Supérieur de l'Aéronautique et de l'Espace (ISAE), 10, avenue Édouard-Belin, 31055 Toulouse Cedex 4, France
e-mail: michel.bousquet@isae.fr

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_13, © Springer Science+Business Media New York 2013

the assignment of radio frequency spectrum to various types of services. Only a limited amount of such spectra is assigned to earth-space radiolinks and thus the available bandwidth must be used with a high degree of efficiency. There are many technical elements associated with the efficient use of RF spectra for satellite communications and navigation and these elements are addressed in some detail in this chapter.

The basic properties of electromagnetic waves are first discussed, together with an overview of the basic electromagnetic phenomena such as reflection, refraction, polarization, diffraction, and absorption useful to define how radio waves travel in free space and in atmosphere. The basic parameters used to characterize the antennas responsible for generating and receiving these waves are introduced.

A survey of the propagation impairments (gas and rain attenuation, scintillation, etc.) due to the nonionized lower layers of the atmosphere from Ku- to Ka- and V-bands is presented. On the other hand, radio waves of Global Navigation Satellite Systems (GNSS) interact with the free electrons of the upper atmosphere ionized layers on their path to the receiver, changing their speed and direction of travel.

Introduction

Radio frequencies allow information (images, sound, and data) to be transmitted over large distances by radio waves. They are the basis of satellite communications. They are a portion of the "electromagnetic spectrum" which is the term used to describe the range of possible frequencies of electromagnetic radiations. Indeed electromagnetic radiation, a form of energy exhibiting wave-like behavior as it travels through space, is classified according to the frequency of its wave. This elemental consideration of the physics of electromagnetic phenomena is discussed in detail in ▶ Chap. 27, "Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing" by Prof. Rycroft. Just a few of the key concepts are reiterated here to explain the exploitation of electromagnetic spectra explicitly used for space communications.

Radio Waves

As discussed in the next section, any variation in time of a charge or of a magnetic moment creates coupled electric and magnetic fields characterizing an electromagnetic field. The variation in time of the electromagnetic field is the image of the time variation of the generating sources. Moreover, the electromagnetic field exhibits a space variation at the same pace: the electromagnetic wave propagates without requiring any physical support and carries energy. If the sources are periodically moving in time, i.e., oscillating at a given frequency, the electromagnetic field oscillates at the same frequency. In homogeneous free space, the electromagnetic field is also periodic in space, and the electromagnetic wave propagates with a constant velocity, the speed of light in vacuum. The wavelength is the period in space of the electromagnetic field, that is the replica of the period in time domain (i.e., proportional

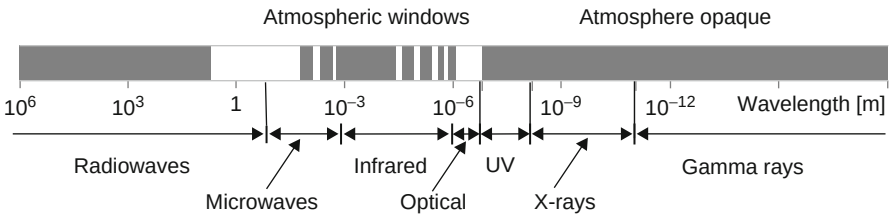


Fig. 11.1 Electromagnetic spectrum

to the inverse of frequency), and is proportional to the wave velocity. Thus either the frequency or the wavelength characterizes an electromagnetic wave.

Possible values of the wave frequency constitute a continuum called "electromagnetic spectrum." The electromagnetic spectrum, in order of increasing frequency and decreasing wavelength, includes radio waves, microwaves, infrared radiation, visible light, ultraviolet radiation, X-rays, and gamma rays (Fig. 11.1). Although, the propagation phenomenon is independent of the value of the frequency, the ability of a material to interact, namely, absorb the energy of an electromagnetic radiation, is strongly frequency dependent. Depending on frequency, some materials could be "seen" by the wave that is reflected or absorbed; others are "transparent" to the wave if no interaction occurs. This is the case with atmosphere that is transparent (or with limited attenuation) or opaque to the electromagnetic waves depending on wavelength.

Radio waves are a type of electromagnetic radiation with wavelengths in the electromagnetic spectrum longer than infrared light. More specifically, the radio spectrum includes radio waves with frequencies between 3 kHz and 300 GHz, corresponding to radio wavelengths from thousands of kilometers to under 1 mm. Naturally occurring radio waves are made by lightning, or by astronomical objects. Artificially generated radio waves are used for fixed and mobile radio terrestrial and satellite communication, broadcasting, radar and navigation systems, etc. When considering terrestrial communications, different frequencies of radio waves have different propagation characteristics in the Earth's atmosphere; long waves may cover a part of the Earth very consistently, shorter waves can reflect off the ionosphere and travel around the world. Earth-space links (satellite communication and navigation) are using shorter wavelengths (centimeter to tens of centimeters) that bend or reflect very little and travel mainly on a line of sight.

Although valid in all the electromagnetic spectrum, electromagnetic field theory has been established by Maxwell in the radio frequency range (Maxwell 1865). Maxwell noticed wavelike properties of light and similarities in electrical and magnetic observations, and proposed equations that described light waves and radio waves as electromagnetic waves that travel in space. In 1887, Heinrich Hertz demonstrated the reality of Maxwell's electromagnetic waves by experimentally generating radio waves in his laboratory. Many inventions followed, making practical the use of radio waves to transfer information through space.

The study of electromagnetic phenomena such as reflection, refraction, polarization, diffraction, and absorption is of critical importance in the study of how radio waves move in free space and over the surface of the Earth. This is the rationale for the overview on these phenomena presented in section "Electromagnetic Waves".

Need for Radioregulations

Radio spectrum is an essential resource underpinning one of world's most dynamic sectors: wireless communications. As well as telecommunications, wireless technologies support services in areas as diverse as transport, security, and environmental protection. But the spectrum is a finite resource, so its allocation requires effective and efficient coordination at global level. Radio regulations are necessary to ensure an efficient and economical use of the radio frequency spectrum by all communications systems, both terrestrial and satellite. While so doing, the sovereign right of each state to regulate its telecommunication must be preserved. It is the role of the International Telecommunication Union (ITU) to promote, coordinate, and harmonize the efforts of its members to fulfill these possibly conflicting objectives. These issues are discussed in greater details in ▶ Chap. 12.

The Radio Regulations refer to the following space radiocommunications services, defined as transmission and/or reception of radio waves for specific telecommunication applications. These are listed in Table 4.1 in ▶ Chap. 4, "Space Telecommunications Services and Applications."

Frequency bands are allocated to the above radiocommunications services to allow compatible use. The allocated bands can be either exclusive for a given service, or shared among several services. Regarding allocations the world is divided into three regions. Frequency allocations are revised regularly at the World Administrative Radio Conference (WARC).

For example, the Fixed Satellite Service (FSS) makes use of the following bands:

(a) Around 6 GHz for the uplink and around 4 GHz for the downlink (systems described as 6/4 GHz or C-band). These bands are occupied by the oldest systems (such as INTELSAT, American domestic systems, etc.) and tend to be saturated.
(b) Around 8 GHz for the uplink and around 7 GHz for the downlink (systems described as 8/7 GHz or X-band). These bands are reserved, by agreement between administrations, for government use.
(c) Around 14 GHz for the uplink and around 12 GHz for the downlink (systems described as 14/12 GHz or Ku-band). This corresponds to current operational developments (such as EUTELSAT, SES, etc.).
(d) Around 30 GHz for the uplink and around 20 GHz for the downlink (systems described as 30/20 GHz or Ka-band). These bands are raising interest due to large available bandwidth and little interference due to present rather limited use. This corresponds to the new developments of high-capacity systems for Internet access (such as VIASAT, KA-SAT, etc.).

The bands above 30 GHz (Q- and V-band, possibly W) will be used eventually in accordance with developing requirements and technology.

The Mobile Satellite Service (MSS) makes use of the following bands:

(a) VHF (very high frequency, 137–138 MHz downlink and 148–150 MHz uplink) and UHF (ultrahigh frequency, 400–401 MHz downlink and 454–460 MHz uplink). These bands are for non-geostationary systems only.

(b) About 1.6 GHz for uplinks and 1.5 GHz for downlinks, mostly used by geostationary systems such as INMARSAT; and 1,610–1,626.5 MHz for the uplink of non-geostationary systems such as GLOBALSTAR.
(c) About 2.2 GHz for downlinks and 2 GHz for uplinks for the satellite component of IMT2000 (International Mobile Telecommunications).
(d) About 2.6 GHz for uplinks and 2.5 GHz for downlinks.
(e) Frequency bands have also been allocated at higher frequencies such as Ka-band.

The Broadcasting Satellite Service (BSS) makes use of downlinks at about 12 GHz. The uplink is operated in the FSS bands and is called a *feeder link*.

The Radio Navigation Satellite service (RNSS) makes use of frequencies in L-band, mainly from space to earth (downlink). Before WARC-2000, RNSS allocations were divided in two parts:
(a) Lower L-band, from 1,215 to 1,260 MHZ, called L2
(b) Upper L-band, from 1,559 to 1,610 MHz, called L1
These bands have been split between the GPS and GLONASS systems, for instance, 1,563–1,587 MHz for GPS and 1,597–1,617 MHz for GLONASS.

At the WARC-2000, new frequency bands have been added to accommodate the needs of new GNSS (Global Navigation satellite Systems) such as GALILEO.
(a) 1,164–1,215 MHZ to be shared between new GPS signals (called L5 from 1,164 to 1,191.795 MHz, carrier frequency 1,176.45 MHz), GALILEO signals (called E5 from 1,164 to 1,215 MHz, carrier frequency 1,191.795 MHz, further subdivided into E5a and E5b with carrier frequencies 176.45 MHz and 1,207.14 MHz, respectively), new GLONASS signals (called L3 from 1,164 to 1,215 MHz), and others (COMPASS B2, China)
(b) 1,260–1,300 MHz called E6, to be shared between GALILEO signals (called E6, with carrier frequency 1,278.75 MHz) and others (COMPASS B3)
(c) 5,010–5,030 MHz at C-band

In addition to frequency bands used by the satellites to transmit navigation signals (downlink), bands 1,300–1,350 MHz and 5,000–5,010 MHz could be used to uplink dedicated signals to the satellites.

It should be noted that the new proposed GNSS systems are planning to share the upper L1-band (called E1 1,559–1,591 MHz with Galileo, B1 with COMPASS, etc.) (Table 11.1).

The following table summarizes the above discussion.

Nomenclature of the Frequency and Wavelength Bands

The ITU Radiocommunications Assembly, in the Recommendation ITU-R V.431, has defined a nomenclature to be used for the description of frequency and wavelength bands, given in Table 11.2.

Certain frequency bands are sometimes designated by letter other than the symbols and abbreviations recommended in the above Table. The symbols in question consist of capital letters which may be accompanied by an index (usually

Table 11.1 Frequency allocations

Radiocommunications service	Typical frequency bands for uplink/downlink	Usual terminology
Fixed satellite service (FSS)	6/4 GHz	C-band
	8/7 GHz	X-band
	14/12–11 GHz	Ku-band
	30/20 GHz	Ka-band
	50/40 GHz	V-band
Mobile satellite service (MSS)	1.6/1.5 GHz	L-band
	30/20 GHz	Ka-band
Broadcasting satellite service (BSS)	2/2.2 GHz	S-band
	12 GHz	Ku-band
	2.6/2.5 GHz	S-band
Radionavigation satellite service (RNSS)	1.164–1.3 GHz (down)	Lower L-band
	1.559–1.617 GHz (down)	Upper L-band
	5.000–5.010 GHz (up)	C-band
	5.010–5.030 GHz (down)	C-band

Table 11.2 Nomenclature of frequency and wavelength bands

Band number	Symbols	Frequency range (lower limit exclusive, upper limit inclusive)	Corresponding metric subdivision
3	ULF	300–3,000 Hz	Hectokilometric waves
4	VLF	3–30 kHz	Myriametric waves
5	LF	30–300 kHz	Kilometric waves
6	MF	300–3,000 kHz	Hectometric waves
7	HF	3–30 MHz	Decametric waves
8	VHF	30–300 MHz	Metric waves
9	UHF	300–3,000 MHz	Decimetric waves
10	SHF	3–30 GHz	Centimetric waves
11	EHF	30–300 GHz	Millimetric waves

a small letter). There is at present no standard correspondence between the letters and the frequency bands concerned, and the same letter may be used to designate a number of different bands. For information, letter designations used by some authors in the field of space communications are indicated in Table 11.1.

Electromagnetic Waves

The basic properties of electromagnetic waves traveling in free space and in atmosphere considered as a uniform medium are provided in the following section and can be specified by a small set of descriptive parameters used to define the behavior of waves.

Radio Frequency Waves Characteristics and Maxwell's Equations

The *Maxwell's equations* (Maxwell 1865) specify the relationships between the variations of the vector electric field E and the vector magnetic field H in time and space within a medium which characterize propagating electromagnetic waves.

Both Faraday's and Ampere's laws are included in Maxwell's equations but, the decisive contribution of Maxwell was to complement the Faraday's law in linking the time variation of the electric induction to the variation in space of the magnetic field.

Maxwell definitely establishes that any time variation in electric induction results in the apparition of time-variant magnetic field and conversely any time variation in magnetic induction results in time-varying electric field.

An oscillating electric field produces a magnetic field, which itself oscillates to recreate an electric field and so on. This phenomenon is represented by the so-called curl equations in the set of Maxwell's equations. Indeed Maxwell describes the way electric and magnetic field lines spread out (<< *diverge* >>) and circle around (<< *curl* >>). Without entering in a rigorous derivation and too many details, the Maxwell's equations could be explained as follows:

- *An electric field is produced by a time-varying magnetic field.*
- *A magnetic field is produced by a time-varying electric field or by a current.*

The interplay between the two fields stores energy and hence carries power.

The E field strength is measured in volts per meter and is generated by either a time-varying magnetic field or by a free charge.

The H field is measured in amperes per meter and is generated by either a time-varying electric field or by a current.

The properties of the propagating medium are incorporated in Maxwell equations through the constants ε, the *permittivity* of the medium, and μ, the *permeability* of the medium, which relate electric and magnetic fields $\vec{\boldsymbol{E}}$ and $\vec{\boldsymbol{H}}$ to electric and magnetic inductions $\vec{\boldsymbol{D}}$ and $\vec{\boldsymbol{B}}$:

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} \text{ and } \boldsymbol{B} = \mu \boldsymbol{H}$$

The scalars ε and μ (tensors in the general case of anisotropic media) model the electric and magnetic properties of the medium as it reacts to the fields producing induction. The permittivity of the medium is expressed in Farad per meter and the permeability of the medium in Henry per meter.

These constants are expressed relative to the values $\mu_0 = 4\pi 10^{-7}$ $\mathrm{Hm}^{-1}$and $\varepsilon_0 = 8.854\ 10^{-12} = 10^{-9}/36\ \pi\ \mathrm{Fm}^{-1}$ in free space as

$$\mu = \mu_r \mu_0 \text{ and } \varepsilon = \varepsilon_r \varepsilon_0$$

where μ_r and ε_r are the relative values ($\varepsilon_r = \mu_r = 1$in free space). Strictly speaking, *free space* provides a reference to vacuum, but the same values can be used as good approximations when performance is compared to dry air at "typical" temperature and pressure.

Maxwell's equations are first-order evolution equations in time and space, symmetrically coupled in an open homogenous space. This symmetry, in free space, allows deriving, for both electric and magnetic fields, uncoupled second-order harmonic equations or wave equations which sustain independent oscillatory or harmonic solutions in both space and time. In complex notation, electric and magnetic field amplitudes read as:

$$E = \bar{E}e^{j\omega t}e^{j2\pi\frac{z}{\lambda}} \text{ and } H = \bar{H}e^{j\omega t}e^{j2\pi\frac{z}{\lambda}}$$

where E and H are the complex amplitudes of the field. The pulsation ω is related to the frequency f of the electromagnetic field by:

$$\omega = 2\pi f$$

and λ, the period in space, is the wavelength that is related to the frequency by:

$$\lambda = \frac{v}{f}$$

with v the phase velocity which takes the value of the speed of light in the considered medium given by:

$$v = \frac{1}{\sqrt{\mu\varepsilon}} = \frac{c}{\sqrt{\mu_r\varepsilon_r}}$$

In free space, the speed of the light is given by:

$$c = \frac{1}{\sqrt{\mu_0\varepsilon_0}} = 3.10^8 m.s^{-1}$$

It is common practice to introduce the *wave number* k or the *propagation constant* $\gamma = jk$, related to the wavelength by $k = \frac{2\pi}{\lambda} = \omega\sqrt{\mu\varepsilon}$.

When considering the range of radio frequencies for satellite communications, say from around 300 MHz to 30 GHz, the free space wavelength varies from 1 m to 1 cm.

The properties of the wave are characterized by a given value of amplitude, frequency, and phase. When one (or a combination) of these parameters is varied according to the amplitude of an information signal, this allows information to be carried in the wave between its source and destination. This process, called carrier *modulation*, is a key concept in satellite communications. Modulation is discussed in some detail in ▶ Chap. 13, "Satellite Radio Communications Fundamentals and Link Budgets" by Daniel Glover.

Field Structure and Plane Wave

Far from the source and from any obstacle, as in free space, fields of the radio waves may be considered having a spherical structure and dispersing energy in the radial direction.

The amplitudes of the fields $\bar{E}$ and $\bar{H}$ in the above expressions of the fields are decaying as function of the inverse of the distance r from the source:

$$\bar{E} = \bar{E}(r) \propto \frac{E_0}{r} \text{ and } \bar{H} = \bar{H}(r) \propto \frac{H_0}{r}$$

This radial variation translates into a bounded value of the integration of the power density over the three-dimensional space, which is equal to the power delivered by the source.

The *power density* at a distance r is related to the amplitude of the electric and magnetic fields components $E_{0\perp}$ and $H_{0\perp}$ on the plane tangent to the sphere of radius r centered on the sources, which represents the wave front. In the plane of the wave front, electric and magnetic field components are perpendicular to each other.

The amplitudes $E_{0\perp}$ and $H_{0\perp}$ are related through the *wave impedance* ζ such as:

$$E_{0\perp} = \zeta H_{0\perp}$$

The wave impedance is a function of the permittivity and the permeability of the medium:

$$\zeta = \sqrt{\frac{\mu}{\varepsilon}}.$$

In vacuum, $\zeta = \sqrt{\frac{\mu_0}{\varepsilon_0}} = 120\pi = 377\Omega$

As far as ε_r and μ_r are real and positive constants, the medium is said to be *lossless* and the amplitudes $E_{0\perp}$ and $H_{0\perp}$ are in phase. The wave front is "locally plane" (*plane wave*).

When the wave interacts with the medium (*lossy media*), energy is removed from the wave and converted to heat. Absorption by the media is taken into account considering a complex permittivity and/or a complex permeability, the imaginary part of which yields to an imaginary part in the wave number, introducing an exponential attenuation α in the spatial variation of fields:

$$E = \bar{E}e^{-\alpha z}e^{j\omega t}e^{j2\pi\frac{z}{\lambda}} \text{ and } H = \bar{H}e^{-\alpha z}e^{j\omega t}e^{j2\pi\frac{z}{\lambda}}$$

It should be noted that, as the wave impedance is complex, the amplitudes $E_{0\perp}$ and $H_{0\perp}$ are no longer in phase.

The constant α is known as the *attenuation constant*, with units of per meter (m^{-1}), which depends on the permeability and permittivity of the medium, the frequency of the wave, and the *conductivity* of the medium, σ, measured in Siemens, or ohm per meter ($\Omega\ m^{-1}$). Together σ, μ, and ε are known as the *constitutive parameters* of the medium. In nonlinear media, ε_r and μ_r depend on the field's intensity.

Poynting Vector and Power Density

The Poynting vector $\boldsymbol{S}$ measured in watts per square meter describes the magnitude and direction of the power flow carried by the wave per unit of surface perpendicular to the direction of propagation z, i.e., the *power density* of the wave. The instantaneous power density is represented by the instantaneous Poynting vector $\boldsymbol{S}$ defined as the vector product between electric and magnetic field:

$$\boldsymbol{S}(t) = \boldsymbol{E}(t) \wedge \boldsymbol{H}(t)$$

The total power leaving a surface A is given by the flux of the Poynting vector through A. It should be noted that the power flux is carried out by the field components tangential to the surface. Field components normal to the surface do not contribute to this flux.

In the case of harmonic variation in time domain, the fields may be written as $\boldsymbol{E} = \boldsymbol{E}(r)e^{j\omega t}$ and $\boldsymbol{H} = \boldsymbol{H}(r)e^{j\omega t}$ and therefore the expression of the instantaneous Poynting vector S reads as:

$$\boldsymbol{S}(t) = \frac{1}{2}\left[\Re e(\boldsymbol{E}(r) \wedge \boldsymbol{H}^*(r)) + \Re e(\boldsymbol{E}(r) \wedge \boldsymbol{H}^*(r))e^{2j\omega t}\right]$$

where the asterisk (*) denotes the complex conjugate.

In practice, the time average of the power flow over one period is considered:

$$S_{av} = \frac{1}{2}\left[\Re e(\boldsymbol{E}(r) \wedge \boldsymbol{H}^*(r))\right]$$

As mentioned above, only the field component in the plane normal to the propagation direction z contributes to the power flow. Introducing the notation of the transverse field of the locally plane waves introduced in the previous session, the average power density could be written as:

$$S_{av} = \frac{1}{2}\left[\Re e\left(E_\perp(r) \wedge H_\perp^*(r)\right)\right] = \frac{1}{2}|E_\perp|^2 \Re e(\zeta)^{-1}$$

E, **H**, and $\boldsymbol{S}_{ave}$ form a right-hand set, i.e., $\boldsymbol{S}_{\text{ave}}$ is in the direction of movement of a right-hand corkscrew, turned from the $\boldsymbol{E}$-direction to the $\boldsymbol{H}$-direction.

Wave Polarization

The time variation of the amplitude and the orientation of the extremity of the transverse component of the electric field define the *polarization* of the wave. In the more general case, the curves traced by the extremity of this field component along the time at a given position or, similarly, along the direction of propagation at a given time, are ellipses. The wave is said to be *elliptically polarized.*

In Fig. 11.2, the electric field is parallel to the x-axis; the wave is *linearly polarized* (x-polarized). This wave could be generated by a straight wire antenna parallel to the x-axis.

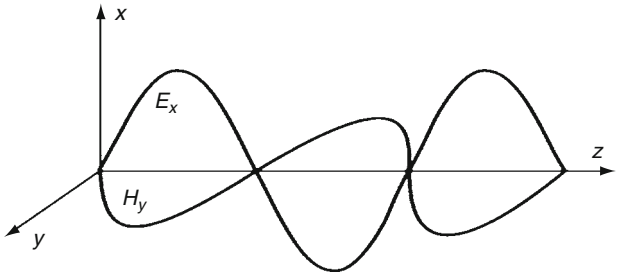


Fig. 11.2 Linearly polarized time harmonic plane wave

An entirely distinct y-polarized plane wave could be generated with the same direction of propagation and recovered independently of the other wave using pairs of transmit and receive antennas with opposite polarization. This principle called "frequency reuse by orthogonal polarizations" is of critical importance in satellite communications in order to provide two independent communication channels in the same portion of allocated bandwidth.

The waves described above are linearly polarized since the electric field vector has a single direction along the whole of the propagation axis. If two plane waves of equal amplitude and orthogonal polarization are combined with a 90° phase shift, the resulting wave is *circularly polarized* (CP), in that the motion of the electric field vector will describe a circle centered on the propagation vector. The field vector will rotate by 360° for every wavelength traveled. Two solutions are possible: either *right-hand circularly polarized* (*RHCP*) or *left-hand circularly polarized* (*LHCP*); RHCP describes a wave with the electric field vector rotating clockwise when looking in the direction of propagation.

LHCP and RHCP waves are orthogonal and can also be considered for *frequency reuse* in satellite communications. Circularly polarized waves are commonly used in satellite communications, since they can be generated and received using antennas that do not require specific orientation along the direction of propagation (in contrast, the use of linear polarization implies that the transmit and receive electric field should match). Another advantage is that circularly polarized communication channels are not influenced by Faraday rotation (see section "Tropospheric Effects on Satellite Communications").

If the component waves have different amplitudes and arbitrary phase difference, the result is the generic elliptically polarized wave introduced above, where the electric field vector still rotates at the same rate but varies in amplitude with time. In this case the wave is characterized by the ratio between the maximum and minimum values of the instantaneous electric field, known as the axial ratio AR (Fig. 11.3):

$$\mathrm{AR} = E_{max}/E_{min}$$

AR is defined to be positive for left-hand polarization and negative for right-hand polarization.

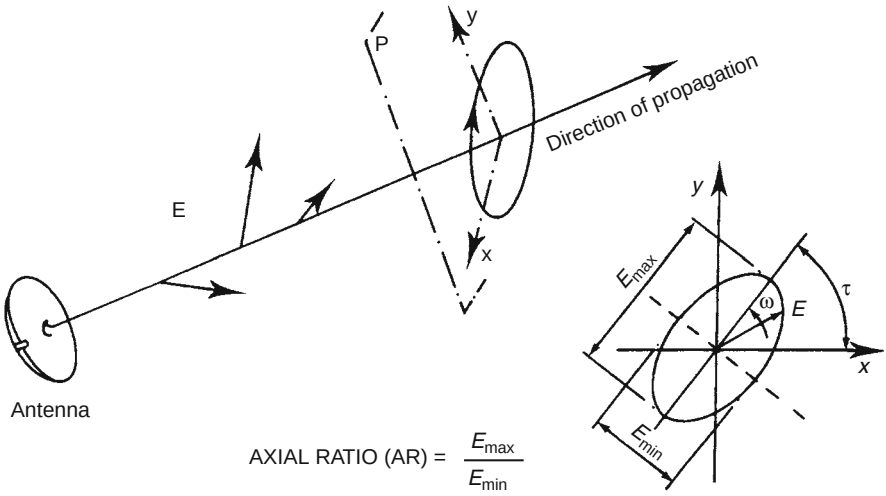


Fig.11.3 Elliptical polarization and axial ratio

Propagation Mechanisms

In the previous section, the environment in which waves were propagating consisted of media within which the constitutive parameters did not vary in space. In practice, the real environment includes boundaries between media (between air and the ground, between, from Earth to space, etc.). These boundary effects give rise to changes in the amplitude, phase, and direction of propagating waves. Four main types of propagation mechanisms exist and can be used to describe the interaction of the electromagnetic waves with the environment. These mechanisms include reflection, refraction, scattering, and diffraction, and they all affect the amplitude, direction, and phase of the propagating radio waves.

Reflection, Refraction, and Transmission

Whenever a radio wave impinges to an obstruction having different material parameters than the propagation medium and greater dimensions than the wavelength, reflection and transmission will occur. Applying Maxwell's equations in the case of smooth homogeneous interface and mediums, the interaction behavior of the propagating wave with these interface mediums results in that two new waves are produced, each with the same frequency as the incident wave. Both waves have their Poynting vectors in the plane that contains both the incident propagation vector and the normal to the surface, called the *scattering plane*. The first wave, called *reflected* wave, propagates within medium 1 away from the boundary, making an angle θ_r to the normal. The second, called *transmitted* wave, results from the mechanism of *refraction*, and travels on into medium 2 making an angle θ_t to the surface normal (Fig. 11.4).

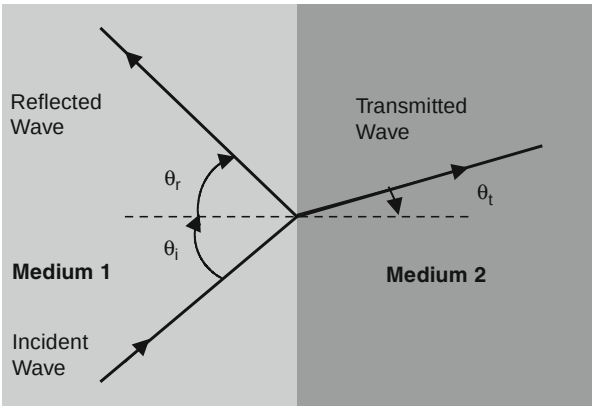


Fig. 11.4 Plane boundary model

Snell's law of reflection states that the angle θ_r of the reflected field is equal to the angle of the incident field, θ_i, to the interface. Snell's law of refraction states that the refracted angle, θ_t, is a function of the incident angle and the materials of the two media:

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{n_2}{n_1}$$

where n_1 and n_2 are the refractive indices of the two media. The *refractive index* is obtained by using the constitutive parameters of the medium:

$$n = \sqrt{\varepsilon_r \mu_r}$$

The *refractive index* is the ratio of the free space phase velocity, c, to the phase velocity in the medium v. Note that the frequency of the wave is unchanged following reflection and transmission; instead the ratio $v/\lambda = f$ is maintained everywhere, then the wave within the denser medium (higher permittivity and permeability) has a smaller phase velocity and longer wavelength than in free space.

In addition to the change of direction, the interaction between the wave and the boundary also causes the energy to be split between the reflected and transmitted waves. Several methods can be used to calculate the reflected and the transmitted field. An introduction to the topic is provided considering the simple boundary model. The amplitudes of the reflected and transmitted waves are given relative to the incident wave amplitude by the Fresnel reflection and transmission coefficients. The Fresnel coefficients depend on the constitutive parameters of the materials, the polarization of the incident field, and the angle of incidence. The coefficients are different for parallel or perpendicular polarization.

	Perpendicular polarization	Parallel polarization
Reflection coefficient	$R_\perp = \dfrac{n_2\cos\theta_i - n_1\cos\theta_t}{n_2\cos\theta_i + n_1\cos\theta_t}$	$R_\parallel = \dfrac{n_2\cos\theta_t - n_1\cos\theta_i}{n_2\cos\theta_t + n_1\cos\theta_i}$
Transmission coefficient	$T_\perp = \dfrac{2n_2\cos\theta_i}{n_2\cos\theta_i + n_1\cos\theta_t}$	$T_\parallel = \dfrac{2n_2\cos\theta_i}{n_2\cos\theta_t + n_1\cos\theta_i}$

If E_i is the incident field on the interface of the media, then the reflected and transmitted field can be calculated by using the Fresnel coefficients. Therefore:

$$E_r = RE_i \text{ and } E_t = TE_i$$

The formulations described until now describe the case of a *lossless medium*. For *lossy medium* the same formulations can be used but the electrical permittivity ε that is used must be replaced by the complex permittivity:

$$\varepsilon_{complex} = \varepsilon' - j\varepsilon''$$

where ε' is the real part of the complex permittivity and ε'' the imaginary part. The *loss tangent* δ of a material or the conductivity σ can also be used.

Rough Surface Scattering

The reflection mechanism discussed so far refers to the case where the surface is considered to be smooth (specular reflection). This condition assumes that the reflection angle is equal to the incidence angle and that all the scattered energy is concentrated in the reflected ray. If the surface is considered to be rougher, the incident ray will be scattered from a large number of positions on the surface. This broadens the scattered energy, which effectively means that the energy in the specular direction is reduced. The degree of scattering depends on the incidence angle and on the roughness of the surface compared to the wavelength of transmission. To account for scattering, the reflection coefficient needs to be multiplied by the surface roughness factor, which is less than unity and depends exponentially on the standard deviation of the surface roughness.

Diffraction

Diffraction occurs when the path between the transmitter and the receiver is obstructed by a physical object. It could be noticed that some field power could be measured in the zone behind the object even though a direct line of sight does not exist between the transmitter and the area behind the obstacle. The effect of diffraction can be explained by Huygens principle, which states that all points on a wave front can be considered as point sources for the production of secondary wavelets, and that these wavelets combine to produce a new wave front in the direction of propagation. Diffraction is caused by the propagation of secondary wavelets into the shadow region, giving rise to a bending of waves behind the obstacle. The field strength of a diffracted wave in the shadowed region is the vector sum of the electric field components of all the secondary wavelets in the space around the obstacle.

Antenna Fundamentals

The previous sections described wave interactions with the propagation media. This section introduces the antennas responsible for generating and receiving these waves. The parameters that characterize an antenna and its application are

discussed further in ▶ Chap. 13, "Satellite Radio Communications Fundamentals and Link Budgets."

Concept of Antenna

Most fundamentally, an antenna is a way of converting currents or guided waves traveling in one dimension present in a waveguide, microstrip, or transmission line, into radiating waves traveling in free space, carrying power away from the transmitter in three dimensions into free space (or vice versa). As a consequence of Maxwell's equations, the radiated wave is caused by current charges oscillating in periodic motion due to the excitation by a sinusoidal transmitter. Close to an antenna, the field patterns change rapidly with distance and include both reactive energy (which oscillates toward and away from the antenna, appearing as a reactance which only stores but does not dissipate energy) and radiating energy. Further away, the reactive fields are negligible and only the radiating energy is present, resulting in a variation of power with direction which is independent of distance. These regions are conventionally divided at a radius $R = 2\,L^2/\lambda$, where L is close to the dimension of the antenna and λ the wavelength. Within that radius is the *near-field* or *Fresnel region*, while beyond it lies the *far-field* or *Fraunhofer region*. Within the far-field region the wave front behaves as spherical waves, so that only the power radiated in a particular direction is of importance, rather than the particular shape of the antenna. Measurements of the power radiated from an antenna have to be made in the far field region (or special account has to be taken of the reactive fields if in the near-field).

Radiation Pattern

The *radiation pattern* of an antenna is a plot of the far-field radiation from the antenna. It is convenient to work in spherical coordinates (r, θ, φ) rather than Cartesian coordinates, with the antenna placed at the origin. The power radiated from an antenna is typically plotted per unit solid angle, that is the *radiation intensity* U (watts per unit solid angle). This is obtained by multiplying the power density at a given distance by the square of the distance r, where the power density S (watts per square meter) is given by the magnitude of the time-averaged Poynting vector:

$$U = r^2 S$$

In such a way, the radiation pattern is the same at all distances from the antenna, provided that r is within the far field. For example, in the case of an antenna radiating equally in all directions, an *isotropic* antenna, if the total power radiated by the antenna is P, then the power is spread over a sphere of radius r, so the power density at this distance and in any direction is:

$$S = P/4\pi r^2$$

The radiation intensity, independent of r, is:

$$U = r^2 S = P/4\pi$$

Radiation pattern is plotted by normalizing the radiation intensity by its maximum value.

Directivity and Gain

The *directivity D* of an antenna, function of the direction, is defined by the ratio of the radiation intensity of the antenna in direction *(θ, φ)* with respect to the mean radiation intensity in all directions, or in other terms, with respect to the radiation intensity of an isotropic antenna radiating same total power. Sometimes directivity is specified without referring to a direction. In this case the term "directivity" implies the maximum value of $D(\theta, \varphi) = D_{max}$.

The *power gain* G, or simply the *gain*, of an antenna is the ratio of its radiation intensity to that of an isotropic antenna radiating the same total power as accepted by the real antenna. The *gain* differs from the *directivity* by the ratio of the power radiated (which is proportional to the so-called *radiation resistance)*, to the power accepted by the antenna (which is proportional to the sum of the radiation resistance and the *loss resistance)*. Indeed, in practice, some power is lost within the antenna itself due to losses (*loss resistance*) in either the conducting or dielectric parts of the antenna.

Also, one should pay attention that transfer of power between microwave components connected together depends on the relative values of source and load impedance ($Z = R + jX$, with R resistance, X reactance).

The maximum of the source power is delivered to the antenna if the source impedance Z_s ($Z_s = R_s + jX_s$) and the total antenna impedance Z_a ($Z_a = R_r + R_l + jX_a$, where R_r is the radiation resistance and R_l is the loss resistance) are complex conjugates.

When antenna manufacturers specify simply the gain of an antenna, they are usually referring to the maximum value of G. The use of the isotropic antenna as a reference is emphasized by giving the gain units of dBi (see end of paragraph):

$$G[\text{dBi}] = 10 log G$$

A radiation pattern plot for a generic directional antenna is shown in Fig. 11.5, illustrating the *main lobe*, which includes the direction of maximum radiation (sometimes called the *boresight* direction) and several *side lobes* separated by *nulls* where no radiation occurs. Note that the drawing represents a cut of the actual three-dimensional pattern in a plane containing the direction of maximum intensity (or maximum directivity or gain).

Some common parameters used to compare radiation patterns are defined as follows:

- The *half-power beamwidth* (HPBW), or θ_{3dB} *beamwidth*, is the angle subtended by the half-power points of the main lobe.
- The *sidelobe level* is the amplitude of the biggest sidelobe, usually expressed in decibels relative to the peak of the main lobe.
- The decrease of envelope of the peak of side lobes as a function of the off-axis angle.

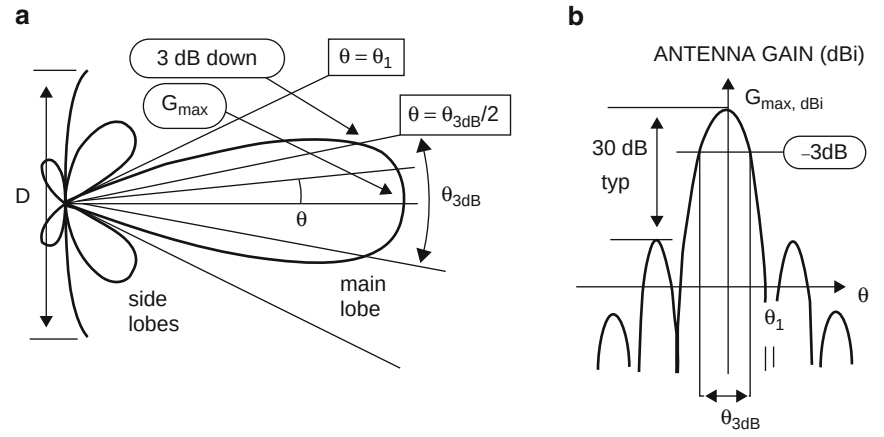


Fig. 11.5 Antenna radiation pattern

The Bel is a logarithmic unit of power ratio, where one Bel corresponds to an increase in power by a factor of 10 relative to a given reference power. It is conventional to express a ratio in decibels (one-tenth of a Bel) relative to some standard reference parameter. This reference is usually indicated by appending an appropriate letter. For instance, the power produced by a transmitter can be expressed in dBW, while here the gain of the antenna is in dBi (referring to the gain of an *isotropic antenna*).

Antenna Aperture

In the previous sections, the antenna has been considered as a transmitting device. It could also be used in receive mode. The performance of a receive antenna can be derived from the *reciprocity theorem:* If a voltage is applied to the ports of an antenna A and the current measured at the ports of another antenna B, then an equal current will be obtained at the ports of antenna A if the same voltage is applied to the ports of antenna B. A useful consequence of this theorem is that the antenna directivity and gain is the same whether used for receiving or transmitting, so all of the gain and pattern characteristics discussed above are fully applicable in receive mode. Another parameter that plays a critical role in the performance of a receiving antenna is the antenna effective aperture A_{eff} [m^2].

If an antenna is used to receive a wave with a power density S watts per square meter, it will produce a power in its terminating impedance (receiver input impedance) of P_r watts. The constant of proportionality between P_r and S is the *effective aperture* of the antenna:

$$P_r = A_{eff} S$$

For some antennas, such as horn or dish antennas, the aperture has some physical interpretation (related to diameter of the physical surface) but the concept is just as valid for all antennas.

It could be shown that the maximum receive antenna gain $G_{r,\,max}$ is related to the effective aperture as follows:

$$G_{r,max} = A_{eff}(4\pi/\lambda^2)$$

Path Loss in Wireless Communications

The path loss between a pair of antennas is the ratio of the transmitted power to the received power. It includes all of the possible elements of loss associated with interactions between the propagating wave and any objects between transmit and receive antennas.

Free Space Loss

Consider two antennas separated by a distance r great enough that the antennas are in each other's far field and oriented such as their direction of maximum gain is aligned. If P_t is the power accepted by the transmit antenna and $G_{r\,max}$ the antenna gain, then the power density incident on the receive antenna is:

$$\Phi = P_t G_{t\,max}/4\pi r^2$$

Considering a receiving antenna with maximum gain $G_{r\ max}$ and with aperture A_{eff}, the received power expresses as:

$$P_r = A_{eff} P_t G_{t\,max}/4\pi r^2$$

Introducing the relationship between A_{eff} and $G_{r\ max}$ and expressing the ratio of the received power to the transmitted power:

$$P_r/P_t = G_{t\,max} G_{r\,max}(\lambda/4\pi r)^2$$

This expression defines the term $(4\pi r/\lambda)^2$ which is called the *free space loss* L_{fs}. Because of square law dependence on both frequency and distance, the free space loss increases by 6 dB for each doubling in either frequency or distance.

With frequency in gigahertz and distance R in kilometers, the free space loss in decibels expresses as:

$$L_{fs}(\mathrm{dB}) = 92.4 + 20\log R + 20\log f$$

Polarization Mismatch Loss

The expression assumes ideal matching of polarization state of the receive antenna with that of the incoming wave. If not, a *polarization mismatch loss* should be accounted for, defined as the ratio between the power received by the antenna and the power which would be received by an antenna perfectly matched to the incident

wave. In free space, the polarization state of the received wave is the same as that of the transmitter antenna. In more complicated media, which may involve polarization-sensitive phenomena such as reflection, refraction, and diffraction, the wave polarization is modified during propagation in accordance with the discussions in the previous sections.

Antenna Gain Fall-Out due to Depointing

The calculation above assumed that the directions of maximum gain of the antennas are aligned. In practice, each antenna will introduce some depointing, that is the angular difference between the direction of maximum gain and the direction of the other antenna. Therefore the antenna gains may not necessarily be the maximum values.

Although some books introduce the gain fall-out due to depointing in the path loss, to describe the propagation medium independently of system component gains, it is preferable that antenna gain values to be used are those corresponding to the direction of the other antenna.

Note that the same concept may apply for the polarization mismatch that could be considered as a reduction of the available antenna gain.

Atmospheric Effects for Earth-Space Radio Links

The concept of free-space loss assumes that the propagation medium between the antennas is lossless. This is a reality in space, for instance, in the case of intersatellite links. However, when considering an earth-satellite link, the radio frequency waves travel through the atmosphere and interact with the propagation medium.

Effects of the nonionized atmosphere (troposphere, i.e., lower layers of atmosphere) need to be considered at all frequencies, but become critical above a few GHz and for low elevation angles. The propagation loss on an Earth-space path, relative to the free-space loss, is the sum of different contributions as follows:

- Attenuation by atmospheric gases
- Attenuation by rain and clouds
- Scintillation

Effects of the troposphere on a satellite link are discussed in some detail in section "Tropospheric Effects on Satellite Communications".

Around and below 1 GHz, ionospheric effects play a particularly important role in the performances of satellite navigation systems. This is the topic addressed in section "Ionospheric Effects on Satellite Navigation Links".

Effect of Environment with Mobile Communications

The propagation effects in the case of satellite mobile communications are a combination of different phenomena. Indeed in addition to the path loss which is an overall decrease in field strength as the distance between the transmitter and the receiver increases, blockage and multipath effects create *shadowing* and *fast fading* which appear as time-varying processes between the antennas. Shadowing is the result of the varying nature of blockages of the radio frequency path by trees and buildings in between the satellite and the receiver. Fast fading results from the

constructive and destructive interference between multiple waves reaching the mobile after reflections on the local environment of the mobile. All of these processes vary as the relative positions of the transmitter and receiver change and as any contributing objects between the antennas are moved. Fast fading involves variations on the scale of a half-wavelength (30 cm at 1.5 GHz) and introduces variations as large as 20–30 dB. These are complex phenomena, the modeling of which is out of the scope of this chapter, and a detailed analysis can be found in (Castanet 2007).

Link Budget

The path loss as discussed above between a pair of antennas is the ratio of the transmitted power to the received power. Therefore, knowing the transmitted power, it is possible to determine how much carrier power will be available at the receiving end input. In practice, the communication link performance is function of the ratio of the amount of available carrier power to the amount of unwanted signals (called noise). Indeed, at the receiver input, the wanted carrier power is not the only contribution. Different unwanted signals contribute to what is called the system noise temperature T. Those sources of noise include the noise captured by the antenna and generated by the feeder, which can actually be measured at the receiver input, plus the noise generated downstream in the receiver, which is modeled as a fictitious source of noise at the receiver input, treating the receiver as noiseless. The system noise temperature T can be used to infer the amount of noise power spectral density No (W/Hz).

The radio frequency link performance is evaluated as the ratio of the received carrier power, C, to the noise power spectral density, No, and is quoted as the C/No ratio, expressed in hertz (Hz).

The calculation of carrier power, noise power spectral density for a complete communication link is the *link budget*, and the influence of the C/No ratio on the end-user performance (e.g., bit error rate, BER) depending on the type of modulation and coding will be discussed in detail in ▶ Chap. 13, "Satellite Radio Communications Fundamentals and Link Budgets."

Tropospheric Effects on Satellite Communications

Satellite telecommunication links operating at frequencies above 10 GHz are disturbed by tropospheric phenomena that degrade link availability and service performance. This section presents a review of the propagation properties for satellite telecommunication systems operating at Ku-, Ka-, and Q/V-bands.

Introduction

Most of today's satellite systems for fixed communications and broadcast are operating in the conventional Ku-band. But it should be noted that there is a strong interest from satellite operators and the space industry for using bands above Ku,

mainly thanks to the wider bandwidths (up to 3 GHz at V-band) allocated to Fixed Satellite Services (FSS) that make design of very high capacity (up to hundreds Gbit/s) satellites possible. In addition, high-frequency bands offer significant technological advantages over conventional ones: reduced RF equipment size, which is particularly important considering the limited room available on board satellites, higher antenna gains, and narrower beams for a given antenna size.

However, a major limitation of high-frequency bands for Earth-space radio links is the effect of radio wave propagation through the lowest layers of the atmosphere, i.e., the troposphere. Indeed, as the operating frequency is increasing, the attenuation and scintillation effects of atmospheric gas, clouds, rain become more severe, together with some depolarization phenomena. Ionospheric effects can be considered as negligible for these frequency bands. Each of these contributions has its own characteristics as a function of frequency, geographic location, and elevation angle. For non-GSO systems, the variation in elevation angle should also be considered.

Attenuation Effects

Attenuation is the main propagation effect affecting satellite links at high-frequency bands. Attenuation is made up with several contributions due to atmospheric gases (oxygen and water vapor) in clear sky conditions, clouds, and precipitation. At Ku-band, only rain attenuation has to be considered, whereas the other contributions already not negligible at Ka-band can affect strongly system performance at Q/V-band.

Attenuation due to Atmospheric Gases

Attenuation by atmospheric gases caused by absorption depends mainly on frequency, elevation angle, altitude above sea level, and water vapor density (absolute humidity). At frequencies below 10 GHz, it may normally be neglected. Its importance increases with frequency above 10 GHz, especially for low elevation angles. Recommendation ITU-R P.676 gives a complete method for calculating gaseous attenuation.

At a given frequency, the oxygen contribution to atmospheric absorption is relatively constant. However, both water vapor density and its vertical profile are quite variable. Typically, the maximum gaseous attenuation occurs during the season of maximum rainfall.

Cloud Attenuation

Clouds are constituted of suspended water drops which are clearly smaller in size than satcom wavelengths considered in this document. Cloud attenuation depends directly on the amount of liquid water, which is a function of the type of cloud present on the propagation path. Cloud attenuation increases with the integrated or total liquid water content (*ILWC* in kg/m^2) of the cloud and becomes noticeable from the Ka-band. Values of ILWC can be obtained everywhere in the world from maps given in Recommendation ITU-R P.840.

While cloud attenuation can be neglected at Ku-band, it has to be considered at Ka-band especially in tropical or equatorial climates or for low elevation links.

At V-band, cloud attenuation can reach large values especially on the uplink which can degrade strongly the link quality.

Rain Attenuation

Attenuation due to rain is the most important propagation effect to be considered in the design of satellite telecommunications systems because it represents high level of attenuation, and may affect communication systems for significant periods of time.

Spatial characteristics of precipitation influence the total attenuation on a link. Two major types of rain may be identified for temperate regions: stratiform and convective. Two more specific types have to be considered for tropical and equatorial regions and refer to tropical storms (hurricanes, typhoons) and monsoon precipitation. The dimensions of the rainy volume and the rain intensity depend on the type of rain: stratiform precipitation is composed of large areas with low rainfall rates, and convective precipitation consists of small areas (horizontal extent lower than 10 km) with intense rainfall (up to 50 mm/h for 0.01% of an average year in temperate areas).

The rainfall rate (mm/h) plays a detrimental role in the amount of attenuation due to rain and is a key parameter in the various models that have been developed to estimate the amount of attenuation that could be expected on an earth-space link. As rain presents a strong variability in space and time, the rainfall rate is a function of the percentage of time and depends on the climatic area. The rainfall rate conditions the specific attenuation (dB/km) of rain that depends on the temperature, size distribution, and shape of the raindrops. Specific attenuation may vary from fractions of dBs to tens of dBs as frequency increases (Ku- to V-bands) depending strongly on the rainfall (from a few mm/h to hundred mm/h in tropical areas). Global information on the cumulative distribution of rainfall rate is available in Recommendation ITU-R P.837 (Fig. 11.6).

The rain height is therefore another important parameter for calculating the influence of rain, and is assimilated to the height of the $-2°C$ isotherm in order to take into account the influence of the Melting Layer (ITU-R Recommendation P.839).

A lot of propagation models for prediction of rain attenuation have been proposed in the literature for more than 30 years. Empirical models relies on the calculation of the rain specific attenuation (expressed in dB/km) and on the estimation of a reduction factor in order to take into account the fact that rain cells have a limited spatial extension (Recommendation ITU-R P.618).

The rain specific attenuation γ_R (dB/km) is obtained from the rainfall rate R (mm/h) using the power-law relationship:

$$\gamma_R = kR^{\alpha}$$

Values for the coefficients k and α are determined as functions of frequency, f (GHz), in the range from 1 to 1,000 GHz (Recommendation ITU-R P.838).

Estimation of the long-term statistics of the slant-path rain attenuation at a given location for frequencies up to 55 GHz for a given percentage of time can be obtained combining information available in Recommendations P.838, P.839,

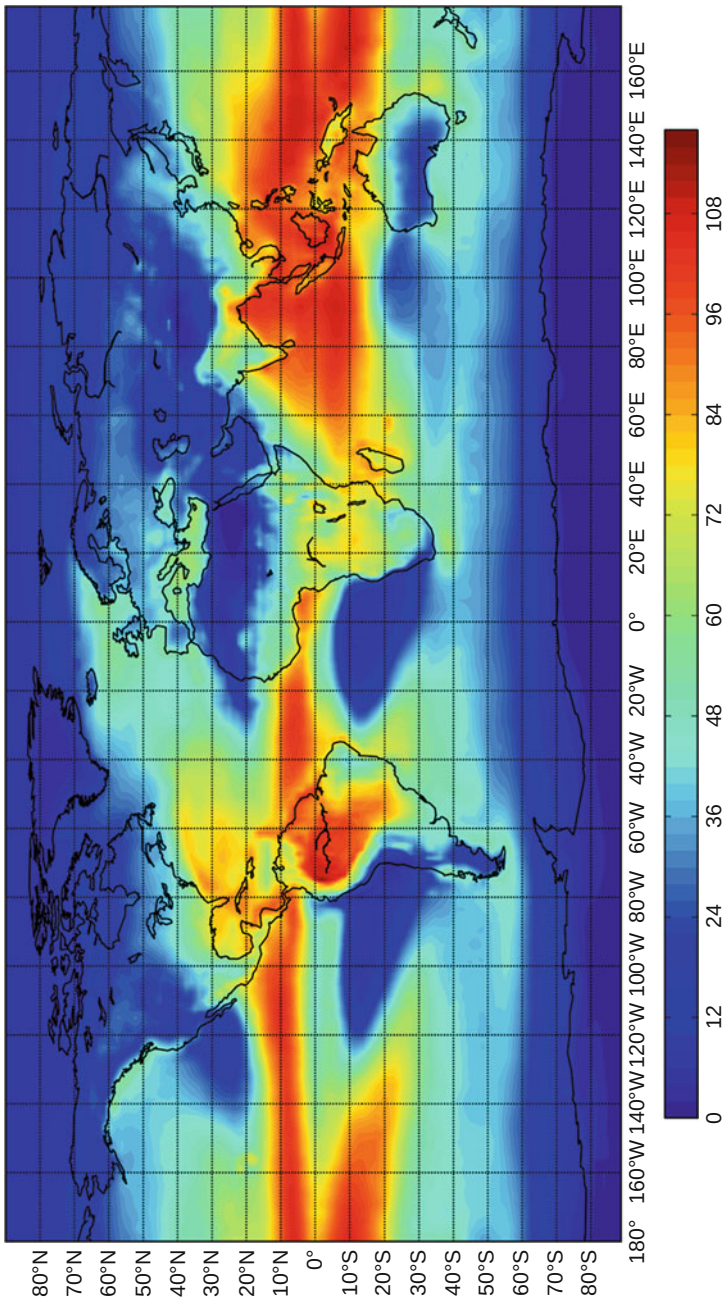


Fig. 11.6 Rain rate (mm/h) exceeded for 0.01% of an average year

and P.618. Information could be synthesized in a step-by-step procedure that allows the calculation of rain attenuation that is overstepped at a given location for a given percentage of time (Maral and Bousquet 2009, ▶ Chap. 4, "Space Telecommunications Services and Applications"). That percentage of time corresponds typically to the accepted unavailability of the link, i.e., one minus the required link availability (the percentage of the time where the link should be able to convey information with a specified quality).

At Ku-band, rain attenuation in a range of a few dBs can be overcome through the use of static margins which allow the required availability (say around 99.95%) to be obtained for the earth-space radio frequency links. Due to technology limitations, a static system margin can no longer be considered as the sole means of compensating propagation disturbances for small percentage of time as the frequency increases to Ka-band and above. Fade mitigation techniques based on adaptive techniques (power control, adaptive coding and modulation, terminal diversity) have to be implemented in order to provide the required availability.

Scintillation Effects

Scintillation corresponds to rapid fluctuations of the received signal characteristics (amplitude, phase, angle of arrival). On satellite links, ionospheric scintillation and tropospheric scintillation can occur, depending on the frequency band.

Tropospheric scintillation is caused by small-scale refractive index inhomogeneities induced by atmospheric turbulence along the propagation path. Above about 10 GHz, tropospheric scintillation will be the unique phenomenon responsible for signal rapid fluctuations. Indeed ionospheric scintillation decreases with frequency and tropospheric scintillation increases with frequency.

Depolarization Effects

Dual polarization transmission is a way to double the system capacity in a given frequency band, which allows the use of the frequency resource to be improved and then to reduce the transmission cost. However, some atmospheric effects limit system performances in terms of isolation between polarizations.

These atmospheric effects regarding wave polarization are different in the ionosphere and in the troposphere. For frequencies above 10 GHz, ionospheric effects (such as Faraday rotation which affects mainly system behavior in linear polarization) can be neglected. Tropospheric effects are caused by the presence of nonspherical hydrometeors (rain drops, ice crystals, snowflakes) with a non-vertical falling direction. It leads to different attenuations and phase shifts on each of the main orthogonal polarization directions, induces a rotation of the polarization plane, and then an increase of the coupling between both orthogonal polarizations.

Recommendation ITU-R P.618 provides a method to calculate long-term statistics of depolarization from rain attenuation statistics.

Conclusion

A survey of the propagation properties of the atmosphere from Ku- to Ka- and V-bands has been presented in this section. The main propagation impairments limiting the performance of satellite communications systems are:
– Attenuation effects due to gas, cloud, rain, and melting layer
– Scintillation in clear sky and inside clouds
– Depolarization due to rain and ice

At Ku-band, the propagation impairments are mainly governed by rain attenuation, which depends on the climatic zone. In temperate climate, typical values of a few dBs are encountered that could be easily mitigated using a static margin (in excess power) in the design of the radio frequency link.

Systems working at Ka-band are submitted to stronger impairments and particularly rain attenuation can reach high levels such as 20 or 30 dB in temperate climate at 30 GHz for percentages lower than 0.1% of an average year, translating in a poor link availability. Fade mitigation techniques have to be implemented into the system in order to improve the link availability and the service performance.

When considering systems operating at V-band, very strong impairments can be obtained, up to several tens of dB at 50 GHz in terms of rain attenuation for low percentages of time. At these frequencies, other effects such as oxygen, water vapor, and cloud attenuation can degrade the system performance even in clear sky conditions (especially for low elevation). The design of such system has to take into account these effects.

Ionospheric Effects on Satellite Navigation Links

The ionosphere is a layer of the upper atmosphere roughly between 50 and 1,500 km above the Earth, which has been ionized by the rays from the sun and, as a result, contains free electrons. Electromagnetic radio waves transmitted from satellites of Global Navigation Satellite Systems (GNSS) such as GPS, GLONASS, COMPASS, and GALILEO interact with the ionospheric plasma on their path to the receiver, changing their speed and direction of travel. The ionosphere has three main effects on satellite to ground signals at the frequencies used for satellite navigation: group delay, scintillations, and Faraday rotation. Since GNSS signals are circularly polarized, Faraday rotation will not be considered here.

It should be noted that troposphere introduces also some delay variation on the radio frequency link. Models are used to mitigate these effects. The availability of GNSS systems results in a rapidly growing number of civilian users who are taking

advantage of the freely available signals in space for a wide range of applications in navigation, surveying, tracking etc., opening a broad market for developers of receivers and applications. Terrestrial overlay systems are being put in place to augment the navigation signals so that they can be used with higher precision or to provide integrity for safety-of-life critical operations such as landing a commercial aircraft (GBAS: Ground-Based Augmentation Systems, or SBAS: Satellite-Based Augmentation Systems).

Basic Concepts

The presence of these free electrons causes a propagation delay indicating a travel distance larger than the real one in the propagation of Global Navigation Satellite Systems (GNSS) signals from satellite to receiver. The amount of delay affecting a particular signal is proportional to the total number of free electrons along the propagation path. At the commonly used L-band frequencies (between 1.2 and 1.6 GHz), the ionosphere causes signal delays that correspond to a range of errors of up to 100 m. As the determination of the terminal position is based on the combination of range estimation to 4 or more satellites, signal delays causing range errors are detrimental to the performance of GNSS systems.

In a first-order approximation, the range error is proportional to the integral of the electron density along the ray path (total electron content – *TEC*). It is called vertical *TEC* when the propagation path has an elevation angle of 90°.

Within the frequency range of GNSS signals, the amount of delay affecting a particular signal is inversely proportional to the square of its frequency ($\sim 1/f^2$). Due to the inertia of the free electrons this propagation error decreases with increasing radio frequency called ionospheric dispersion, and has a measurable impact on radio signals up to frequencies of about 10 GHz.

Irregularities in the distribution of the free electrons along the propagation path can cause rapid fluctuations in the amplitude and phase of received signal, a phenomenon known as scintillation. The dispersive interaction of the navigation signal with the ionospheric plasma impinges space-based radio systems working at L-band frequencies. Scintillation can cause GNSS receivers to temporarily lose lock on one or more of the satellite signals, depending on the number of signals affected by scintillation and on the intensity of the resulting amplitude and phase fluctuations.

Satellite-Based Navigation Technique and Influence of Ionosphere

Satellite navigation relies on precise measurement of the time it takes for a signal to propagate from a spacecraft to the antenna of the navigation receiver. If there is no refractive medium on the path, the location can be determined by calculating the distance to a number of visible satellites by multiplying the time observed by the vacuum speed of light. With the position and clock offsets of the satellites precisely

known, the user of the navigation receiver can easily establish his own position in a given coordinate system.

In reality there are some additional propagation delays, which have to be taken into account. These delays are caused by the interaction of the electromagnetic wave field with the charged particles of the ionospheric plasma, the neutral but polarized particles in the troposphere and with obstacles. The resulting observation equation can be written for the code and the carrier phase ϕ and ϕ, respectively, as:

$$\phi = \rho + c(dt - dT) + d_I + d_T + d_{MP} + dq + dQ + \varepsilon$$

$$\phi = \rho + c(dt - dT) - d_I + d_T + d_{MP} + dq + dQ + N\lambda + \varepsilon$$

Where ρ is the true geometrical range between GPS satellite and receiver, c is the vacuum speed of light, dt and dT are the satellite and receiver clock errors, d_I is the ionospheric delay along the ray path s, d_T is the delay caused by the troposphere, d_{MP} is the multipath error, dq and dQ are the instrumental satellite and receiver biases, λ is the wave length of the carrier, N is the phase ambiguity number (integer), and ε is the residual error.

The space weather-sensitive ionospheric propagation term d_I is a function of the refraction index and related effects such as diffraction and scattering.

In a first-order approximation of the refraction index, the ionospheric delay is proportional to the integral of the electron density along the ray path according to:

$$d_I = \frac{K}{f^2} \int_S^R n_e ds$$

with $K = 40.3\ \mathrm{m^3 s^{-2}}$.

The integral of the local electron density n_e along the ray path between satellite S and receiver R is called total electron content (*TEC*).

Typical vertical *TEC* values range from a few *TEC* units (1 TECU = 10^{16} electrons/$\mathrm{m^2}$) for nighttime conditions at high latitudes to well above 200 TECU at equatorial latitudes and during ionospheric storm conditions.

At the L_1 GPS frequency of 1.575 GHz, 1 TECU is equivalent to a path length increase of 0.162 m indicating maximum ionospheric vertical delays of up to about 30 m under heavily perturbed conditions. The delay may even increase by a factor of more than 2 at low elevation angles, i.e., when the path length through the ionosphere increases considerably.

The first-order ionospheric range error (*IRE*) typically varies from 1 to 100 m at the GNSS frequencies.

Ionospheric Effects as a Function of Latitude

The world can be divided basically into three regions as the ionosphere is concerned.

They are:
– The low latitudes which include the equatorial and equatorial "anomaly" regions
– The midlatitudes
– The high latitudes which include auroral and polar caps

In practice, each of these major ionospheric regions can be further broken down into subregions. During geomagnetically disturbed periods, the auroral regions expand and move equatorward, thus expanding the size of the auroral regions, and shrinking the size of the midlatitude regions. Thus these regions are called transitional regions.

The largest region is the equatorial and so-called equatorial anomaly region, the effects of which can be measured up to at least $\pm 20^\circ$ from the magnetic equator. Note also that most of the continent of South America is located in the equatorial region, as is much of Africa, all of India, and portions of many other countries. Also, ground stations in countries that are located in the lower midlatitudes can view the ionosphere in the equatorial region or into the auroral region when monitoring GNSS satellites at relatively low elevation angle. Thus, the nature of the ionospheric effects affecting a GNSS receiver is not a simple function of the location of that receiver, but of the geographic extent of the regions crossed by the lines of sight to the satellites in view.

Depending on the ionospheric regions, range delay characteristics and scintillation effects display different characteristics. In midlatitude regions, equivalent vertical delays form fairly smooth planar surfaces. Scintillation effects are in general insignificant. Strong phase scintillation occurs only during severe geomagnetic and ionospheric storms under solar maximum conditions. The midlatitude ionosphere normally has the smallest temporal and spatial gradients, has a rare scintillation occurrence, and the 5° by 5° grid of ionospheric corrections that is discussed into the next section has already been shown to work well in this region. The other world regions, however, have larger range delays as well as larger spatial and temporal gradients, a large occurrence of scintillation effects, and, in some cases, a larger day-to-day variability in range delay than the midlatitude ionosphere.

Mitigation Techniques

Single Frequency Receivers

Using an ionospheric model, single-frequency users can reduce the ionospheric range error. Basic GNSS receivers obtain ionospheric corrections from simple models that were derived from analyses of historical data. One such model is the single-frequency GPS ionospheric model. This model is a vertical *TEC* model that uses a mapping function and a thin layer ionosphere assumption to convert vertical to slant *TEC*. It has been shown to statistically correct approximately 50% rms of the actual vertical ionospheric delays. This estimate was done using Faraday rotation *TEC* data from stations at different geographical locations (Klobuchar 1987). The equations for the model are implemented in the receiver, but they operate on a small set of model coefficients broadcast by GPS satellites. These

coefficients are regularly updated by the GPS Ground Control Segment based on observations of the state of the ionosphere during the previous few days. The accuracy of the corrections is limited by the fact that the model is simple, and therefore can only account for first-order variations in the ionosphere, not being able to reproduce basic spatial variations like those found at low latitudes. In particular, the model cannot account for possible irregularities, whether small or large, that may exist at the time the corrections are applied. This model performs best when the ionosphere is a quiet state. It tends to underestimate the magnitudes of delays during ionospheric storms.

The single frequency ionospheric correction algorithm proposed for the Galileo system is based on the 3D NeQuick model (Radicella and Leitinger 2001) driven by an "effective ionization level" Az, valid for the whole world and applicable for a period of typically 24 h. The global Az is given in terms of three coefficients, function of the modified dip latitude or "MODIP" (that depends on the true magnetic dip and the geographic latitude). The Az coefficients are transmitted in the navigation message to the user to allow Az calculation at any wanted location. It must be noted that this model calculates the slant *TEC* by means of integration along the ray path. Therefore, no mapping function or thin layer approximation for the ionosphere is applied.

Residual range errors resulting from the limited accuracy of ionospheric delay corrections with single frequency receivers have to be accounted for when evaluating the accuracy, integrity, availability, and continuity performance of GNSS navigation solutions. The main causes of large-scale variations in vertical *TEC* are related to the 11-year solar cycle, seasonal changes, day-to-day changes, and time of day.

Multiple Frequency Receivers

The frequency dependency characterizes the ionosphere as a dispersive medium as already mentioned above. Since the amount of delay affecting a particular GNSS signal is inversely proportional to the square of the frequency of that signal, accurate calculations of *TEC* (or ionospheric range delay) along the line of sight between a receiver and a satellite are possible using two signals with different frequencies. However, the accuracy of such calculations depends on the calibration technique used and has to take into account effects like multipath.

The differential code phase $\Delta\Phi = \Phi_2 - \Phi_1$ measured at the two frequencies L_1 and L_2 reveal a first-order estimation of *TEC*:

$$TEC = \frac{\Delta\Phi \cdot f_1^2 f_2^2}{K \cdot (f_1^2 - f_2^2)} + TEC_{cal}$$

Besides random errors the calibration term TEC_{cal} includes specific satellite and receiver code phase delays that do not cancel out in $\Delta\Phi$.

The availability of a third or more carrier frequencies improves the performance. In case of a third frequency, the simplest approach is the linear combination of three different frequency pairs for computing ionosphere-free solutions.

Real-Time Ionospheric Measurements

To achieve high-precision positioning, propagation errors have to be mitigated as accurately as possible. Terrestrial overlay systems are used to augment the navigation signals so that they can be used with higher precision or to provide integrity for safety-of-life critical operations such as landing a commercial aircraft (GBAS: Ground-Based Augmentation Systems, or SBAS: Satellite-Based Augmentation Systems).

Indeed errors of several tens of meters with partially corrected ionospheric delays cannot be tolerated for approach operations during which vertical guidance is provided to the aircraft such as Approach with Vertical Guidance (APV) and Precision Approach (PA) operations.

SBAS receivers can correct for ionospheric delays more accurately than basic GNSS receivers because they use information derived from real-time ionospheric delay measurements (see following section). These measurements are collected by a network of reference stations and used by the SBAS ground system to estimate and broadcast vertical delays at the nodes of a standardized ionospheric grid defined on the thin shell. For each line of sight to a satellite, the receiver interpolates among the nearest grid nodes to the location of the ionospheric pierce point (IPP), then converts the interpolated vertical delay to a range (or slant) delay by applying a standardized "obliquity factor" that accounts for the angle at which the line of sight pierces the thin shell. With this type of augmentation, the accuracy of the corrections is limited by:

- The relatively sparse sampling of the ionosphere available to the SBAS ionospheric delay estimation process
- The reduction of the ionosphere (in which free electrons are distributed over a wide range of altitudes) to a two-dimensional model (ionospheric grid) associated with a fixed one-to-one mapping between vertical delays and range (slant) delays
- Time delays between the collection of ionospheric delay measurements by the SBAS ground infrastucture, the broadcast of ionospheric grid information, and the application of the corrections by the SBAS receiver

The monitoring of ionosphere over a region of interest is a key component of SBAS systems. Several techniques have been considered for ionospheric grid estimation. Some techniques proceed on the basis of local models that are separately estimated at each ionospheric grid point (IGP). These techniques rely on low-degree polynomials such as a simple constant, a planar surface, or a quadratic surface to represent variations in vertical ionospheric delays in the local area of each IGP. Other techniques model the ionosphere over the entire service area with a unique model such as high-degree polynomial surfaces or spherical harmonics.

Current SBAS implementations rely on low-degree polynomial surfaces that are separately estimated at each IGP. Indeed, the accuracy of these simple models is good, they are easy to implement (and therefore also verify and certify), and formal error formulas corresponding to a prescribed level of confidence can be easily computed.

However, techniques that work well in midlatitude regions such as the planar fit approach to estimating vertical grid delays, and the simple obliquity factor which converts slant to vertical delays, then back from vertical to slant delays, may not

perform adequately well to ensure a high availability of APV service in the equatorial region. Indeed spatial and temporal changes in the ionosphere that cannot be adequately represented by simple models, and with the possible presence of narrow ionospheric structures that may develop between measurements, the accuracy of the ionospheric delay corrections broadcast to the users will be limited.

Ionospheric scintillation can cause many receivers in the region affected by it to lose lock on one or several satellite signals. Receiver design is the primary source of mitigation against scintillation effects.

GBAS provides pseudorange measurement corrections that include ionospheric corrections. These corrections are derived from real-time measurements and correct for the combined effects of various sources of errors including ionospheric delays. The main limitation to the accuracy of these corrections is the spatial separation between the GBAS ground station and the GBAS receiver.

Ground-Based Ionosphere Monitoring

As outlined in the previous sections, dual-frequency GNSS measurements enable the estimation of the total electron content (*TEC*) along the observed ray paths providing valuable information on the ionospheric state.

To obtain absolute *TEC* values, the ionospheric travel time delay or range error has to be calculated from differential code phases. Since the link-related travel times are biased by the instrumental delays at the satellite and at the receiver, the derived *TEC* data must be calibrated.

To convert the numerous slant *TEC* measurements into the vertical for reference, the ionosphere is assumed to be compressed in a thin single layer at the height h_I. The location of the piercing point of the radio link with the ionospheric layer is considered to be the "point" of measurement. Assuming a single-layer ionospheric shell at a fixed height of about 350–400 km (single layer approximation), the slant *TEC* values are mapped to the vertical at the piercing points by a geometrical mapping function taking into account the elevation angle of the ray path from the satellite, the single layer height, and the radius of the earth (see Arbesser-Rastburg and Jakowski 2007).

Vice versa, this function is also used to compute the slant ionospheric propagation error if grid-based vertical *TEC* information is provided.

For ionospheric imaging, the availability of a sufficient number of measurements is required.

Acknowledgments The works of the European COST 296 Group and of the International SBAS-IONO group (White paper) represent important source contribution to this chapter section. I would also like to acknowledge the material provided by Bertram Arbesser-Rastburg, Head of the Propagation Division at ESA/ESTEC and Chair of the ITU-R Group 3. Similarly, the efforts of the ITU and its members and the quality of their publications, from which a lot of material for the Tropospheric section has been extracted, are fully acknowledged by the author. All ITU References are available on the web at www.itu.int/publications. Finally, I would like to thanks my colleague Prof. Damienne Bajon for her inputs regarding the electromagnetic wave section.

Cross-References

- ▶ History of Satellite Communications
- ▶ Regulatory Process for Communications Satellite Frequency Allocations
- ▶ Satellite Applications Handbook: The Complete Guide to Satellite Communications, Remote Sensing, Navigation, and Meteorology
- ▶ Satellite Communications Overview
- ▶ Satellite Radio Communications Fundamentals and Link Budgets
- ▶ Space Telecommunications Services and Applications

References

B. Arbesser-Rastburg, N. Jakowski, Effects on satellite navigation, chapter 13, in *Space Weather – Physics and Effects*, ed. by V. Bothmer, I. Daglis (Springer, Berlin, 2007)

L. Castanet, *Influence of the Propagation Channel on Satellite Communications – Channel Dynamic Effects on Mobile, Fixed and Optical Multimedia Applications* (Satnex E-book, Shaker, 2007)

COST 255: Radiowave propagation modelling for new satcom services at Ku-band and above, COST 255 final report. ESA Publications Division, SP-1252, March 2002

J.A. Klobuchar, Ionospheric time-delay algorithm for single-frequency GPS users. IEEE Trans. Aerosp. Electron Syst. AES **23**(3), 325–331 (1987)

J.D. Kraus, R.J. Marhefka, *Antennas for All Applications* (McGraw-Hill, New York, 2002)

G. Maral, M. Bousquet, *Satellite Communications Systems*, 5th edn. (Wiley, Chichester, 2009)

J. Maxwell, A dynamical theory of the electromagnetic field. R. Soc. Trans. **CLV**, 459 (1865)

S.M. Radicella, R. Leitinger, The evolution of the DGR approach to model electron density profiles. Adv. Space Res. **27**(1), 35–40 (2001)

Recommendation ITU-R P.618–9:2007, Propagation data and prediction methods required for the design of Earth-space telecommunication systems

Recommendation ITU-R P.676–8:2009, Attenuation by atmospheric gase

Recommendation ITU-R P.836–4:2009, Water vapour: surface density and total columnar content

Recommendation ITU-R P.837–5:2007, Characteristics of precipitation for propagation modelling

Recommendation ITU-R P.838–3:2005, Specific attenuation model for rain for use in prediction method

Recommendation ITU-R P.839–3:2001, Rain height model for prediction method

Recommendation ITU-R P.840–5:2012, Attenuation due to clouds and fog

Recommendation ITU-R P.841–4:2005, Conversion of annual statistics to worst-month statistics

Recommendation ITU-R V.431–7:2000, Nomenclature of the frequency and wavelength bands used in telecommunications

Regulatory Process for Communications Satellite Frequency Allocations

12

Ram S. Jakhu

Contents

Introduction 272
Structure and Functioning of the ITU 273
Purposes of the ITU 273
Legal Instruments Governing the ITU 273
ITU Membership 274
Organizational Structure of the ITU 275
Processes for Obtaining Radio Frequencies and Orbital Slots 279
International Notification, Coordination, and Registration of Radio Frequencies and Orbital Positions 280
Problem of Interference 286
Cross-References 288
Notes 288
References 292

Abstract

Ready access to radio frequencies with limited interference and appropriate orbital positions are indispensable and highly valuable tools for all satellite communications. However, radio frequencies are limited, natural, and international resources. Furthermore, the global demand for radio spectrum has been increasing exponentially. Acting primarily through the International Telecommunication Union (ITU), the international community has developed a very complex regulatory regime that provides detailed rules and processes that govern the international allocation and allotment of radio frequencies and orbital positions. This chapter briefly describes those regulatory processes as well as the manner in which they are created as part of the functioning of the ITU.

R.S. Jakhu
Institute of Air and Space Law, McGill University, 3690 Peel Street, H3A 1W9, Montreal, QC, Canada
e-mail: ram.jakhu@mcgill.ca

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_14, © Springer Science+Business Media New York 2013

Keywords

Coordination and registration processes • Frequency allocations • Frequency allotments • Geostationary or geosynchronous Orbit • International telecommunication union • ITU administrative regulations • ITU constitution • ITU convention • Master international frequency register • National frequency assignments • Plenipotentiary conferences • Radio communications bureau • Radio frequencies • Radio regulations board, radio stations • Satellite communications • Spectrum • Table of frequency allocations • World radio communications conference

Introduction

Radio frequencies and orbital positions are indispensable tools for satellite communications; i.e., in order to function properly all satellites need appropriate orbits (paths in outer space) and interference-free radio frequencies. However, if they are not properly used, harmful interference could occur which, in turn, might reduce the quality of communications.

The starting point for a satellite operator is always to meet the mandatory requirements for a national radio license under national regulatory procedures. National assignment[1] of radio frequencies and orbital positions are carried out through applicable national licensing systems and procedures, which vary from country to country. However, once radio frequencies and orbital positions are secured by a satellite operator under a national license, the initiation of, and follow-up for international coordination and registration processes to ensure interference-free use of radio frequencies and orbital positions through the International Telecommunication Union (ITU) are undertaken on behalf of that operator by the respective State (i.e., the national communications regulatory Administration). It is interesting to note that in 2009, the ProtoStar communication satellite system (worth about $500 million), which was established to provide direct-to-home television and broadband internet services in Asia using Ku-band and C-band radio frequencies, had to declare bankruptcy because it could not find any State (Administration) to properly register with ITU the radio frequencies and orbital positions it intended to use.[2]

The general perception is that the need for telecommunications (especially those involving the use of satellites and radio frequencies) will expand rapidly. However, the level of that expansion will be greatly determined by the availability of radio frequencies and orbital positions. All satellite operators will try to secure as many radio frequencies and orbital positions as possible in order to maintain and expand their market share. Accordingly to Adrian Ballintine, chief executive of NewSat a company that recently acquired seven geostationary orbital slots for its new satellites, "[t]he slots are extremely valuable assets with senior filing status, outstanding geographic footprint and certainly enough capacity to see NewSat's long-term future assured. We are now in a position to launch multiple satellites, each of which could generate in excess of $100 million of (earnings) per year."[3]

However, radio frequencies and orbital positions are limited natural resources of an international character that must be shared among several radio services and all countries. Therefore, a very complex and extensive international regulatory regime has been established through an intergovernmental organization; i.e., the ITU. New challenges in accessing appropriate radio frequencies arise due to increasing privatization, competition, and globalization and even abuse of the ITU regulatory processes. One can expect more rigid and extensive international regulations and procedures in the not too distant future.

This chapter discusses (1) the structure and functioning of the ITU in order to explain how rules and procedures are adopted and (2) the processes for obtaining radio frequencies and orbital slots through coordination within, and registration with, the ITU.

Structure and Functioning of the ITU

Purposes of the ITU

The purposes of the ITU, *inter alia*, are: (a) to maintain and extend international cooperation between all Members for the improvement and rational use of telecommunications of all kinds; (b) to promote and to offer technical assistance to developing countries in the field of telecommunications; (c) to promote the development of technical facilities and their most efficient operation with a view to improving the efficiency of telecommunication services; and (d) to promote, at the international level, the adoption of a broader approach to the issues of telecommunications in the global information economy and society, by cooperating with other world and regional intergovernmental organizations and nongovernmental organizations concerned with telecommunications.[4] Pursuant to these purposes, in the field of satellite telecommunications, the ITU (a) effects the allocation[5] of bands of the radio frequency spectrum, the allotment[6] of radio frequencies and registration of radio frequency assignments and any associated orbital positions in the geostationary-satellite orbit in order to avoid harmful interference between radio stations; (b) coordinates efforts to eliminate harmful interference between radio stations and to improve the use made of the radio frequency spectrum and of the geostationary-satellite orbit; (c) facilitates the worldwide standardization of telecommunications for a satisfactory quality of service; and (d) fosters international cooperation in the delivery of technical assistance to developing countries and the creation, development, and improvement of telecommunication equipment and networks in developing countries.[7]

Legal Instruments Governing the ITU

The ITU functions under three basic international legal instruments (i.e., treaties). They are: (a) the Constitution of the International Telecommunication Union

(hereinafter referred to as the ITU Constitution); (b) the Convention of the International Telecommunication Union (hereinafter referred to as the ITU Convention); and (c) the Administrative Regulations.[8] The Administrative Regulations, which regulate the use of telecommunications and are binding on all Members of ITU,[9] are: (a) the International Telecommunication Regulations and (b) the Radio Regulations.[10] In the event of an inconsistency between a provision of the Constitution and a provision of the Convention or of the Administrative Regulations, the Constitution prevails.[11] In the case of inconsistency between a provision of the Convention and a provision of the Administrative Regulations, the Convention prevails.[12] Ratification, acceptance, approval, or accession of the Constitution and the Convention, also constitute consent to be bound by the Administrative Regulations adopted prior to the date of signature of the Constitution and the Convention.[13] A Member State must notify its consent to be bound by a partial or complete revision of the Administrative Regulations to the ITU Secretary-General.

The Members of ITU are bound to abide by the provisions of the Constitution, the Convention, and the Administrative Regulations in all telecommunication offices and stations established or operated by them which engage in international services or which are capable of causing harmful interference to radio services of other Administrations (countries).[14] The Members are also bound to take the necessary steps to ensure the observance of the provisions of the Constitution, the Convention, and the Administrative Regulations by operating agencies (including private companies) authorized by them to establish and operate telecommunications and which engage in international services or operate stations capable of causing harmful interference to the radio services of other countries.[15] No communication transmitting station ought to be established or operated by a private person without a license issued in conformity with the provisions of the Radio Regulations by the government of the Administration to which the station in question is subject.[16] Therefore, Member States must require their operating agencies to obtain radio licenses from the appropriate national regulatory authorities (e.g., the Federal Communications Commission in the United States). However, under Article 48 of the ITU Constitution, Member States are exempted from the application of ITU Radio Regulations with respect to their military radio installations, although these installations must observe regulatory provisions relative to giving assistance in case of distress and to the measures to be taken to prevent harmful interference. In addition, when these installations are used for the provision of public correspondence service or other services governed by the Radio Regulations, they must, in general, comply with the regulatory provisions for the conduct of such services.

ITU Membership

Membership in the ITU, the oldest specialized agency of the United Nations (UN), is based on the principle of universality; i.e., the rule is: once a member, always a member. Thus, membership cannot be canceled or suspended. However,

a Member may be barred from attending a particular meeting or all meetings or conferences of the ITU. There are two types of members of ITU, i.e., State Members and Sector Members-Associates. As of February 2011, there were 193 Member States and over 700 Sector Members of the ITU.[17]

State membership: A State that is a Member of the UN becomes State Member of the ITU by acceding to its Constitution and Convention.[18] If a State is not a Member of the UN, its application for membership needs to be approved by two-thirds of the Member States of the ITU.[19] An "Instrument of accession" covering both the Constitution and the Convention must be deposited with the Secretary-General of the ITU in order for a State to become a Member State of the ITU.

Sector membership: Global communication industry representatives are allowed to join the ITU as Sector Members. This supports the purposes and activities of the Sector of which they become Members and contributes to defraying the expenses of the Sector concerned. The entities that are eligible to become Sector Members include: (a) recognized operating agencies, scientific or industrial organizations, and financial or development institutions which are approved by the Administration of the Member State concerned; (b) other entities dealing with telecommunication matters which are approved by the Administration of the Member State concerned; and (c) regional and other international telecommunication, standardization, and financial or development organizations.[20] With the prior approval of the Administration (Government) of the Member State in which it has its headquarters, an entity may submit an application to the Secretary-General of ITU, on the receipt of which that entity becomes a Sector Member.[21] Sector Members are entitled to participate (a) in the work of Study Groups and subordinate groups and (b) in the process of preparing recommendations and comments before the adoption of recommendations, if any. However, they are not allowed to be involved in voting for, or approval of recommendations.[22] As from October 2010, academic institutions, universities, and research establishments that are concerned with the development of communications are allowed to be admitted to participate without voting rights in the work of the three Sectors of ITU for a trial period of 4 years.[23]

Organizational Structure of the ITU

Since its birth in 1865,[24] the ITU has evolved to become a very complex international organization with a multifaceted organizational structure, which currently includes: (a) the Plenipotentiary Conference; (b) the Council; (c) World Conferences on International Telecommunications; (d) the Telecommunication Standardization Sector; (e) the Telecommunication Development Sector; (f) the Radiocommunication Sector; and (g) the General Secretariat.[26] See Fig. 12.1.

The plenipotentiary conference, the supreme organ of the Union, is composed of delegations representing all Member States.[27] It is convened every 4 years. Exceptionally, Extraordinary Plenipotentiary Conference may be convened with a restricted agenda to deal with specific matters by a decision of the preceding

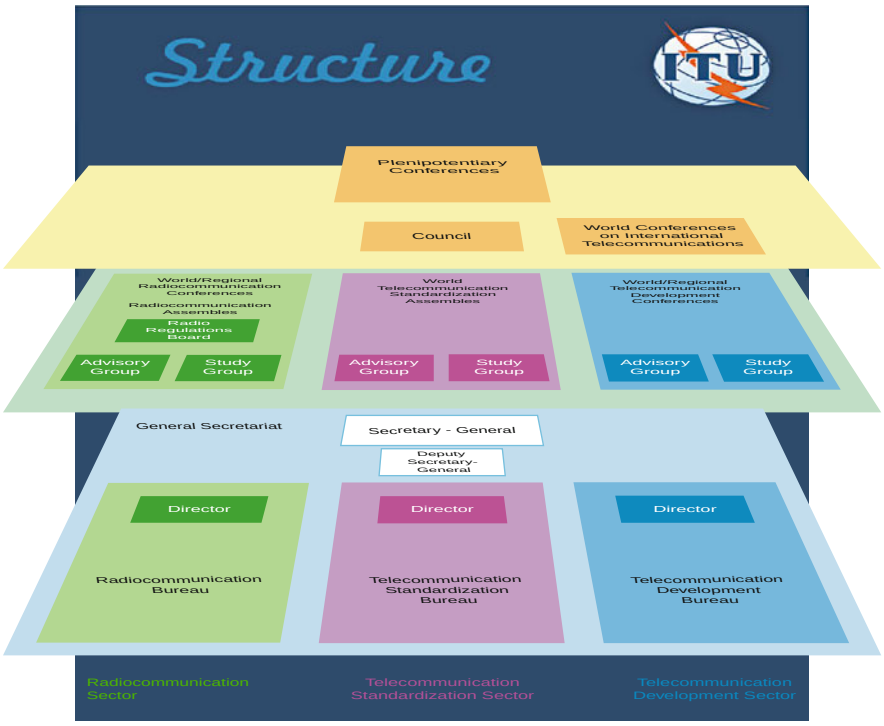


Fig. 12.1 Structure of international telecommunication union[25] (Reproduced with the kind permission of ITU)

ordinary Plenipotentiary Conference or if two-thirds of the Members of the Union so request.[28] The Plenipotentiary Conference, *inter alia*, determines the general policies of the Union; elects the Members of the ITU Council, the Secretary-General and the Deputy Secretary-General, the Directors of the Bureaux of the three ITU Sectors, and the members of the Radio Regulations Board; considers and adopts, if appropriate, proposals for amendments to the Constitution and the Convention; and concludes or revises, if necessary, agreements between the ITU and other international organizations.[29]

The council is the governing body of the ITU during the intervening periods between the Plenipotentiary Conferences. The Council considers "broad telecommunication policy issues in accordance with the guidelines given by the Plenipotentiary Conference to ensure that the Union's policies and strategy fully respond to changes in the telecommunication environment."[30] It drafts the agenda for administrative conferences. Currently, there are 48 Member States represented on the Council and they were elected during the 2010 Plenipotentiary Conference.[31]

The telecommunication standardization sector (*ITU-T*) studies technical, operating, and tariff questions and adopts recommendations on them with a view

to standardizing telecommunications on a worldwide basis.[32] World telecommunication standardization conferences are convened every 4 years; however, an additional conference may be held in accordance with the relevant provisions of the Convention.[33]

The telecommunication development sector (*ITU-D*) facilitates and enhances telecommunications development by offering, organizing and coordinating technical cooperation and assistance activities. It is required to (a) raise the level of awareness of decision-makers concerning the important role of telecommunications in national economic and social development programs; (b) encourage participation by industry in telecommunication development in developing countries; and (c) offer advice on the choice and transfer of appropriate technology.[34]

The radiocommunication sector (*ITU-R*) is the main unit within the ITU that adopts the detailed rules, procedures, and standards for ensuring interference-free use of radio frequencies and orbital positions and their routine implementation. For this purpose, the Sector functions under a very complex organizational structure and by following intricate decision-making processes. See Fig. 12.2.

The primary function of the Radiocommunication Sector is to ensure the rational, equitable, efficient, and economical use of the radio frequency spectrum by all radiocommunication services, including those using the geostationary-satellite orbit.[36] It works through: (a) world and regional radiocommunication conferences; (b) the Radio Regulations Board; (c) radiocommunication assemblies, which are associated with world radiocommunication conferences; (d) radiocommunication study groups; and (e) the Radiocommunication Bureau, headed by an elected Director.[37] The members of the Radiocommunication Sector are: (a) of right, the Administrations of all Member States of the Union and (b) entities or organizations that become Sector Members in accordance with the relevant provisions of the Convention.[38]

World radiocommunication conferences (*WRCs*) are exclusively entitled to revise the ITU Radio Regulations, to deal with any question related to matters of worldwide character involving allocation and allotment of radio frequencies and orbital positions,[39] and to instruct the Radio Regulations Board and Radiocommunications Bureau and review their activities. They are normally convened every 3 - 4 years and are the main bodies of the ITU which establish international regulatory processes related to satellite communications.

The radio regulations board (*RRB*) is constituted by 12 part-time individuals elected by the Plenipotentiary Conference.[40] The members of the RRB are required to serve as custodians of an international public trust and not as representatives of their respective Member States countries or regions.[41] They must refrain from intervening in decisions directly concerning their own respective Administrations and must not request or receive instructions relating to the exercise of their duties from any government or any public or private organization or person. Similarly, Member States and Sector Members are obligated to respect the exclusively international character of the duties of the members of the RRB.[42] The Board meets four times a year, in stark contrast to the more permanent nature of its predecessor, the International Frequency Registration Board (IFRB). Fewer

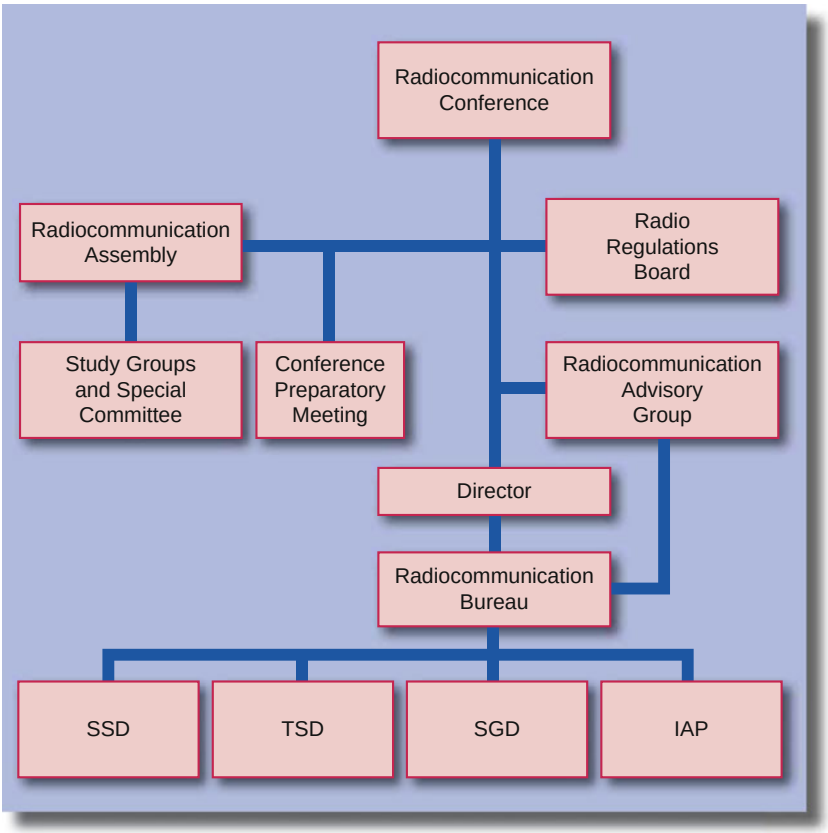


Fig. 12.2 Structure of radiocommunications sector[35] (Reproduced with the kind permission of ITU)

functions and powers have been entrusted to the RRB as compared to those of the former IFRB.

The RRB develops Rules or Procedures, which are submitted to the upcoming World Radiocommunication Conference for the necessary modifications to the ITU Radio Regulations. The RRB's Rules of Procedure are used by the Radiocommunications Bureau for the purpose of registering radio frequencies and orbital positions in the Master International Frequency Register (Master Register).[43]

In brief, it should be noted that the organizational structure of and the procedures followed by the ITU Radiocommunication Sector are complex and intricate. See Fig. 12.3.[44]

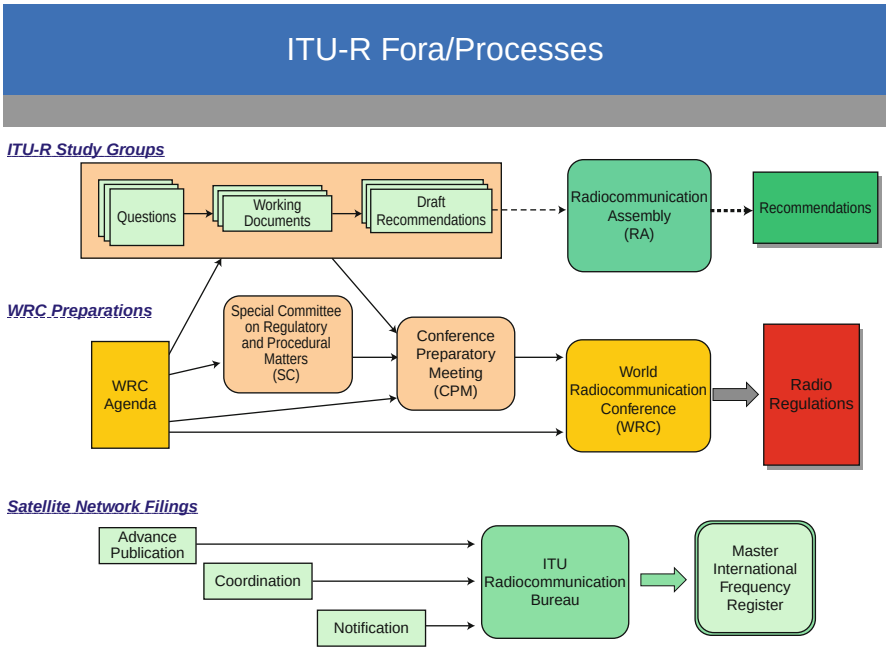


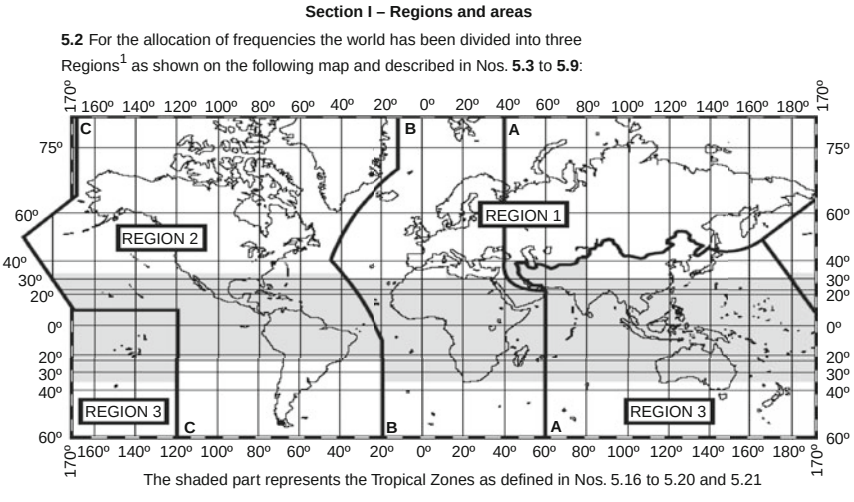
Fig. 12.3 ITU radiocommunications sector and processes (Reproduced with the kind permission of Boeing)

Processes for Obtaining Radio Frequencies and Orbital Slots

The main goals of the ITU's international regulatory regime governing satellite communications are to avoid harmful interference and to ensure equitable access to radio frequencies and satellite orbital slots. According to Article 45 of the ITU Constitution:

- All radio stations must be established and operated in such a manner as not to cause harmful interference to the radio services of other Members which are operated in accordance with the provisions of the ITU Radio Regulations.
- The Members are required to take all practicable steps to prevent the operation of electrical apparatus and installations of all kinds from causing harmful interference to the radio services.
- Each Member undertakes to require their recognized operating agencies to respect the above-mentioned obligations.

The avoidance of harmful interference is achieved through processes of allocation and registration of radio frequencies and orbital positions with the ITU. In this context, it is important to keep in mind the earlier-mentioned specific definition of



[1] **5.2.1** It should be noted that where the words "regions" or regional" are without a capital "R" in these Regulations, they do not relate to the three Regions here defined for purposes of frequency allocation.

Fig. 12.4 Regions and areas of the world for purposes of allocation of radio frequencies[45] (Reproduced with the kind permission of ITU)

"allocation" as included in the ITU Radio Regulations. For the purpose of allocation, the ITU has divided the world into three Regions. See Fig. 12.4.

As determined periodically, World Radiocommunication Conferences add, delete and modify all allocations in a complex "Table of Frequency Allocations" under Article 5 of the ITU Radio Regulations (Table 12.1). This Article provides the categories of all radio services as well as their respective rights or priorities to use.[46] Table 12.1 (i.e., one page extracted from the Table of Frequency Allocations in Article 5 of Radio Regulations) provides an example of allocations within the 2,520–2,700 MHz band to various services within the three ITU Regions. In order to fully understand the precise nature and scope of specific allocations, one needs to keep in mind the definitions, categories of services and their operational priorities, and hundreds of footnotes mentioned under almost each and every allocation. These footnotes serve as explanations and exceptions to the general rule of allocations made in each category and Region.

International Notification, Coordination, and Registration of Radio Frequencies and Orbital Positions

According to Article 8 of the ITU Radio Regulations, international rights and obligations in respect of radio frequency assignments and orbital positions are

Table 12.1 Table of frequency allocations (Reproduced with the kind permission of ITU)

2,520–2,700 MHz		
Allocation to services		
Region 1	Region 2	Region 3
2,520–2,655 FIXED MOD 5.410 MOBILE except aeronautical mobile 5.384A BROADCASTING-SATELLITE 5.413 MOD 5.416	**2,520–2,655** FIXED MOD 5.410 FIXED-SATELLITE (space-to-Earth) 5.415 MOBILE except aeronautical mobile 5.384A BROADCASTING-SATELLITE 5.413 MOD 5.416	**2,520–2,535** FIXED MOD 5.410 FIXED-SATELLITE (space-to-Earth) 5.415 MOBILE except aeronautical mobile 5.384A BROADCASTING-SATELLITE 5.413 MOD 5.416 5.403 5.415A ADD 5.4A01
		2,535–2,655 FIXED MOD 5.410 MOBILE except aeronautical mobile 5.384A BROADCASTING-SATELLITE 5.413 MOD 5.416
5.339 5.405 5.412 5.417C 5.417D 5.418B 5.4I8C	5.339 5.417C 5.417D 5.418B 5.418C	5.339 5.417A 5.417B 5.417C 5.417D MOD 5.418 5.418A 5.418B 5.418C
2,655–2,670 FIXED MOD 5.410 MOBILE except aeronautical mobile 5.384A BROADCASTING-SATELLITE 5.347A 5.413 MOD 5.416 Earth exploration-satellite (passive) Radio astronomy Space research (passive)	**2,655–2,670** FIXED MOD 5.410 FIXED-SATELLITE (Earth-to-space) (space-to-Earth) 5.347A 5.415 MOBILE except aeronautical mobile 5.384A BROADCASTING-SATELLITE 5.347A 5.413 MOD 5.416 Earth exploration-satellite (passive) Radio astronomy Space research (passive)	**2,655–2,670** FLXED MOD 5.410 FIXED-SATELLITE (Earth-to-space) 5.415 MOBILE except aeronautical mobile 5.384A BROADCASTING-SATELLITE 5.347A 5.413 MOD 5.416 Earth exploration-satellite (passive) Radio astronomy Space research (passive)
5.149 5.412	5.149	5.149 5.420

derived from the recording of those assignments and positions in the Master International Frequency Register (Master Register or MIFR) in accordance with the provisions of ITU Radio Regulations. Only those radio frequency assignments and associated orbital positions that have been properly recorded in the Master Register are entitled to the right to international

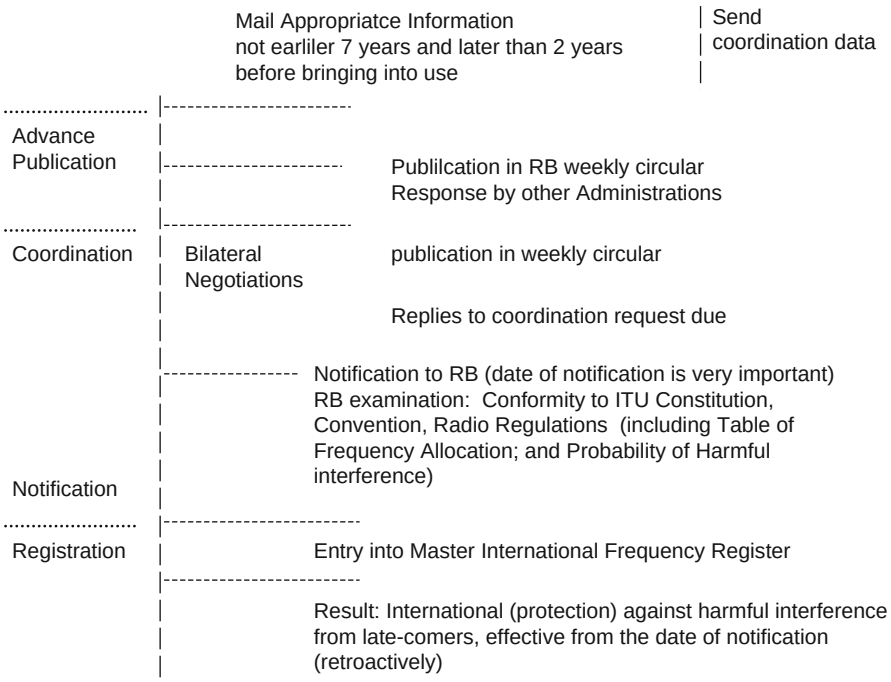


Fig. 12.5 First-come, first-served procedure (applicable to all space services in all bands except those for which a priori plans are adopted)

recognition; i.e., protection against harmful interference. If harmful interference to any station, whose assignment has been properly registered, is actually caused by the use of a frequency assignment which is not in conformity with the ITU Radio Regulations, the radio station using the latter frequency assignment must, upon receipt of advice thereof, immediately eliminate this harmful interference.[47]

There are two methods for international notification, coordination, and registration of radio frequencies and orbital positions; i.e., the so-called first-come, first-served method and the a priori planning.

First Come: First Served

In order to be entitled to a right of international protection against harmful interference the assignments must be properly coordinated within ITU and registered with the ITU, *inter alia*, if (a) international protection against harmful interference is desired, (b) the assignment will be used for international service, or (c) it is believed that the use of a new assignment will cause harmful interference.[48]

Ensuring Equitable Access: A Priori Planning

In order to ensure equitable access to radio frequencies and orbital positions, Article 44. 2 of the ITU Constitution specifies that "In using frequency bands for radio services, Member States shall bear in mind that radio frequencies and any associated orbits, including the geostationary-satellite orbit, are limited natural resources and that they must be used rationally, efficiently and economically, in conformity with the provisions of the Radio Regulations, so that countries or groups of countries may have equitable access to those orbits and frequencies, taking into account the special needs of the developing countries and the geographical situation of particular countries." This Article introduced the concept of equity or equitable access with respect to the use or the sharing of the radio frequencies and orbital positions. Giving effect to this concept, the ITU has a priori allotted (distributed) radio frequencies and associated orbital positions among its Member States at World Radiocommunications Conferences. However, so far, it has only been implemented in connection with a limited number of allotment plans, e.g., (a) provision and associated frequency allotment plan for the aeronautical mobile (OR) service in the bands allocated exclusively to that service[49]; (b) frequency allotment plan for the aeronautical mobile (R) service and related information[50]; (c) broadcasting-satellite service (BSS) operating in 12 GHz band and associated feeder links[51]; and (d) fixed-satellite service (FSS) operating in 6/4 GHz and 14/11 GHz bands,[52] etc. The rarity of such plans is attributable to the desire of ITU State Members to retain their freedom in the use of radio frequencies and orbital positions (Fig. 12.6).

Decision-Making Process

For the purpose of registration, the notifying Administration is required to send to the ITU the required information that is published by the organization. This process is called "advance publication." An Administration which considers that the newly notified and published satellite system might interfere with its already registered radio frequencies is entitled to object to the registration of the later satellite system. In such cases, the notifying Administration may ask for coordination with the objecting Administration. The purpose of this advance publication is to give other Administrations the opportunity to look at the information and to communicate with the publishing Administration within a period of 4 months if there is a risk of interference.[53]

Applying Rules of Procedures, the RB examines the notifications with respect to their conformity to ITU Constitution, Convention, and the Radio Regulations (including Table of Frequency allocations, allotment plans, and probability of harmful interference). In this regard, if the findings of the RB are positive, it registers the notified assignments in the Master Register.

According to Article 14 of the ITU Radio Regulations, any Administration may request to the RB for a review of its finding(s) or any other decision(s) of the RB. If the outcome of the review does not successfully resolve the matter, or if it would adversely affect the interests of other Administrations, the RB is required to prepare a report and send it to the Administration which requested the review

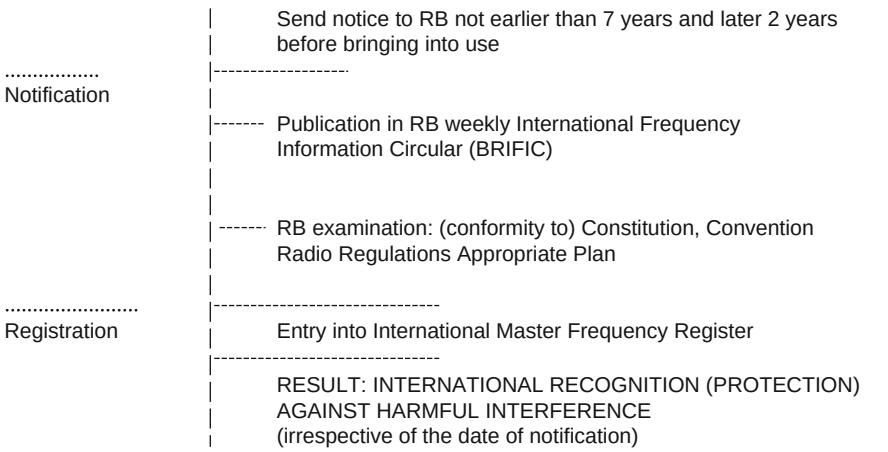


Fig. 12.6 An a priori plan procedure (applicable to BSS in 12 GHz band and FSS in 6/4 GHz and 14/11 GHz bands)

and to any others concerned in order to enable them, if they so desire, to address the RRB. The RB then sends the report with all supporting documentation to the RRB. The decision of the RRB on the review, to be taken in accordance with the Convention,[54] is regarded as final in so far as the RB and the RRB are concerned. However, if the Administration which requested the review disagrees with the RRB's decision, it may raise the matter at a World Radiocommunication Conference.

In order to reduce the abuse of regulatory processes,[55] the ITU has adopted several mechanisms, the first one of which fixes a period of 7 years for bringing into use a communication satellite system from the date of submission of the advance publication information.[56] "If, after the expiry of the period of 7 years from the date of receipt of the relevant complete information. . ., the administration responsible for the satellite network has not brought the frequency assignments to stations of the network into use, the corresponding (published) information. . ., shall be cancelled, but only after the administration concerned has been informed at least 3 months before the expiry date."[57] However, the ITU Radio Regulations do not place any time limitations upon the right of States to continue occupying radio frequencies and orbital slots after they have started using them. When the notifying Administration suspends the use of a recorded assignment to a communication satellite system for a period not exceeding 18 months, it is obliged to inform the RB about the date on which the suspended assignment will be brought back into regular use.[58] The RB is also entitled to review periodically the Master Register with the aim of maintaining or improving its accuracy.[59] Using this power, the RB can reduce (if not eliminate) the hoarding of recorded radio frequency assignments, which are not

actually used. (See ITU Circular Letter CR/301, "Removal of unused frequency assignments (Space Services) from the Master Register", 1 May 2009).

Second: Each notifying Administration is required to provide evidence of seriousness of its intention of establishing a satellite network. ITU Radio Regulations Resolution 49,[60] Annex 2, requires that the following information be provided to the ITU at the time of submission of information of radio frequency assignments:

- Identity of the satellite network: (a) Identity of the satellite network; (b) Name of the Administration; (c) Country symbol; (d) Reference to the advance publication information. . .; (e) Reference to the request for coordination; (f) Frequency band(s); (g) Name of the operator; (h) Name of the satellite; and (i) Orbital Characteristics
- Spacecraft Manufacturer: (a) Name of the spacecraft manufacturer; (b) Date of execution of the contract; (c) Contractual "delivery window" (planned period, beginning and end dates); and (d) Number of satellites procured
- Launch Services provider: (a) Name of the launch vehicle provider; (b) Date of execution of the contract; (c) Anticipated launch or in-orbit delivery window; (d) Name of the launch facility; and (e) Name and location of the launch facility

At the 2007 World Radiocommunication Conference, the filing with the ITU by the notifying Administration of the required documents showing seriousness of its planned satellite system was made mandatory. If this is not completely done in a timely fashion, the satellite notified to ITU for registration "shall no longer be taken into account and shall not be recorded in the MIFR" (Master International Frequency Register).[61]

Third: In order for the ITU to recover its processing costs for the communication satellite networks, each notifying Administration is required to pay "filing fee." This measure is a market mechanism in line with the "user-pay" principle; so that ITU is in a position to recover administrative expenses from the users of the radio frequencies and orbital slots. The 1998 ITU Plenipotentiary Conference held in Minneapolis, adopted a decision to charge for all satellite filings received by ITU after 7 November 1998. This decision was implemented by the Council in its Decision 482 (adopted at its 2002 session), which has been further modified several times. The ITU has started charging a nonrefundable filing fee per satellite network.

The revised Article 9.2B1 of the ITU Radio Regulations now provides that "If the payments are not received (by the ITU) in accordance with the provisions of Council Decision 482, as amended, on the implementation of cost recovery for satellite network filings, the (Radiocommunication) Bureau shall cancel the publication, after informing the Administration concerned. The Bureau shall inform all Administrations of such action, and that the network specified in the publication in question no longer has to be taken into consideration by the Bureau and other Administrations." Almost every year, a number of communication satellites filings are canceled as a result of nonpayment of ITU processing fee invoices.[62] It should be noted that only those satellites that have been properly registered with ITU are entitled to protection against harmful interference.

Problem of Interference

Irrespective of the efforts made by Administrations and the ITU for the avoidance of harmful interference, problems of interference often arise. Article 1.166 of the ITU Radio Regulations defines "harmful interference" as "the effect of unwanted energy due to one or a combination of emissions, radiations, or inductions upon reception in a radiocommunication system, manifested by any performance degradation, misinterpretation, or loss of information which could be extracted in the absence of such unwanted energy."

There exists no compulsory international dispute settlement machinery within the ITU with respect to the resolution of interference problems. Under the ITU agreements, dispute resolution is dealt with under Article 56 of the ITU Constitution. According to this Article, Member States may settle their disputes on questions relating to the interpretation or application of the ITU Constitution, of the Convention or of the Administrative Regulations (including Radio Regulations governing space communications) by negotiation, through diplomatic channels, or according to procedures established by bilateral or multilateral treaties concluded between them for the settlement of international disputes, or by any other method mutually agreed upon. If none of these methods of settlement is adopted, any Member State party to a dispute may have recourse to arbitration in accordance with the arbitration procedure as specified in Article 41 of the ITU Convention. Member States have also concluded an Optional Protocol on the Compulsory Settlement of Disputes Relating to the ITU regulatory regime[63] which is applicable among Member States parties to that Protocol. This Protocol essentially makes the arbitration procedure as defined in Article 41 of the ITU Convention compulsory for the settlement of disputes among the States Parties to the Protocol, numbering 64 at present. In practice, neither Article 56 of the ITU Constitution, nor Article 41 of the ITU Convention or the Optional Protocol has ever been used. Therefore, all the harmful interference problems have been, and are, resolved according to the provisions of Article 15 of the ITU Radio Regulations.

Under Article 15 of Radio Regulations, cases of harmful interference are resolved exclusively through bilateral negotiations between the concerned Administrations who are obliged to exercise the utmost goodwill and mutual assistance in the application of the provisions of the ITU Constitution and Radio Regulations to the settlement of problems of harmful interference. In this regard, if considered appropriate, the concerned States may seek the administrative support of the RB. It is the RB which acts as an executive arm of the RRB conducting investigations into harmful interference allegations and registering frequency assignments. The RRB provides recommendations to the concerned Administrations in cases of harmful interference after a report has been received from the Director of the RB.[64]

Frequency jamming amounts to intentionally caused interference which is illegal. Article 45 of ITU Constitution specifies that "All stations, whatever their purpose, must be established and operated in such a manner as not to cause harmful interference to the radio services or communications of other Members... which carry on a radio service, and which operate in accordance with the provisions of the Radio

Regulations." In addition, according to Article 15 of ITU Radio Regulations, "(1) All stations are forbidden to carry out unnecessary transmissions, or the transmission of superfluous signals. . .. (2) Transmitting stations shall radiate only as much power as is necessary to ensure a satisfactory service." Jamming could also have serious consequences for national and international security because it is considered to be an unfriendly act or a sort of war action. The incidence of possible interference with, or jamming of, radio signals of satellites has been well-known in the operation of both civilian and military satellites, including global navigation satellite systems, like the Russian GLONASS and American Global Positioning System (GPS) (Butsch 2011). For instance, in 2009 and 2010, Iran intentionally jammed television and radio signals transmitted by the EUTELSAT communication satellite system to Europe and the Middle East (Brown 2010).

The ITU does not possess any mechanism or power of enforcement or imposition of sanctions against the violators of its rules, regulations, and processes. It is true that the voluntary compliance approach worked has well in the past and that States have largely been following the ITU rules mainly due to the fact that noncompliance would not augur well for their own individual and collective self-interest. It is doubtful whether this tradition will work well in the future as the number of State and non-State players has been increasing and the competition for scarce resources is becoming severe.

In order to diagnose the problem of harmful interference and its elimination or reduction, it is imperative to have an appropriate monitoring system. Currently, there exists no independent international monitoring system. Under Article 16 of the ITU Radio Regulations, Administrations have agreed to continue the development of their monitoring facilities. These stations may be operated by an Administration or, in accordance with an authorization granted by the appropriate Administration, by a public or private enterprise, by a common monitoring service established by two or more countries, or by an international organization. However, the so-called international monitoring system *currently* comprises *only* those national monitoring stations which have been so nominated by Administrations in the information sent to the ITU Secretary-General. Lack of independent international monitoring system(s) inhibits the availability of objective information which is critical in unbiased settlement of disputes related to harmful interference.

Conclusion

More and more telecommunication satellites are being, and will be, established worldwide. This means that there is, and will be, increased pressure on the already scarce and seriously congested radio frequency spectrum without which no radio telecommunications system can be operated. The competition for very limited radio frequencies and also orbital positions will only grow fierce as the demand for telecommunications increases. Costs for accessing and using these resources to operators and regulators are expected to amplify as access becomes difficult and cases of harmful interference increase.

Good faith has been the main basis for the implementation of the ITU procedures and regulations (there are no sanctions against violations). In recent

years, due to rapidly increased demands and competition among applicants, ITU's coordination and registration procedures have been abused. Thus, the ITU registration processing system is seriously clogged and it takes a few years for an application to get processed. The problem of orbital congestion is particularly significant in the geostationary orbit. The above-discussed three steps taken by the ITU in order to reduce the registration of "paper or virtual satellites" may be small but are important developments for enhancing the effectiveness of the ITU regulations and processes. It is in the interest of all communication satellite operators and States to comply fully with the ITU regulatory regime and processes.

Cross-References

- ▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ▶ Economics and Financing of Communications Satellites
- ▶ Fixed Satellite Communications: Market Dynamics and Trends
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends
- ▶ Satellite Communications and Space Telecommunications Frequencies
- ▶ Satellite Communications Antenna Concepts and Engineering
- ▶ Satellite Communications Overview
- ▶ Satellite Communications Regulatory, Legal and Trade Issues
- ▶ Satellite Orbits for Communications Satellites
- ▶ Satellite Radio Communications Fundamentals and Link Budgets
- ▶ Space Telecommunications Services and Applications
- ▶ Trends and Future of Satellite Communications

Notes

1. ITU Radio Regulations, Article 1.18, (as amended by World Radiocommunication Conference in 2007; hereinafter referred to as the ITU Radio Regulations) define "assignment" (of a radio frequency or radio frequency channel) as an "Authorization given by an Administration (State) for a radio station to use a radio frequency or radio frequency channel under specified conditions."
2. S. Nadgir, UPDATE 2-ProtoStar files for bankruptcy, to sell satellites (2009), http://www.reuters.com/article/2009/07/29/protostar-idUSBNG6434120090729. Accessed 25 March 2011; Venture Capital Dispatch, VC-Backed ProtoStar falls out of orbit and into bankruptcy (2009), http://blogs.wsj.com/venturecapital/2009/07/29/vc-backed-protostar-falls-out-of-orbit-and-into-bankruptcy/. Accessed 25 March 2011. The ITU Radiocommunication Bureau, expressed its concern about the operation of ProtoStar satellite network without its proper registration with the ITU, particularly due to possible harmful interference as well as "risk of physical collision between the "Thuraya-3" satellite – operating on the assigned frequencies of the Emarsat-4S satellite network, which is located at 98.5° East – and the ProtoStar-1 satellite." The Bureau was of the opinion that there was "no information concerning the associated ITU satellite network filing, or the responsible administration under which the ProtoStar-1 satellite would be brought into use and operated, the Bureau was extremely

concerned and alarmed to be the witness of a situation in which a satellite, in this case, ProtoStar-1, could be operated in contravention of the ITU Constitution, particularly No. 196 and No. 18.1 of Article 18 on Licences of the Radio Regulations, and this without a responsible Administration and by an unknown operating agency not duly authorized by an ITU Member State." See Director, Radiocommunication Bureau, Report to the 48th meeting of the Radio Regulations Board, ITU Document: RRB08-3/3-E, page 4, of 4 August 2008.

3. NewSat price, turnover soar on satellite news, http://www.smh.com.au/business/newsat-price-turnover-soar-on-satellite-news-20110201-1abv4.html. Accessed 25 Feb 2011
4. Article 1.2 of the Constitution of the International Telecommunication Union (Geneva, 1992) as amended by the Plenipotentiary Conferences in 1994, 1998, 2002, 2006 and 2010 (hereinafter referred to as the ITU Constitution).
5. ITU Radio Regulations, Article 1.16, define "allocation" (of a frequency band) as an "Entry in the Table of Frequency Allocations of a given frequency band for the purpose of its use by one or more terrestrial or space radiocommunication services or the radio astronomy service under specified conditions. This term shall also be applied to the frequency band concerned."
6. ITU Radio Regulations, Article 1.17 define "allotment" (of a radio frequency or radio frequency channel) as an "Entry of a designated frequency channel in an agreed plan, adopted by a competent conference, for use by one or more administrations for a terrestrial or space radiocommunication service in one or more identified countries or geographical areas and under specified conditions."
7. ITU Constitution, Article 1.3.
8. ITU Constitution, Article 4.1.
9. ITU Constitution, Article 54.1.
10. ITU Constitution, Article 54.1.
11. ITU Constitution, Article 4.4.
12. ITU Constitution, Article 4.4.
13. ITU Constitution, Article 54.2.
14. ITU Constitution, Article 6.1.
15. ITU Constitution, Article 6.2.
16. ITU Radio Regulations, Article 18.1.
17. http://www.itu.int/members/index.html.
18. ITU Constitution, Article 2.
19. ITU Constitution, Article 2.
20. http://www.itu.int/members/sectmem/categories.html. Accessed 25 Feb 2011
21. http://www.itu.int/members/sectmem/categories.html. Accessed 25 Feb 2011
22. http://www.itu.int/members/sectmem/participation.html. Accessed 25 Feb 2011
23. http://www.itu.int/members/academia/index.html. Accessed 25 Feb 2011
24. For details about ITU's history, visit: http://www.itu.int/en/history/Pages/default.aspx
25. ITU, http://www.itu.int/net/about/structure.aspx. Accessed 25 Feb 2011
26. ITU Constitution, Article 7.
27. ITU Constitution, Article 8.1.
28. ITU Constitution, Article 8.3.
29. ITU Constitution, Article 8.2.
30. ITU Constitution, Article 10.4.
31. They are: Mexico (votes received 143), Brazil (135), Canada (135), Argentina (131), Cuba (125), Venezuela (119), United States (114), Costa Rica (93), Paraguay (91), Spain (138), Italy (136), France (135), Germany (130), Sweden (126), Turkey (125), Greece (109), Russian Federation (123), Bulgaria (116), Romania (114), Poland (107), Czech Republic (93), Egypt (122), Kenya (119), Algeria (114), Morocco (114), Ghana (112), Tunisia (111), South Africa (105), Mali (101), Burkina Faso (97), Nigeria (95), Rwanda (93), Senegal (93), Cameroon (83), Indonesia (135), China (134), Japan (133), Malaysia (127), Korea (Rep. of) (125), Bangladesh (123), Thailand (121), Australia (119), India (119), United Arab Emirates (114), Kuwait (108), Saudi Arabia (105), and Philippines (97).

32. ITU Constitution, Article 17.1.
33. ITU Constitution, Article 18.
34. ITU Constitution, Article 21.2.
35. ITU, http://www.itu.int/ITU-R/index.asp?category=information&rlink=sector-organization&lang=en
36. ITU Constitution, Article 12.1.
37. ITU Constitution, Article 12.2.
38. ITU Constitution, Article 14.
39. ITU Constitution, Article 14.1.
40. Elected by the 2010 plenipotentiary conference, the current members of the radio regulations board are: Ricardo Luis Terán (Argentine Republic); Julie Napier Zoller (United States); Alfredo Magenta (Italy); Mindaugas Zilinskas (Lithuania); Victor Strelets (Russian Federation); Baiysh Nurmatov (Kyrgyz Republic); Stanley Kaige Kibe (Kenya); Mustapha Bessi (Morocco); Simon Koffi (Côte d'Ivoire); Yasuhiko Ito (Japan); Ali R. Ebadi (Malaysia); and P. K. Garg (India).
41. ITU Constitution, 13.3.
42. ITU Constitution, 13.3.
43. ITU Constitution, Article 14.2.
44. Audrey Allison, "Latest developments in ITU radiofrequency regulations and procedures including coordination and registration, for interference-free operation of satellites", a lecture given at the Institute of Air and Space Law, McGill University, Montreal, Canada, 26 January 2008.
45. ITU Radio Regulations, Article 5.2.
46. ITU Radio Regulations, Article 5, specify that:

    5.23 *Primary and secondary services*

    5.24 1) Where, in a box of the Table in Section IV of this Article, a band is indicated as allocated to more than one service, either on a worldwide or Regional basis, such services are listed in the following order:

    5.25 *a)* services the names of which are printed in "capitals" (example: FIXED); these are called "primary" services;

    5.26 *b)* services the names of which are printed in "normal characters" (example: Mobile); these are called "secondary" services (see Nos. 5.28 to 5.31).

    5.27 2) Additional remarks shall be printed in normal characters (example: MOBILE except aeronautical mobile).

    5.28 3) Stations of a secondary service:

    5.29 *a)* shall not cause harmful interference to stations of primary services to which frequencies are already assigned or to which frequencies may be assigned at a later date;

    5.30 *b)* cannot claim protection from harmful interference from stations of a primary service to which frequencies are already assigned or may be assigned at a later date;

    5.31 *c)* can claim protection, however, from harmful interference from stations of the same or other secondary service(s) to which frequencies may be assigned at a later date.

    5.34 *Additional allocations*

    5.35 1) Where a band is indicated in a footnote of the Table as "also allocated" to a service in an area smaller than a Region, or in a particular country, this is an "additional" allocation, i.e. an allocation which is added in this area or in this country to the service or services which are indicated in the Table (see No. 5.36).

    5.36 2) If the footnote does not include any restriction on the service or services concerned apart from the restriction to operate only in a particular area or country, stations of this service or these services shall have equality of right to operate with stations of the other primary service or services indicated in the Table.

    5.38 *Alternative allocations*

    5.39 1) Where a band is indicated in a footnote of the Table as "allocated" to one or more services in an area smaller than a Region, or in a particular country, this is an "alternative"

allocation, i.e. an allocation which replaces, in this area or in this country, the allocation indicated in the Table (see No. 5.40).

5.40 2) If the footnote does not include any restriction on stations of the service or services concerned, apart from the restriction to operate only in a particular area or country, these stations of such a service or services shall have an equality of right to operate with stations of the primary service or services, indicated in the Table, to which the band is allocated in other areas or countries.

5.41 3) If restrictions are imposed on stations of a service to which an alternative allocation is made, in addition to the restriction to operate only in a particular country or area, this is indicated in the footnote.

5.50 5) The footnote references which appear in the Table below the allocated service or services apply to more than one of the allocated services, or to the whole of the allocation concerned. (WRC-2000)

5.51 6) The footnote references which appear to the right of the name of a service are applicable only to that particular service.

47. ITU Radio Regulations, Article 8.5.
48. ITU Radio Regulations, Articles 11.1 to 11.8.
49. ITU Radio Regulations, Appendix 26 (3,025 and 18,030 KHz).
50. ITU Radio Regulations, Appendix 27.
51. ITU Radio Regulations, Appendix 30A.
52. ITU Radio Regulations, Appendix 30B.
53. ITU Radio Regulations, Article 9.3.
54. ITU Constitution, Article 14.2.
55. Particularly to reduce the registration of so-called "paper or virtual satellites," the notifications for which are filed with the ITU without any serious plans for the acquisition and launch of these satellites.
56. ITU Radio Regulations, Article 11.44.
57. ITU Radio Regulations, Article 11.48.
58. ITU Radio Regulations, Article 11.49.
59. ITU Radio Regulations, Article 11.50.
60. Entitled "Administrative due diligence applicable to some satellite radiocommunication services."
61. ITU Radio Regulations, Resolution 49, Annex 1, para. 11.
62. For example: OPTOS satellite system (see Director, Radiocommunication Bureau, Report to the 56th meeting of the Radio Regulations Board, ITU Document: RRB11-1/3-E, Annex 4, of 1 March 2011); LARKSAT-IORR, LARKSAT-AOR2R, LARKSAT-NAR, LARKSAT-PORR, INSAT-NAV-A-GS, SWANSAT-3A, THAICOM-LS2, and THAICOM-LS3 satellite systems (see Director, Radiocommunication Bureau, Report to the 55th meeting of the Radio Regulations Board, ITU Document : RRB10-3/4-E, Annex 4, of 29 October 2010); INSAT-EXC55E (see Director, Radiocommunication Bureau, Report to the 50th meeting of the Radio Regulations Board, ITU Document: RRB09-1/1-E, Annex 4, of 4 February 2009); and GOES-89.5W and GOES-105W satellite systems (see Director, Radiocommunication Bureau, Report to the 48th meeting of the Radio Regulations Board, ITU Document: RRB08-3/3-E, Annex 4, of 4 August 2008).
63. As of February 2011, there are 64 States Parties to this Protocol. They are Australia, Austria, Bahrain, Barbados, Belarus, Belgium, Belize, Benin, Bosnia and Herzegovina, Botswana, Canada, Chile, Colombia, Congo (Rep. of the), Cyprus, Denmark, Egypt, El Salvador, Estonia, Finland, Greece, Guinea, Iceland, Ireland, Italy, Japan, Jordan, Kenya, Kiribati, Korea (Rep. of), Kuwait, Lao P.D.R., Latvia, Libyan Arab Jamahiriya, Liechtenstein, Lithuania, Luxembourg, Madagascar, Malta, Mauritius, Mexico, Monaco, Netherlands, New Zealand, Oman, Panama, Peru, Philippines, Portugal, San Marino, Slovenia, South Africa, Sudan, Sweden, Switzerland, Togo, Tunisia, Turkey, United Arab Emirates,

United Kingdom, Uruguay, Uzbekistan, Viet Nam and Zimbabwe. http://www.itu.int/cgibin/htsh/mm/scripts/mm.final-acts.list?_languageid=1&_agrmts_type=PROT-92. Accessed 25 Feb 2011.

64. Article 10.2 of the ITU Convention of the International Telecommunication Union (Geneva, 1992) as amended by the Plenipotentiary Conferences in 1994, 1998, 2002, 2006 and 2010.

References

P.J. Brown, Iranian in a jam over satellite blocking, ATIMES (2010), http://www.atimes.com/atimes/Middle_East/LC25Ak04.html. Accessed 26 March 2010

F. Butsch, GPS and GLONASS radio interference in Germany, http://elib.uni-stuttgart.de/opus/volltexte/1999/278/pdf/278.pdf. Accessed 25 Feb 2011

Satellite Radio Communications Fundamentals and Link Budgets

13

Daniel R. Glover

Contents

Introduction 294
Basic Systems Concepts 294
Transponders 299
Antennas 301
Digital Communications 303
Modulation and Coding 304
Shannon's Law 308
Link Budgets 309
Understanding Decibels 311
Link Budget Calculation 315
Cross-References 323
References 324

Abstract

Satellite communications makes use of radiofrequency links. Particular frequencies are allocated for satellite communications through international regulatory registration and coordination processes which prevents interference between systems. In typical operation, a satellite's transponder receives an uplinked signal from Earth, changes its frequency slightly to avoid self-interference, and retransmits it on a downlink to Earth. Antennas provide gain by focusing the transmitted energy. Path loss describes a natural spreading out of the transmitted wave front as it travels through space. A link budget is an accounting of gains and losses throughout a system that is used as a design tool to provide sufficient power (or gain) to allow a satellite connection to be established. The link margin is the

D.R. Glover
International Space University, 28505 Lincoln Road, Bay Village, OH, 44140, USA
e-mail: danglover@gmail.com

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_17, © Springer Science+Business Media New York 2013

excess amount of received signal power above what is required. Shannon's Law implies that there are trade-offs possible in a communications system design between power, bandwidth, and complexity.

Keywords

Bandwidth efficiency • Bit error rate • Coding • Decibel • Effective isotropic radiated power (EIRP) • Forward error correction • Free space loss • Gain • Line-of-sight • Link budget • Link margin • Modulation • Path loss • Satellite slot • Shannon's law • Signal-to-noise ratio • Spectrum • Trade-off • Transponder • Wavelength

Introduction

This chapter provides an overview of radio communications as they apply to commercial satellite communications systems. This chapter is not intended to be a manual for designing satellite communications systems, but rather provides the basic concepts needed to understand a system design and link budget. The general concepts are presented briefly and simply and may only hint at the actual complexity of a satellite communications system. As in any system design, satellite communications must trade-off several parameters where a choice for a given parameter results in constraining the choices for the other parameters. The major trade-offs are between power, bandwidth, and system complexity.

Basic Systems Concepts

One of two images may spring to mind when the term "satellite communications" is encountered. The image may be that of a "satellite dish" antenna on the ground. Or one might picture a spacecraft in space with its antennas and solar arrays deployed. These two images represent two major subdivisions of a satellite communications system: the flight segment (or satellite) and the ground segment (or the ground station with its associated antenna being the predominant feature). The flight segment is the part of the system that makes it a satellite communications system, but it is the ground segment that interfaces with users and with terrestrial communications systems. The ground segment is becoming a more familiar sight, with large dishes at "satellite farms" or gateways, medium sized dishes atop retail stores, and small dishes for satellite TV becoming a ubiquitous sight. To provide useful communications capabilities, a system must have at least two ground stations (one providing a transmission and another receiving the transmission) and one satellite that relays the information between the two ground stations. In addition, there is a control segment (which is usually transparent to the user) that controls the spacecraft operations. Of course, a particular system may have many satellites

and many ground stations (or user terminals), but for our purposes it is best to start with the simplest picture.

A satellite is an object (in our case, a spacecraft) that orbits around another object (in our case, the Earth). The communications spacecraft has several design constraints placed upon it (and thus, on the overall communications system) because it must be placed in orbit. Spacecraft designs are limited in their mass and volume in order to fit on the launch vehicle that places them into orbit. The mass and volume limits affect the size of the power system on the spacecraft, and so the power available is also constrained. In addition, the space environment (thermal, radiation, atomic oxygen, space debris, micrometeoroids, etc.) imposes constraints on the design (such as parts and material selection).

A spacecraft may be considered as consisting of two parts: the spacecraft "bus" and the payload. The spacecraft bus provides support services to the payload, while the payload provides the useful, or money-making, part of the satellite. Examples of payloads are: scientific instruments, remote sensing instruments, navigation service transmitters, or (in our case) communications equipment. A satellite may have one type of payload or a combination of payload types.

The spacecraft bus provides services to the communications payload including: power, structural support, attitude control and pointing, propulsion and station keeping, thermal control, commands, and telemetry. These services are provided by spacecraft bus subsystems which may vary slightly in name and content among different organizations or spacecraft manufacturers. The spacecraft bus typically has its own communications system separate from the communications payload which is used to control the spacecraft bus and the payload operations. The bus communications system is spacecraft part of the control segment. The spacecraft bus communications system receives commands from the control segment ground station and also transmits data concerning the state and status of the spacecraft bus and payload to the control segment ground station.

A typical communications payload consists of antennas and electronics designed for the reception and transmission of radio signals. Some processing of these signals is done on board which may be simple or complex depending on the system. In effect, the communications payload is a radio relay station in space with the satellite bus analogous to a tower.

Let us take a look at generic radio communications, then see where satellites fit in. Two radio stations are able to communicate when they are in "line-of-sight" or have no obstructions or barriers between them. Line-of-sight is a term that is used to describe the path between two stations along which an electromagnetic wave may travel without obstruction. As long as electromagnetic waves can propagate from one station's antenna to the other and can be distinguished from background noise or other interfering signals, communications is possible. Due to the curvature and topographic features of the Earth, as the separation distance on the ground between two locations increases, eventually the two stations will not be within line-of-sight and communications will not be possible. For two ground stations that are not within line-of-sight of each other, a relay station may be used to enable communications if the relay station is within line-of-sight of each of the two ground

stations. The relay station receives a signal from one ground station and retransmits it to the other ground station. Additional relay stations can be used to extend the distance between the two ground stations.

A relay station might be placed on a tower to allow the line-of-sight to reach farther over the horizon. The higher the tower, the farther the horizon is extended (or the line-of-sight can reach) and so the farther apart the two ground stations can be and still communicate. For two ground stations on either side of the Atlantic Ocean, a relay tower would have to be over 600 km high (above sea level) in the middle of the ocean to have a line-of-sight to each shore. Since such a large tower greatly exceeds the height of the largest structure ever built, it makes sense to use a satellite as a platform for a radio relay station. Even a series of small relay towers across an ocean would be prohibitively expensive, although undersea cables (originally telegraph cables, but now using optical fibers and repeaters) have been used since the nineteenth century and are still an effective means of bridging the oceans.

An example of a passive relay is a phenomenon that occurs at certain frequencies at night. Radio waves are affected by the ionosphere, an electrically conductive layer of the atmosphere that contains charged particles from ionization by solar radiation. The ionosphere is dynamic, as illustrated by the aurora, and changes from day to night. Certain frequencies of radio waves have their paths bent by the ionosphere to the point of being reflected when conditions are right at night, and so can propagate much farther around the world than normal. Another example of a reflective relay was the experimental Echo satellite (1960) which was a metalized balloon 30 m in diameter placed in Low Earth Orbit. Other frequency bands can penetrate the ionosphere with minimal effect and can be used for communication with spacecraft.

One parameter that describes electromagnetic waves (such as radio waves) is the frequency. An alternate parameter is the inverse of the frequency, wavelength. The frequency (f) of a wave in a vacuum is related to its wavelength (λ) by the relationship $f\lambda = c$ where c is the speed of light. Communications engineers tend to use frequency except when a physical representation is useful, as in antenna design, where wavelength is typically employed.

So, a satellite in orbit can act as a relay by receiving signals from one ground station and transmitting them back to another ground station that may not be in the line-of-sight of the first ground station. A satellite in Geosynchronous Earth Orbit (GEO) can "see" a little less than half the Earth's surface, so one satellite is not enough to provide global coverage. The path from one ground station to a satellite then on to another ground station is known as a "hop." It may take more than one hop through more than one satellite to communicate from a ground station to another on the other side of the world.

Three satellites spaced 120° around the equator at GEO can provide almost total coverage of the globe (with the poles being just out of sight). This idea was first put forward by Arthur C. Clarke in 1945 in an article in the magazine *Wireless World* entitled "Extra Terrestrial Relays." An example implementation of this concept is the NASA Tracking and Data Relay Satellite System (TDRSS) which, among other uses, provides global coverage for International Space Station and Shuttle operations.

A ground station or terminal may be a building full of electronics with a large parabolic dish antenna or as simple as a mobile telephone handset or a satellite television receiver in a home. A ground terminal might transmit, receive, or both transmit and receive. A transmitting ground station requires attention to detail in the ground terminal setup and operation as improper transmissions can interfere with other systems. Terminals may be set up as nodes for point-to-point communications between two stations, broadcast, point-to-multipoint, or in a mesh. Small ground stations that are used in satellite communications networks were named VSATs (very small aperture terminals) when they were introduced. The name comes from their relatively small antennas, although more recent ground terminals (such as satellite television receivers) have even smaller antennas.

Two communications systems that try to use the same frequency at the same time in the same location may interfere with one another such that reliable communications is not possible. Sharing frequencies requires careful coordination to prevent interference. Purposeful interference is known as jamming.

The radiofrequency (RF) spectrum is a portion of the electromagnetic spectrum that is in great demand for uses including communications. Many different uses (applications) require some amount of spectrum in order to operate. To prevent interference and to allow efficient use of spectrum, national, and international regulations have been put in place to coordinate the assignment of frequencies to applications. The international body facilitating this coordination is the International Telecommunications Union (ITU), a specialized agency of the United Nations (UN). National regulations control the use of spectrum within the borders of a country and supersede international regulations as long as other countries' spectrum use is not affected.

Figure 13.1 shows the frequency allocations for spectrum in the United States between 10 and 30 GHz. The purpose of this illustration is to indicate the incredible intricacy of the allocations within just a single country. Applications that use a portion of the RF spectrum include television broadcast, radio broadcast, mobile telephones, wireless telephones, consumer goods, microwave ovens, wireless networks, aeronautical and maritime communications, radio navigation, meteorological radar, radio astronomy, space research, microwave relays, multipoint distribution systems, fixed satellite, mobile satellite, direct broadcast television by satellite, and satellite radio broadcast. Frequency allocations are negotiated among the administrations of the world via a process established by the International Telecommunication Union (ITU) that is described separately in the chapter on this subject by Dr. Ram Jakhu. Satellite communications operators, in short, are in competition for spectrum with many other applications.

There are other types of communications besides commercial. The military may use commercial systems or their own satellite systems. There are search and rescue locating systems that use communications technology. Global navigation satellite systems are essentially highly specialized communications systems. There are satellite communications systems for collecting scientific data on Earth and deep space communications systems for scientific spacecraft exploring the solar system.

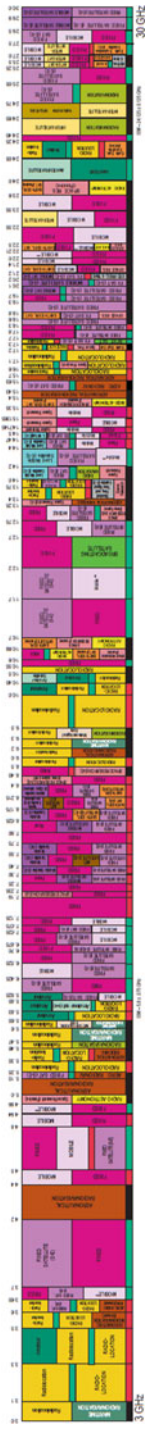


Fig. 13.1 United States frequency allocations from 10-30 GHz, (from US Department of Commerce chart "United States Frequency Allocations, The Radio Spectrum" available at http://www.ntia.doc.gov/osmhome/allochrt.pdf)

Table 13.1 Radar band designations for some frequencies used for satellite communications

Letter designation	Frequency band (approx.)
L	0.4–1.55 GHz
C	6/4 GHz
Ku	14/12 GHz
Ka	30/20 GHz

On a deep space mission, the communications system is somewhat different from that of a commercial spacecraft. There is typically a "low gain" and a "high gain" system named after the characteristics of their respective antennas. The low gain system performs the functions that the spacecraft bus communications subsystem does on a commercial satellite, namely, command and data handling, tracking functions for navigation, and housekeeping telemetry. The high gain system provides a wide bandwidth "pipe" for sending back large amounts of scientific data (e.g., image data, video, radar data, etc.). A deep space probe is like a television station in space that needs to transmit a large amount of data back to Earth. The radio system is also sometimes used as a science instrument itself as is planned with the European Space Agency's Mars Radio Science Experiment (MaRS). MaRS will use the radio signals that convey data and instructions between the spacecraft and Earth to probe the planet's ionosphere, atmosphere, surface, and even the interior. The low gain antenna also can serve as a backup to the high gain antenna as with NASA's Galileo spacecraft when a failure in the high gain antenna resulted in the science data being sent back by the low gain system (allowing much less data to be returned than would have been possible with the high gain system, but much better than nothing).

Satellite communications engineers frequently make use of old radar band designations when discussing frequency bands. Table 13.1 lists a few common letter designations for bands useful for commercial communications. Another convention is to list the uplink frequency first, then the downlink frequency when a band is used for both uplink and downlink. Two different frequencies are used to keep the uplink and downlink from interfering with each other. Thus, in C-band, a frequency band (of, say, 500 MHz) around 6 GHz is typically used for the uplink and 4 GHz for the downlink (Gagliardi 1984).

Transponders

A *transponder* is the electronics portion of the communications relay station that receives a signal, changes the frequency, and retransmits it. The word transponder is a contraction of the words transmitter and responder. Sometimes one might hear the word "repeater" used, but that is something of a misnomer in this case because a transponder intentionally changes the frequency of the received signal while a repeater usually refers to electronics that receive and retransmit a signal without any intended changes to it other than amplification.

The transponder is the basic unit of the communications payload. Another way of thinking of a transponder is as a communications channel of the satellite. A communications payload may have from one to over 100 transponders, with a typical number being a couple of dozen. Each transponder operates in a portion of the total frequency band available for use by the satellite. The bandwidth of an individual transponder might be designed to be 20 MHz or as much as 500 MHz. For many years, the typical bandwidth of a transponder was set at 36 MHz; because of this history, one may still speak of 36 MHz equivalent transponders as a unit of comparison when discussing satellite capacity. The bandwidth required to transmit an analog, studio quality video signal was 36 MHz which was a major driver in satellite communications design in the early days of the business when analog television distribution was an important application. Today, transponders can be found with bandwidths of 26, 27, 36, 54, 72 MHz, or even larger bandwidths.

A transponder that receives and retransmits a signal without affecting it (other than amplifying it and changing its carrier frequency) is said to be operating in "bent-pipe mode." This term is an analogy that compares information flowing in a communications channel to fluid flowing in a pipe. A "bent-pipe" merely changes the direction of the flow without performing any processes on the "fluid." This is not a perfect analogy (since we know the signal is amplified and the carrier frequency changed), but is good enough to be widely used. A satellite in "bent-pipe mode" does not process the signals in the channels, but is only a conduit that provides a little boost (perhaps a better analogy would include a pump at the bend, but that might be making the analogy a little too clumsy to use and still does not incorporate an analogy for the frequency translation).

Originally, satellite transponders were analog devices. With the advent of digital signals, the use of error correction coding was introduced. A regenerative transponder demodulates the signal and provides error correction on board the satellite before modulating and transmitting back to the ground. This has the advantage of correcting errors introduced on the uplink, thus regenerating the original signals before exposing them to potential errors on the downlink.

In contrast to a bent-pipe satellite is the satellite that includes onboard processing of signals. An early example of the use of onboard processing in a commercial communications satellite is the Iridium system. Iridium uses a set of satellites (known as a "constellation") in low Earth orbit (LEO) to enable personal communications through satellite telephones. Telephone calls are switched on board the satellites and relayed through the constellation and the ground system. Iridium uses inter-satellite links between the satellites in the constellation to route an individual call to its destination.

A signal is a representation of information; that information could be a voice signal, video, data, or a combination. A channel (in our case, a particular transponder's bandwidth) is spaced around a carrier frequency and has a capacity to contain a signal or collection of signals with a certain amount of information within that channel's frequency bandwidth.

Antennas

An antenna radiates and/or captures radio frequency waves to and from free space. It is the interface that couples our electronics to free space and enables telecommunications without wires. Antenna design is a specialized topic within telecommunications engineering.

Different antennas have different patterns of radiation. An isotropic antenna radiates equally in all directions (a pattern described as "omnidirectional"). Other antennas concentrate their radiation in patterns known as "lobes" or "beams." A design may attempt to concentrate all of the radiated energy into one main beam, but there are typically many additional smaller beams, known as side lobes, radiating in unwanted directions. The main lobe of the pattern has a beamwidth that is usually measured as the angle between points on that pattern that are at half the power of the peak of the main lobe. To simplify matters, we will ignore the side lobes and concentrate on the main lobe of the pattern when we talk about the antenna beam.

The antenna pattern characterizes the radiated energy (transmission) of a given antenna as well as its ability to collect energy (receive signals). If you are familiar with optics, it may help to think of an antenna as the radio equivalent of a telescope at optical wavelengths. There are similarities such as diffraction limits, diameter versus field-of-view, etc., since light and radio waves are both electromagnetic waves (just at different wavelengths). Telecommunications engineers tend to use frequency except in a few instances such as antenna design. With antenna engineers, wavelength is a more pertinent view since they are dealing with physical dimensions of shapes and sizes.

There are many different types of antennas, but the one most commonly associated with satellite communications is the parabolic dish antenna. These dish antennas have a narrow beamwidth, concentrating the energy of the radiated main beam into a smaller solid angle. This means more of the radiated energy reaches, or "illuminates," the satellite when using a dish antenna as compared to an omnidirectional (or "omni" for short) antenna. A useful analogy is: a dish antenna is to an omni as a searchlight is to a bare light bulb. This advantage of the dish relative to the omni (that is due to focusing the energy) is known as "antenna gain." The antenna gain of an omni is 1 (or 0 dB) while the antenna gain of a dish antenna is greater than one and related to the diameter (or aperture) of the dish in wavelengths.

Antenna beams allow us to reuse spectrum without interfering with other satellite systems. By using a narrow enough beam, we can communicate with one satellite and not illuminate its neighbors with unwanted electromagnetic energy. The beamwidth of ground antenna systems is the key parameter in defining the separation of spacecraft placed in the Geosynchronous Earth Orbit (GEO) arc 36,000 km above the equator at sea level. The spacing supporting a single communicating spacecraft at a particular frequency is commonly referred to as a *satellite slot*.

A crude analogy may help to illustrate the concept: Imagine communicating by Aldis lamp (a flashing signal lamp used for naval communications in the twentieth century) using Morse code. In flashing the lamp, the sender illuminates an area (at a given distance) that is a cross-section of the beamwidth of the lamp. That illuminated area hopefully includes the intended receiver, but may include other unintended receivers. Now imagine being surrounded by a ring of signalers, only one of which is the intended communication partner (or node). Now there are two problems with communicating: first is illuminating only the desired node in the ring (corresponding to a satellite in a GEO slot) and the other problem is filtering out the signal flashes coming from non-desired nodes in the ring and only receiving flashes from the desired node. The first problem is easily solved by using a lamp with a narrow enough beam such that only the desired node is illuminated (perhaps using a flashing laser if necessary). The second, reception, problem can be solved by using a mask with an aperture that emphasizes the desired node (say by cutting a hole in a piece of cardboard that masks the neighbors of the desired node).

If you were being indiscriminant in your illumination and inadvertently included neighbor nodes that you were not trying to communicate with, you would be a bad neighbor. By illuminating non-desired nodes, you are introducing distracting flashes (or interfering noise) into your neighbor's systems. To be a good neighbor and avoid jamming the communications of others, it is required to use a narrow enough beam to illuminate the desired node and not its neighbors.

In a satellite system, the antenna beamwidth solves both problems by illuminating only the desired satellite in the main lobe and masking the neighbors (since they are not in the main lobe of the antenna pattern) in the uplink. In the downlink, the ground station receives strong signals from the desired satellite and only small amounts of energy (which count as noise for our system) from the neighbors because the desired satellite is located within the main lobe (beam) and the neighbors are not. The separation between adjacent satellites needs to be at least the same amount that corresponds to the beamwidth of the antennas that are practical at that frequency (say 3° for C-band).

Polarization can be used to cram a few more channels into a given amount of bandwidth. There are two types of polarization: linear and circular. Linear polarization is the same effect that is used in sunglasses to reduce glare. A filter can be used to allow a horizontally polarized wave through while blocking a vertically polarized one. In communications, polarized antennas can be used to capture or reject polarized radiofrequency carriers.

There are two bands of frequencies in the electromagnetic spectrum that are said to be windows through the atmosphere because electromagnetic waves can pass through relatively unchanged. One window is the radio window used for satellite communications (the other is the visible window). At low frequencies, the ionosphere bends the paths of electromagnetic waves to the point of reflection, while at higher radio frequencies, the atmosphere absorbs the waves energy. Water vapor absorbs radio waves, and at high frequencies, such as at Ka band,

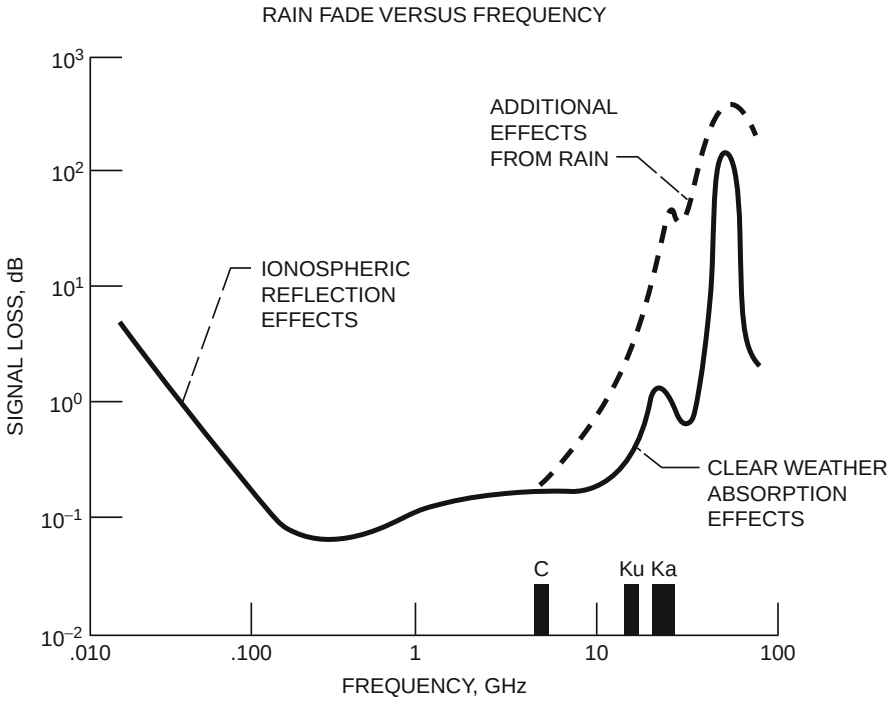


Fig. 13.2 Rain fade versus frequency (Courtesy NASA)

clouds and rain can produce significant signal loss (see Fig. 13.2) known as rain fade.

An antenna located at the equator pointing straight up at a satellite in the GEO arc will have a shorter path and look through less atmosphere (optical depth) than an antenna located at a higher geographic latitude. This distance from the ground station to the satellite is the slant range. Another parameter of ground station antenna pointing is the look angle (azimuth and elevation) by which the ground station antenna must be pointed to view a particular satellite. An antenna pointed at a GEO satellite will also be pointed at the sun for a short period of time around the equinox twice a year. The sun is a significant noise source at all frequencies and will cause interference (known as a sun outage) at those two short periods (typically 15 min) each year.

Digital Communications

Initially, satellite systems started out as analog systems and over the years have increasing become digital systems. Digital communications allows flexibility in design through the use of compression and error correction coding. Digital signals

can be perfectly reconstructed when a system is designed properly, while analog signals tend to become more corrupted in each stage of processing. Converting an analog signal to a digital signal does introduce quantization noise, but the amount of noise added can be selected (by finer quantization and sampling) and is only added once (at the time of conversion).

A signal can be converted from analog to digital through a process called pulse code modulation (PCM) that includes sampling (Jayant and Noll 1984). The analog signal is measured periodically at some sample rate. The samples are quantized to numerical values. The analog signal is reconstructed by essentially connecting the dots of the values at the sampling times. The reconstruction will be poor unless enough samples have been taken and the samples were quantized to a suitable range of values. The sample values can be represented by binary digital values (bits) and strung together into a bitstream. The bitstream (with values of 1 or 0) represents the signal.

The analog signal at its original frequency location and bandwidth, typically from 0 Hz up to some frequency that includes most of the signal energy, is the baseband signal. When digitized, the signal will typically take up much more bandwidth than the original analog signal, but that can be reduced through digital data compression. For example, a typical voice signal for telephony would have a baseband bandwidth of 4 kHz and would require 64 kbps in digital form after PCM at 8 bits/sample and 8 k samples/s. Using compression, the 64 kbps PCM data rate could be reduced to 10 kbps or even 4 kbps (with a quality vs data rate trade-off). In the digital world, data rate is sometimes referred to as bandwidth and they are closely related. Approximate data rates for some types of signals range from 32 to 320 kbps for MP3 (Motion Picture Experts Group 1 Layer III) compressed audio, 1,411 kbps for CD audio, and 2–40 Mbps for MPEG-2 compressed video (typically ~5 Mbps for standard video resolutions, ~19 Mbps for high definition TV, ~36 Mbps (1x) for Blu-ray video).

Modulation and Coding

As presented in the previous chapter, modulation is the technique of imposing a signal (say, a baseband signal) on to a radiofrequency carrier (or carriers) for transmission by varying the carrier's amplitude, frequency, or phase in accordance with the signal. A few modulation schemes are demonstrated in the figures below. These highly simplified representations are not to scale; the carrier frequency is typically much higher than the baseband signal.

Figure 13.3 depicts analog Frequency Modulation (FM). The carrier wave's frequency is varied in accordance with the change in the information signal's amplitude. Figure 13.4 depicts digital Amplitude Shift Keying (ASK, or on-off keying) modulation. In this case, a digital string of data is represented by a digital waveform. The carrier amplitude is varied according to the digital information bit stream, on for a 1 and off for a 0. The envelope for this modulated carrier is obviously not constant which would present a problem for most transponders. Figure 13.5 depicts the digital version of frequency modulation, Frequency Shift

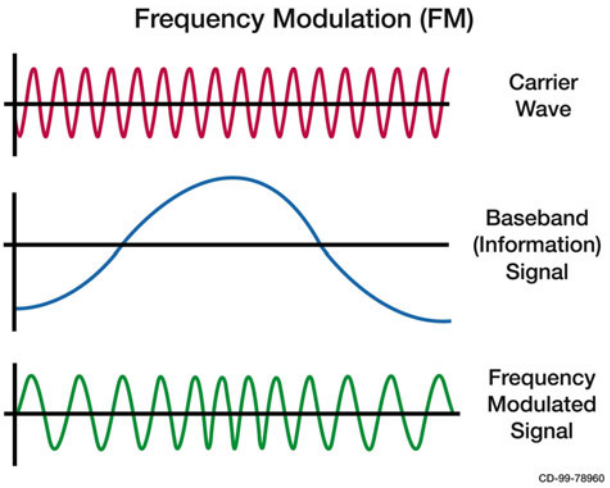


Fig. 13.3 A cartoon representation of analog frequency modulation, not to scale (Courtesy NASA)

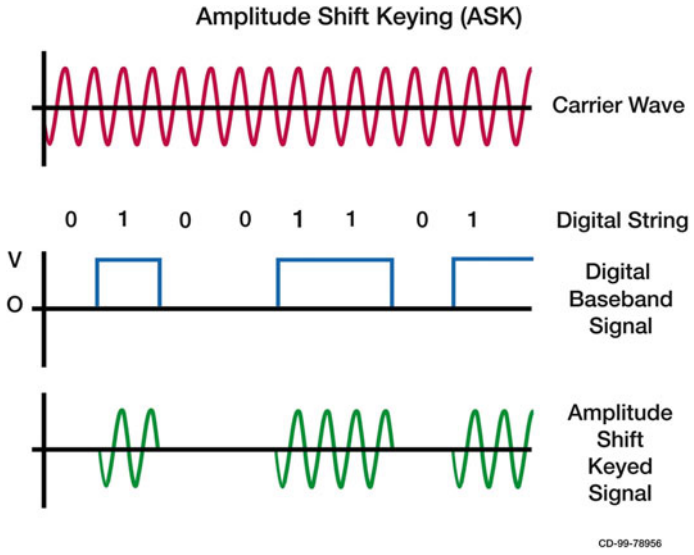


Fig. 13.4 A cartoon representation of digital amplitude shift keying modulation, not to scale (Courtesy NASA)

Keying (FSK). Figure 13.6 depicts digital Phase Shift Keying (PSK). The digital waveform used is typically a non-return-to-zero (NRZ) representation where a bit value of 1 is represented by a positive value and a bit value of 0 is represented by a negative value, but here a bit value of 0 is represented as 0 for illustrative

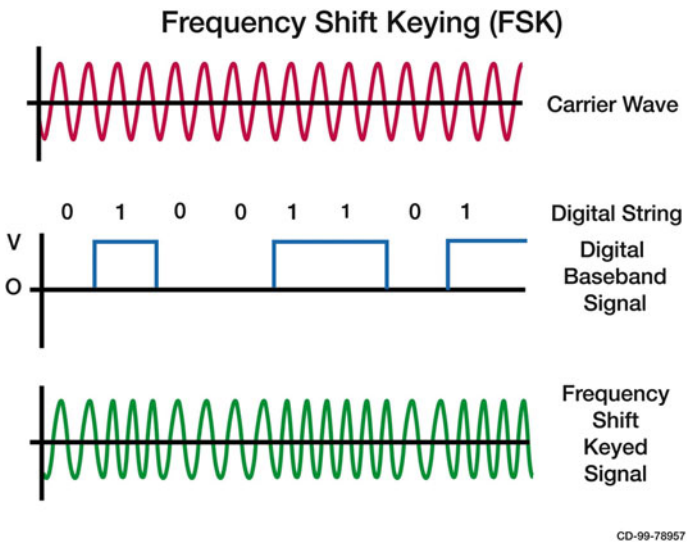


Fig. 13.5 A cartoon representation of digital frequency shift keying modulation, not to scale (Courtesy NASA)

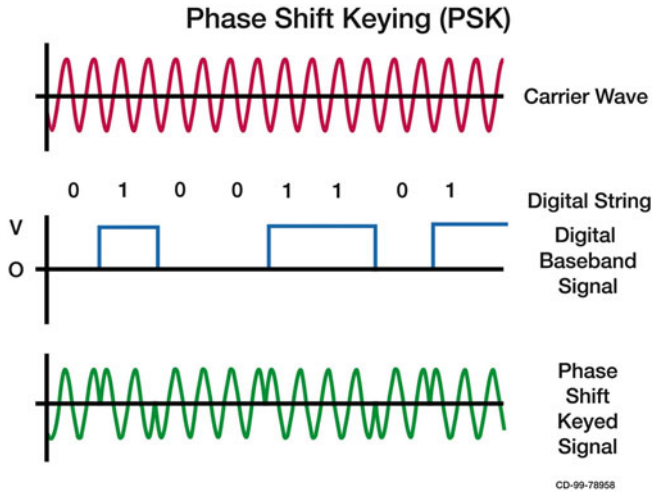


Fig. 13.6 A cartoon representation of digital phase shift keying modulation, not to scale. V represents a positive voltage and O is usually a negative voltage, but shown here as 0 for illustrative purposes (Courtesy NASA)

purposes only. There are many other methods of representing a bit stream as a waveform such as Manchester (or bi-phase) encoding.

A version of PSK, Quadrature Phase Shift Keying (QPSK) is widely found in satellite systems. Biphase PSK (BPSK) is sometimes encountered. QPSK is better

Table 13.2 Bandwidth efficiency of various digital modulation schemes

Modulation scheme	Theoretical bandwidth efficiency
BPSK	1 bit per second (bps)/Hz
QPSK	2 bps/Hz
8-PSK	3 bps/Hz
16-QAM	4 bps/Hz

from the standpoint of bandwidth efficiency as it is equivalent to two independent (orthogonal) BPSK schemes combined. Although BPSK is a simpler method, QPSK chipsets have become widely available making it a cost effective choice. Bandwidth efficiency refers to the amount of bandwidth required to accommodate a given amount of data. The more bandwidth efficient a scheme, the more data that can be accommodated in a given channel bandwidth, but generally this results in a more complicated scheme. In other words, there is usually a complexity versus bandwidth trade-off to be made in selecting a modulation scheme, amongst other considerations.

Advanced modulation schemes include 8-PSK, Quadrature Amplitude Modulation (QAM) including 16-QAM, 64-QAM, and 256-QAM where the numbers refer to the number of symbol states in the scheme. Table 13.2 shows the theoretical bandwidth efficiency for various modulation schemes. A system's bandwidth efficiency is affected by more than just the modulation scheme and will be less than that shown in the table.

System bandwidth efficiency will be less than that shown in Table 13.7 due to things like pulse shaping (typically raised cosine filtering (Hayken 1983)) and coding. A QPSK system might have a bandwidth efficiency of 1.5 bps/Hz or less.

An important measure of quality in a digital system is the bit error rate (BER). The BER is the number of bits in error divided by the total number of bits sent. Bit errors can occur randomly spaced or in blocks. The BER is usually assumed to be an average over some period of time. For example, if one bit is in error on average for every million bits sent, the BER would be $1/10^6$ or 10^{-6}. Error correction coding, also known as channel coding, can be used to correct some errors. Coding requires that additional bits are inserted which adds information redundancy to the bitstream at the cost of increasing the number of bits that must be sent.

In a simple example of error correction coding, each original data bit could be sent three times. The original data bit is combined with two parity bits that provide a three bit representation in the transmission data stream. One bit error can be detected and corrected since it will conflict with two unchanged bits. This triples the number of bits that must be sent compared with the original data, however. If the original bit was only sent once, then a bit error could be detected, but not corrected since the receiver will not know which bit is the correct value. It turns out that there are more complicated codes that allow various levels of error detection and correction that use fewer than three bit symbols. The amount of redundancy that a particular code adds is described by the code rate where the rate is the number of original data bits divided by the total number of bits (data and parity) that are sent. For example, the rate of our inefficient tripling code above is 1/3.

There are several types of error correction coding including block codes, convolutional codes, concatenated codes, turbo codes, low density parity check (LDPC) codes, etc (Peterson and Weldon 1991). They add varying amounts of complexity and provide varying levels of protection. Source coding refers to data compression coding. Channel coding adds redundancy to the bitstream while source coding removes redundancy. A bit error in a compressed data stream may cause damage to more than one data bit when uncompressed.

Coding gain is a measure of the power savings that a coding scheme achieves compared to an uncoded signal in the same system. The power savings comes at the cost of increased bandwidth and complexity as coding adds redundancy to the original bitstream.

Multiple access (Bhargava et al. 1981) refers to the sharing of a communications channel by more than one signal. It is similar to multiplexing, but is a bit more complicated because of the interactions between the signals. There are three main types of multiple access: Time Division Multiple Access (TDMA), Frequency Division Multiple Access (FDMA), and Code Division Multiple Access (CDMA). In TDMA, each signal (which may itself be a collection of signals) uses the channel for a specified period of time. This requires complex time synchronization. In FDMA, each signal gets its own frequency band subchannel. In CDMA, each signal is encoded with a pseudo-random code which spreads the signal over a larger bandwidth (Dixon 1976) (CDMA is a spread spectrum technique) and is extracted using the same code. Multiple access usually takes place at an intermediate radio frequency and there are complex interactions between signals that can affect them.

Intermodulation distortion can occur when several carriers share a nonlinear device, such as some power amplifiers. An amplifier may have a linear transfer function in a portion of its operating curve, but may be nonlinear at the high power output portion of its operating curve. The amplifier may be operated at a back off point compared to its maximum power output to ensure that it is operating as a linear device to minimize intermodulation distortion effects if several carriers are sharing it.

Shannon's Law

Claude Shannon was one of the seminal thinkers of the twentieth Century. He worked as a researcher at Bell Labs in its heyday. By applying Boolean algebra to telephone switching circuits, he set the stage for digital electronics design. He developed a mathematical theory of communication that included the foundation of information theory (Hamming 1986). One result of the theory is a description of the limit of the amount of information that can be transmitted on a given channel (that is sometimes known as "Shannon's Law" or the "Shannon–Hartley Theorem") commonly called the channel capacity. The "Shannon limit" refers the theoretical best rate of transmission for a given channel. The channel capacity equation is:

$$R \leq W \log2(P_R/P_N + 1)$$

where
R = theoretical maximum transmission rate, bits per second
W = usable bandwidth, Hertz
P_R = power of received signal
P_N = power of noise

The Shannon limit implies that a system can be designed (i.e., with sufficient complexity, for example, using error-correcting codes) that can allow transmission on a channel that approaches the rate of the Shannon limit. For error free communications, E_b/N_0 has to be at least −1.6 dB, even with infinite bandwidth. Error correction coding can reduce the power required, but uses more bandwidth. By adding coding to the system, complexity, and bandwidth are increased, but power required is decreased (i.e., complexity and bandwidth are traded for power).

Link Budgets

Communications is achieved by introducing into free space a modulated radio wave (transmitting) which propagates over a distance to a point where a receiving station captures a portion of the transmitted energy and extracts information from it. As the radio wave propagates through space, it spreads out and the energy transmitted is spread across an increasingly larger cross-sectional area; less energy is available to be captured by a given antenna as the distance increases. Although we have been talking about beamwidth as a two-dimensional measurement characterizing an antenna, the transmitted wave actually propagates in a three-dimensional solid angle. A solid angle can be visualized as a cone whose apex is at the transmitting antenna having a base, or cross-section of the beam, at any given distance from the antenna (this approximation is an illustrative aid, since waves are usually assumed to propagate spherically which would mean that the base bulges out as a spherical section). The area of the cross-section of the beam gets larger as the distance from the antenna increases and the energy of the transmitted wave is spread over that increasingly larger cross-section. The received power flux density drops off inversely proportional to the square of the distance.

As the distance between the transmitter and receiver increases, the signal power decreases (all else being equal). At some point the received signal power will be less than the noise that is received/generated by the receiver and reliable communications will not be possible. The greatest contribution to the decrease in signal power received is the loss due to the propagation distance known as the path loss or free space loss. The free space loss, L_s is given by:

$$L_s = (4\pi d)^2/\lambda^2$$

where
d is the distance in meters (or whatever unit is used for λ)
λ is the wavelength in meters

The wavelength term is not part of the physical path loss, but is related to the receiving antenna's aperture (and related gain) and is included with the physical path loss.

The propagation distance from the ground to a geostationary satellite is around 36,000 km (the distance varies depending on the latitude and longitude of the ground station and the location of the satellite in the geostationary arc). At a frequency of 6 GHz, the free space loss is approximately 1×10^{20}. Because this is a loss, this means that the signal will be 10^{-20} times weaker when the path loss is accounted for.

Although it is usually the largest factor by far, the path loss is not the only contribution to the received power; there are many factors that can boost or decrease the power of the transmitted wave or that can introduce noise to the system. The ability to communicate depends on being able to receive the desired signal at a level above the received and generated noise. Factors that decrease signal (loss) or increase noise are detrimental to reliable communications, whereas factors that increase signal (gain) or decrease noise are beneficial.

The system *link budget* is used to account for the various gains and losses across a communications link between two nodes. Part of the link budget is calculated according to a conceptually simple equation accounting for gains and losses:

$$T_p * G_{link}/L_{link} = R_{\mathbf{p}}$$

where
T_p is the transmitted power
G_{link} is the product of all the gains in the link
L_{link} is the product of all the losses in the link
R_p is the received power

Once we have the received power, we can calculate the signal-to-noise ratio (SNR):

$$SNR = S/N$$

where, in the simplest case, the received signal power, $S = R_P$ and N is the received noise (including noise generated by the receiver). The signal-to-noise ratio is the key parameter in determining whether or not a communication link will be successful (or in the vernacular, whether it is possible to "close the link"). The higher the SNR, the better the link. The signal is modulated onto a radiofrequency (RF) wave known as a *carrier*. Sometimes the Carrier to Noise ratio, C/N or CNR, is used in satellite system design to characterize a link at the RF level before demodulation, whereas the SNR is a measure of the signal strength after demodulation. An RF carrier will usually carry many baseband signals. In a digital system, the signal-to-noise quantity we are looking for is the ratio of the energy per bit to noise density, E_b/N_0.

In a satellite relay system, there are two links that need to be analyzed: (1) the uplink, or the link from the ground to the satellite, and (2) the downlink,

or the link from the satellite to a second ground station. The uplink and the downlink are analyzed separately and the results combined to get the overall link budget.

The key result from a link budget is a quantity known as the link margin. This is the difference between the actual received power and the required received power in decibels (see below for a discussion of decibels). Hopefully the link margin is a positive quantity; a negative quantity indicates that the link will not be closed and communications will not be successful. Generally the higher the link budget the better from an engineering standpoint (too high and a cost accountant somewhere should be complaining).

Understanding Decibels

The range in values of the various factors involved in calculating the link budget is very large. Engineers use a dimensionless unit, the *decibel* or dB (named after Alexander Graham Bell), to simplify the calculations. A result in decibels, R, is defined as a ratio of two powers, P1 and P2 such that:

$$\mathrm{R\ dB} = 10\ \log_{10}(\mathrm{P_1}/\mathrm{P_2})$$

or ten times the logarithm (to the base 10) of the ratio of the two power quantities. (The logarithm of a number is the exponent to which the base is raised to get that number.) Some examples of the conversion to decibels:

Let the ratio $\mathrm{P_1}/\mathrm{P_2} = 10$ (that is $\mathrm{P_1} = 10 \times \mathrm{P_2}$). Then

$$\mathrm{R} = 10\ \log_{10}(10) = 10\ \mathrm{dB}$$

since $\log_{10}(10) = 1$. Admittedly, this example is not a very exciting transformation. Let us try some more.

Let the ratio $\mathrm{P_1}/\mathrm{P_2} = 1$ (that is $\mathrm{P_1} = \mathrm{P_2}$). Then

$$\mathrm{R} = 10\ \log_{10}(1) = 0\ \mathrm{dB}$$

since $\log_{10}(1) = 0$.

Let the ratio $\mathrm{P_1}/\mathrm{P_2} = 1/10$ (that is $10 \times \mathrm{P_1} = \mathrm{P_2}$). Then

$$\mathrm{R} = 10\ \log_{10}(1/10) = -10\ \mathrm{dB}$$

since $\log_{10}(10^{-1}) = -1$. That is still not very impressive. Let us look at something a little bigger.

Let the ratio $\mathrm{P_1}/\mathrm{P_2} = 1{,}000{,}000$ (that is $\mathrm{P_1} = 1{,}000{,}000\ \mathrm{P_2}$). Then

$$\mathrm{R} = 10\ \log_{10}(1,000,000) = 60\ \mathrm{dB}$$

since $\log_{10}(10^6) = 6$. Ah, now we are getting somewhere. Looking at the result (60 dB), we can see that there should be six zeros in the original ratio because of the

Table 13.3 Decibels provide compact representations of large dynamic ranges

Ratio	Exponent	Decibels
1	10E0 or 10^0	0 dB
10	10E1	10 dB
100	10E2	20 dB
1,000	10E3	30 dB
1,000,000	10E6	60 dB
1,000,000,000	10E9	90 dB
1,000,000,000,000	10E12	120 dB
100,000,000,000,000,000,000	10E20	200 dB

logarithmic nature of the dB. Table 13.3 shows that working with decibels makes it easier to work with power ratios having large dynamic ranges.

Dealing with large ranges of ratios cannot be the only reason for using decibels since scientific notation also makes it easier to work with large ranges of numbers and does not require calculations. Another reason for using decibels is that the equivalent function to multiplying ratios is adding decibels and the equivalent function to dividing ratios is subtracting decibels. This is because of the logarithm in the definition. That means that instead of multiplying gains and dividing losses to get a link budget, we can convert to decibels and add gains and subtract losses. For example:

$$1/10 = 0\ \text{dB} - 10\ \text{dB} = -10\ \text{dB}$$

as we saw before. We will come back to link budgets, but first a few more examples.

Let the ratio $P_1/P_2 = 2$ (that is $P_1 = 2P_2$). Then

$$R = 10\ \log_{10}(2) = 3.0\ 1\text{dB}$$

since $\log_{10}(2) = 0.30102999\ldots$. Most engineers round off and use 3 dB to mean a ratio of 2. The power ratio equivalent of 3 dB is calculated:

$$10E(R/10) = P_1/P_2$$

$$10^{(3/10)} = 1.99526\ldots$$

which is approximately 2 as we would expect. Some other useful values to remember are given in Table 13.4.

Knowing the conversions for 2, 3, 7, and 10 we can get the rest from easy manipulation of the decibel values (the examples below use various levels of rounding off):

$$4 = 2 \times 2 = 3\ \text{dB} + 3\ \text{dB} = 6\ \text{dB}$$

$$5 = 10/2 = 10\ \text{dB} - 3\ \text{dB} = 7\ \text{dB}$$

Table 13.4 Converting numbers from 1 to 10 to decibels (rounded to two decimal places)

Ratio	dB
1	0
2	3.01
3	4.77
4	6.02
5	6.99
6	7.78
7	8.45
8	9.03
9	9.54
10	10.00

Table 13.5 Decibel values from 0 to 10 and their corresponding ratio values to two decimal places

Decibels	Ratio
0	1
1	1.26
2	1.58
3	1.99
4	2.51
5	3.16
6	3.98
7	5.01
8	6.31
9	7.94
10	10

$$6 = 2 \times 3 = 3.0 \text{ dB} + 4.8 \text{ dB} = 7.8 \text{ dB}$$

$$8 = 2 \times 4 = 3 \text{ dB} + 6 \text{ dB} = 9 \text{ dB}$$

$$9 = 3 \times 3 = 4.77 \text{ dB} + 4.77 \text{ dB} = 9.54 \text{ dB}$$

With these values, it is easy to put together whatever number we need in decibels. For example:

$$2,000,000 = 2 \times 1,000,000 = 3 \text{ dB} + 60 \text{ dB} = 63 \text{ dB}$$

$$400 = 4 \times 100 = 6 \text{ dB} + 20 \text{ dB} = 26 \text{ dB}$$

Table 13.5 shows the case for the inverse conversion from decibels to power ratios rounded off to two places. As we have already seen, 3 dB is typically rounded off to a value of 2, which would have happened if we rounded these values off to one decimal place.

To convert, say, 26 dB to a ratio, one could use the definition and calculate:

$$10\mathrm{E}(26/10) = 398$$

or use some simple arithmetic and some memorized conversions (i.e., that 6 dB is approximately a ratio of 4):

$$26\ \mathrm{dB} = 20\ \mathrm{dB} + 6\ \mathrm{dB} = 100 \times 4 = 400$$

which is hopefully close enough for the desired application.

Up to this point we have been talking about converting power ratios to decibels because that is the way decibels are defined. Of course, if one needs to convert a number, the special case where $P_2 = 1$ can be used, so that the ratio P_1/P_2 becomes P_1. At the risk of confusing the reader, there are cases where dB are used for amplitude ratios where the definition is $20\ \log_{10}\ (A_1/A_2)$. This definition is not used in the link budget discussion below, but is provided for completeness in the decibel discussion. The amplitude definition of decibels, dBV (V for Volts in this case), can be encountered by replacing the power quantities in the power ratio by voltages:

$$P = V^2/Z$$

where P = power, V = voltage, and Z = impedance, so that

$$\begin{aligned} R &= 10\ \log_{10}(P_1/P_2) \\ &= 10\ \log_{10}\left(\left(V_1{}^2/Z\right)/\left(V_2{}^2/Z\right)\right) \\ &= 10\ \log_{10}(V_1/V_2)^2 \\ &= 20\ \log_{10}(V_1/V_2) \end{aligned}$$

because $\log_{10} X^2 = 2\ \log_{10} X$.

The decibel is not an official unit of the International System of Units (SI), however it is approved by the International Committee for Weights and Measures (CIPM, *Comité International des Poids et Mesures*) for use with the SI. A common practice, which is not in accordance with official SI rules, is to attach a reference unit indicator to the dB symbol, such as dBW, dBm, dBHz, etc. as a shorthand way of specifying a reference level (Ambler and Taylor 2008). These units indicate a reference value in order to use the decibel definition in a specialized way as an absolute unit rather than a ratio (relative) unit. For example, dBW indicates a measurement relative to a value of 1 W (or $P_2 = 1$ W). Thus:

$$R\ \mathrm{dBW} = 10\ \log_{10}(P_1/1\mathrm{W})$$

so that, if $P_1 = 1$ W, then R = 0 dBW. In SI units, the reference value of 1 W would have to be specified explicitly each time, i.e., R (re 1 W) = 0 dB or $R_{(1\ \mathrm{W})} = 0$ dB.

Table 13.6 Informal decibel reference indications

Informal reference unit	Reference value
dBW	1 W
dBm	0.001 W or 1 milliWatt
dBi	Isotropic antenna gain
dBHz	1 Hz

Common reference values that may be encountered in non-SI literature are shown in Table 13.6.

Link Budget Calculation

The simplified power balance equation (see above):

$$T_p * G_{link}/L_{link} = R_{\mathbf{p}}$$

using decibels becomes:

$$T_p dB + G_{link} dB - L_{link} dB = R_{\mathbf{p}} dB$$

where
T_p dB is the transmitted power in decibels (referenced to W or mW)
G_{link} dB is the sum of all the gains in the link in decibels
L_{link} dB is the sum of all the losses in the link in decibels
R_p dB is the received power in decibels (referenced to W or mW as in T_p)

Although this equation is very simple, the challenge is in finding all of the significant gains and losses and keeping their signs correct. The definition of a loss, for example, could be as a positive quantity to be subtracted or as a negative quantity to be added. For example, the free space loss to GEO at 6 GHz (which we saw above is 10^{20}) is 200 dB which is subtracted in the link budget equation.

The transmitted power and transmitter antenna gain are usually combined in a quantity called the Effective Isotropic Radiated Power (EIRP). As the name implies, EIRP is the power that would have to be radiated from an isotropic antenna (an antenna that radiates equally in all directions) to provide the same power flux density at the receiver:

$$EIRP = P_T - L_T + G_T dBW$$

where
P_T = transmitter power
L_T = transmitter losses
G_T = transmit antenna gain

The important parameter to the receiver is the received power. Different combinations of transmitter hardware can provide the same received power and EIRP is

a way of representing the transmitted power in a standard way. For example, if we ignore the losses, an EIRP of 50 dBW could be obtained by any number of combinations of transmit power and antenna gain (e.g., $P_T = 15$ dBW and $G_T = 35$ dB, or $P_T = 20$ dBW and $G_T = 30$ dB, etc.). Higher EIRP results in higher receiver power, all else being equal, but there are regulatory limits on the amount of power that can be radiated onto the Earth. These regulatory limits are for the purpose of limiting interference to other communication systems.

Another concept, G/T, is used to characterize the receiver. The receiver figure of merit is the ratio of the receiving antenna gain, G, to the receiver system noise temperature, T (in Kelvins). The higher the G/T the better the receiver. The noise temperature is a way of lumping together various noise contributions many of which are thermal in origin and the remainder of which are modeled as thermal. The received noise power is related to the noise temperature by Boltzmann's constant and the bandwidth of interest.

The simplified link budget equation (Miller et al. 1993) can be written using EIRP and G/T:

$$C/N = EIRP - L_s + 10\log_{10}(G/T) - 10\log_{10}(kB)$$

where k is Boltzmann's constant and B is the equivalent noise bandwidth of the receiver.

For the purposes of this chapter, it is sufficient to realize that there are many quantities that need to be determined for entry into a link budget calculator, some of which are combinations of lower level details of the system design. The level of detail in the link budget depends on its purpose. As a design tool, the systems engineer will have a link budget that may have dozens of separate items that allow trade-offs to be made. Once a system is built and losses become more determined and tightly range bound, some quantities can be combined into a simplified link budget for use in, say, determining a particular ground station design.

A detailed design link budget may include gains and losses due to items such as: circuits (may include cables, connectors, waveguides, switches, etc.), antenna size, antenna polarization, antenna alignment, weather, atmospheric gasses, thermal noise, active device noise, interference, back off or attenuation, coding, etc.

There are many free and commercial automated tools for calculating radiofrequency link budgets and especially link budgets for satellite systems. Many link budget tools for educational purposes require the user to enter all parameters, but may provide a limited number of variable parameters by lumping detailed parameters together. This provides a nice platform for getting the feel of a link budget without getting bogged down in the details of a design (Fig. 13.7).

A systems engineer doing a detailed design will probably use a spreadsheet to build their own tool so that they can introduce parameters as needed. Some of the parameters in the link budget are related to the satellite design (e.g., transponder bandwidth, transmitted power, antenna gain) and some are related to the ground station design (e.g., terminal location, antenna gain, receiver sensitivity). To add services with an existing satellite, the satellite design parameters are already

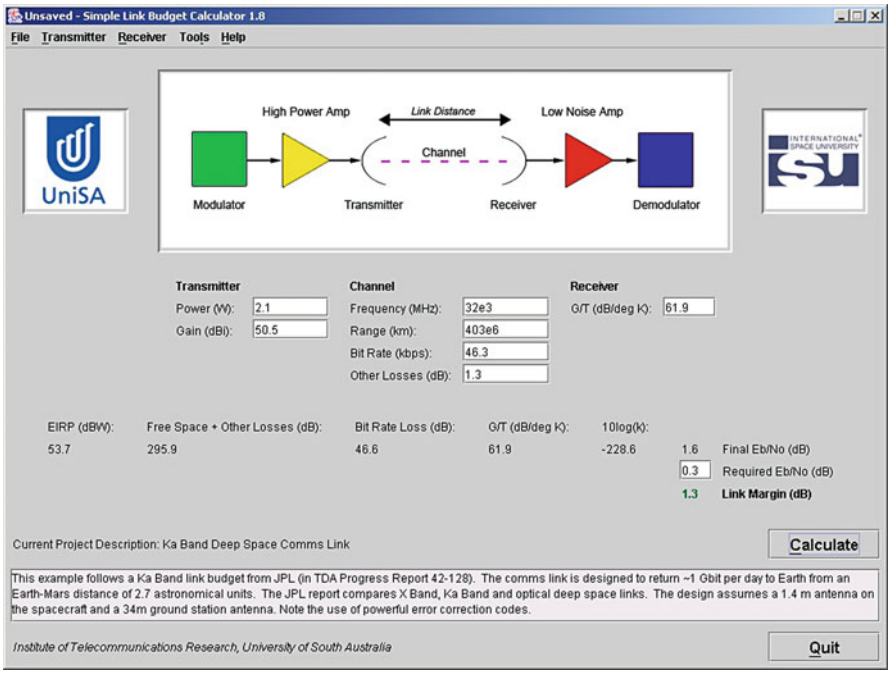


Fig. 13.7 Example of a free tool for demonstrating link budget concepts (Courtesy UniSA, from http://www.itr.unisa.edu.au/itrusers/bill/public_html/software/)

determined since it is already deployed in space and can be obtained from the satellite service provider or from a database.

Commercial tools allow potential users to investigate system options by including a database of existing satellite and ground systems with parameters needed to conduct link budget analyses in addition to a tool for doing the calculations.

Some software packages provide a useful representation of a satellite transponder's coverage called a footprint map (see Fig. 13.8). A footprint is the intersection of a satellite's beam with the surface of the Earth. The contours of the map show the areas of signal strength due to the antenna pattern. Note in Fig. 13.4 that the designer of the spacecraft antenna shaped the antenna coverage contour to evenly spread the maximum EIRP output over the US northern continental area and the populous areas of Canada. Presumably, this is the area of the primary intended customer base.

Table 13.7 shows sample output from a commercial software package called SatFinder. This table indicates the considerable complexity found in an actual link budget calculation for a particular satellite. SatFinder has a database with information on over 500 satellites (including orbital location, frequencies, EIRP, G/T, bandwidth, etc.). Similar satellite link budget tools include products that can be found on line such as Satmaster or Customizable Link Budget Tool.

Some commercial network modeling tools, such as OpNet, have satellite link modeling capability and can integrate with other tools, such as Analytical

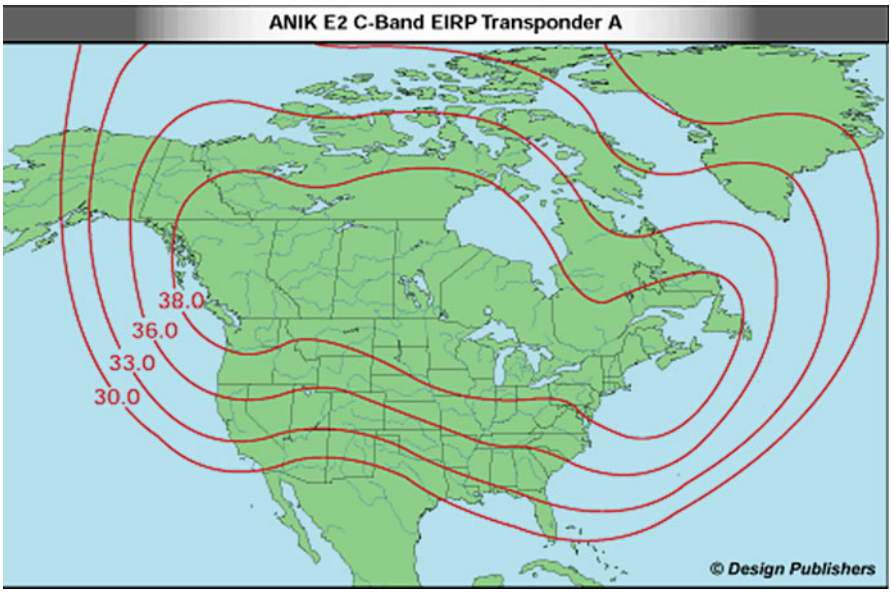


Fig. 13.8 Example footprint map from SatFinder (Courtesy SatNews.com)

Graphics, Inc.'s Satellite Tool Kit (STK), to provide analysis of satellite systems. STK has several nice communications functions and tutorials that can help in understanding link budgets and system trade-offs. There are tools from other companies that are specialized for regulatory coordination to avoid interference between systems. Examples include Visualyse (from Transfinite) and Sat-Coord (from RPC Telecommunications).

The main result from a link budget analysis is the link margin. This is the amount of received power over and above the minimum amount required to close the link and meet the system requirements for quality. A good value for a link margin in a fixed satellite system (where the ground stations are well characterized and design parameters do not change radically) is around 3 dB. For high frequency systems (Ka band), rain fade is an important design consideration. From Fig. 13.2 we can see that rain fade may require 10 dB or more of margin, but there are techniques for dealing with rain including site diversity and adaptive power control.

For a mobile satellite system, at least one ground terminal is a mobile device, perhaps handheld. In that case, the design parameters may be more variable depending on how and where the handset is being used. Handheld devices generally need to be low power (to minimize radiation of the user) and have low gain antennas (to avoid having to point the antenna at the satellite). In an urban environment, there is shadowing from buildings as well as reflections that cause multipath distortions. Mobile systems may make use of LEO constellations or a GEO satellite. In a LEO system, there may be a need for some margin for handoffs between satellites. In general, a link margin of 10–15 dB might be appropriate for

Table 13.7 Example output from SatFinder link budget software

Digital link budget produced using satfinder (http://www.satnews.com/linksample.htm)			
Service Name	Unspecified		
Coverage	Unspecified		
Uplink earth station	Denver		
Downlink earth station	Los Angeles		
Satellite name	Telstar 402R		
Link Input Parameters	*Up*	*Down*	*Units*
Site latitude	39.73 N	34.05 N	degrees
Site longitude	105.00 W	118.25 W	degrees
Site altitude	1	0.1	km
Frequency	14.472	12.172	GHz
Polarization	Vertical	Vertical	–
Rain model	ITU (30.3)	ITU (19.7)	(mm/h or zone)
Availability (average year)	99.9	99.9	%
Water vapor density	3	10	gm/m^3
Surface temperature	10	20	°C
Antenna aperture	10	0.9	meters
Antenna efficiency/gain	65	70	% (+ prefix dBi)
Coupling loss	0.2	0	dB
Antenna tracking/mispoint error	0	0.3	dB
LNB noise figure/temp	–	0.75	dB (+ prefix K)
Antenna noise	–	27	K
Adjacent carrier interference	30	30	dB
Adjacent satellite interference	30	12	dB
Cross polarization interference	200	200	dB
Uplink station HPA output back off	3	–	dB
Number of carriers/HPA	1	–	–
HPA C/IM (up)	200	–	dB
Uplink power control	0	–	dB
Uplink filter truncation loss	0	–	dB
Required HPA power capability	MAX	–	W
Satellite Input Parameters	*Value*	*Units*	
Satellite longitude	89.00 W	degrees	
Transponder type	TWTA	–	
Receive G/T	1	dB/K	
Saturation flux density	−86.05	dBW/m^2	
Satellite attenuator pad	0	dB	
Satellite ALC	0	dB	
EIRP (saturation)	47	dBW	
Transponder bandwidth	27	MHz	
Input back off total	1	dB	
Output back off total	0.3	dB	
Intermodulation interference	30	dB	

(*continued*)

Table 13.7 (continued)

Digital link budget produced using satfinder (http://www.satnews.com/linksample.htm)				
Number of transponder carriers	AUTO	–		
Carrier/Link Input Parameters	*Value*	*Units*		
Modulation	4-PSK	–		
Required bit error rate performance	10^{-9}	–		
Required Eb/No without FEC coding	20	dB		
Required Eb/No with FEC coding	5.1	dB		
Information rate	23.6	Mbps		
Overhead	0	%		
FEC code rate	0.59	–		
Spreading gain	0	dB		
Reed Solomon code	1	–		
(1 + Roll off factor)	1.2	–		
Carrier spacing factor	1.3	–		
Bandwidth allocation step size	0.1	MHz		
System margin	0.9	dB		
Calculations at Saturation	*Value*	*Units*		
Gain 1 m^2	44.67	dB/m^2		
Uplink C/No	98.88	dBHz		
Downlink C/No	89.72	dBHz		
Total C/No	79.59	dBHz		
Uplink EIRP for saturation	76.65	dBW		
General Calculations	*Up*	*Down*	*Units*	
Elevation	41.14	39.61	degrees	
True azimuth	155.84	134.99	degrees	
Compass bearing	145.77	121.46	degrees	
Path distance to satellite	37692.47	37809.06	km	
Propagation time delay	0.125728	0.126117	seconds	
Antenna efficiency	65	70	%	
Antenna gain	61.75	39.65	dBi	
Availability (average year)	99.9	99.9	%	
Link downtime (average year)	8.766	8.766	hours	
Availability (worst month)	99.615	99.615	%	
Link downtime (worst month)	2.809	2.809	hours	
Spectral power density	−59.11	−26.31	dBW/Hz	
Uplink Calculation	*Clear*	*Rain Up*	*Rain Dn*	*Units*
Uplink transmit EIRP	75.65	75.65	75.65	dBW
Transponder input back off (total)	1	1	1	dB
Input back off per carrier	1	4.32	1	dB
Mispoint loss	0	0	0	dB
Free space loss	207.18	207.18	207.18	dB
Atmospheric absorption	0.1	0.1	0.1	dB
Tropospheric scintillation fading	0.08	0.08	0.08	dB

(*continued*)

Table 13.7 (continued)

Digital link budget produced using satfinder (http://www.satnews.com/linksample.htm)				
Atmospheric losses total	0.18	0.18	0.18	dB
Total path loss (excluding rain)	207.36	207.36	207.36	dB
Rain attenuation	0	3.32	0	dB
Uplink power control	0	0	0	dB
Uncompensated rain fade	0	3.32	0	dB
C/No (thermal)	97.88	94.56	97.88	dB.Hz
C/N (thermal)	24.08	20.76	24.08	dB
C/ACI	30	26.68	30	dB
C/ASI	30	26.68	30	dB
C/XPI	200	196.68	200	dB
C/IM	200	200	200	dB
Eb/(No + Io)	13.26	13.26	13.26	dB
Downlink Calculation	*Clear*	*Rain Up*	*Rain Dn*	*Units*
Satellite EIRP total	47	47	47	dBW
Transponder output back off (total)	0.3	0.3	0.3	dB
Output back off per carrier	0.3	3.62	0.3	dB
Satellite EIRP per carrier	46.7	43.38	46.7	dBW
Mispoint loss	0.3	0.3	0.3	dB
Free space loss	205.71	205.71	205.71	dB
Atmospheric absorption	0.1	0.1	0.1	dB
Tropospheric scintillation fading	0.3	0.3	0.3	dB
Atmospheric losses total	0.4	0.4	0.4	dB
Total path loss (excluding rain)	206.11	206.11	206.11	dB
Rain attenuation	0	0	2.74	dB
Noise increase due to precipitation	0	0	4.14	dB
Downlink degradation (DND)	0	0	6.89	dB
Total system noise	81.67	81.67	212.02	K
Figure of merit (G/T)	20.23	20.23	16.09	dB/K
C/No (thermal)	89.42	86.1	82.53	dB.Hz
C/N (thermal)	15.61	12.29	8.73	dB
C/ACI	30	26.68	30	dB
C/ASI	12	8.68	12	dB
C/XPI	200	196.68	200	dB
C/IM	30	26.68	30	dB
Eb/(No + Io)	−4.15	−4.15	−4.15	dB
Totals per Carrier (End-to-End)	*Clear*	*Rain Up*	*Rain Dn*	*Units*
C/No (thermal)	79.59	79.59	79.59	dB.Hz
C/N (thermal)	5.78	5.78	5.78	dB
C/ACI	23.19	23.19	23.19	dB
C/ASI	6.15	6.15	6.15	dB
C/XPI	193.19	193.19	193.19	dB
C/IM	26.2	26.2	26.2	dB

(*continued*)

Table 13.7 (continued)

Digital link budget produced using satfinder (http://www.satnews.com/linksample.htm)				
C/(No + Io)	59.96	59.96	59.96	dB.Hz
C/(N + I)	−13.85	−13.85	−13.85	dB
Eb/(No + Io)	−13.77	−13.77	−13.77	dB
System margin	0.9	0.9	0.9	dB
Net Eb/(No + Io)	−14.67	−14.67	−14.67	dB
Required Eb/(No + Io)	5.1	5.1	5.1	dB
Excess margin	−19.77	−19.77	−19.77	dB
Earth Station Power Requirements	*Value*	*Units*		
EIRP per carrier	75.65	dBW		
Antenna gain	61.75	dBi		
Antenna feed flange power per carrier	13.9	dBW		
Uplink power control	0	dB		
HPA output back off	3	dB		
Waveguide loss	0.2	dB		
Filter truncation loss	0	dB		
Number of HPA carriers	1	–		
Total HPA power required	17.1016	dBW		
Required HPA power capability	10	W		
Spectral power density	−59.11	dBW/Hz		
Space Segment Utilization	*Value*	*Units*		
Overall link availability	99.8	%		
Information rate (inc overhead)	23.6	Mbps		
Transmit rate	40	Mbps		
Symbol rate	20	Mbaud		
Occupied bandwidth	24	MHz		
Noise bandwidth	73.8	dB.Hz		
Minimum allocated bandwidth required	26	MHz		
Allocated transponder bandwidth	26.1	MHz		
Percentage transponder bandwidth used	96.67	%		
Used transponder power	46.7	dBW		
Percentage transponder power used	100	%		
Max carriers by transponder bandwidth	1.03	–		
Max carriers by transponder power	1	–		
Max transponder carriers limited by	Power	[1.00]		
Power equivalent bandwidth usage	27	MHz		

mobile systems. However, one must examine the link budget to see where various effects are accounted for. Assumptions and confidence in models used to arrive at parameters used in the link budget calculation (such as rain fade) will affect the amount of link margin required.

Adaptive power control can be used to vary the transmitted power to overcome short-term increases of the link path loss (known as fading). When conditions are good, the transmitted power can be reduced, and when conditions deteriorate,

transmitted power levels can be increased. Some link margin is still required to accommodate dynamic changes in losses that occur faster than the adaptive feedback can respond.

Availability and signal quality are two requirements that will determine an acceptable link margin. Availability refers to the percentage of time that a satellite link can be closed during a certain time period (a year is typically used). An availability of 99.95% implies that a link is down for a total of around 4 h per year. Signal quality depends on the type of signal being transmitted. For digital signals, the signal quality parameter of interest is the bit error rate. For other signals it might be the SNR or C/N.

Conclusion

Link budgets are an accounting of the gains and losses in a communications system. The link margin is a measure of the robustness of the link. The channel capacity equation implies that a communications engineer can trade off power, bandwidth, and complexity to achieve various system requirements in a design. Establishing link budget and link margins for different types of satellite systems represents a variety of challenges. As one moves to higher frequencies and one moves from fixed to mobile satellite systems the more difficult the challenge becomes and the higher the link margin must be set to provide reliable service.

Cross-References

- ▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ▶ Appendix 3a: Global Communications Satellite Systems
- ▶ Appendix 3b: US Domestic Communications Satellite Systems
- ▶ Common Elements versus Unique Requirements in Various Types of Satellite Application Systems
- ▶ Economics and Financing of Satellite Communications
- ▶ Fixed Satellite Communications: Market Dynamics and Trends
- ▶ Geographic Information Systems and Geomatics
- ▶ History of Satellite Communications
- ▶ Introduction to Satellite Navigation Systems
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends
- ▶ Overview of the Spacecraft Bus
- ▶ Regulatory Process for Communications Satellite Frequency Allocations
- ▶ Satellite Antenna Systems Design and Implementation Around the World
- ▶ Satellite Communications and Space Telecommunications Frequencies
- ▶ Satellite Communications Antenna Concepts and Engineering
- ▶ Satellite Communications Modulation and Multiplexing
- ▶ Satellite Communications Overview
- ▶ Satellite Communications Regulatory, Legal and Trade Issues
- ▶ Satellite Communications Video Markets: Dynamics and Trends

- ▶ Satellite Earth Station Antenna Systems and System Design
- ▶ Satellite Orbits for Communications Satellites
- ▶ Satellite Radio Communications Fundamentals and Link Budgets
- ▶ Satellite Transmission, Reception and On-Board Processing Signaling and Switching
- ▶ Space Telecommunications Services and Applications
- ▶ Telemetry, Tracking, and Command (TT&C)
- ▶ Trends and Future of Satellite Communications

References

T. Ambler, B.N. Taylor, *Guide for the Use of the International System of Units (SI)*. NIST Special Publication 811, 2008 edn. (U.S. Department of Commerce, Technology Administration, National Institute of Standards and Technology, Gaithersburg, 2008), p. 30

V.K. Bhargava, D. Haccoun, R. Matyas, P.P. Nuspl, *Digital Communications by Satellite* (Wiley, New York, 1981), p. 24

R.C. Dixon, *Spread Spectrum Systems* (Wiley, New York, 1976), p. 13

R.M. Gagliardi, *Satellite Communications* (Van Nostrand Reinhold, New York, 1984), pp. 17–20

R.W. Hamming, *Coding and Information Theory*, 2nd edn. (Prentice-Hall, Englewood Cliffs, 1986), p. 2

S. Hayken, *Communications Systems*, 2nd edn. (Wiley, Chichester, 1983), p. 469

N.S. Jayant, P. Noll, *Digital Coding of Waveforms* (Prentice-Hall, Englewood Cliffs, 1984), p. 221

M.J. Miller, B. Vucetic, L. Berry (eds.), *Satellite Communications Mobile and Fixed Services* (Kluwer, Boston, 1993), p. 51

W.W. Peterson, E.J. Weldon Jr., *Error-Correcting Codes*, 2nd edn. (MIT Press, Cambridge, 1991), p. 6

Satellite Communications Modulation and Multiplexing

14

Paul T. Thompson

Contents

Introduction 326
Modulation 327
Key Aspects of Modulation 327
Digital Modulation 329
Performance of Modulation Schemes 333
Filtering 335
Modulation Summary 337
Coding 337
- Block Codes 338
- Convolutional Codes 338
- Decoding Methods 338
- Serial Concatenation of Codes 339
- Determining the Required C/N from the Required E_b/N_o and Vice Versa 340
- Symbol Rate 341
- Occupied Bandwidth 341
- Evolution of Coding 341
- Low Density Parity Check (LDPC) 342
- Turbo Codes 342

Coding Summary 343
Multiple Access 344
- Introduction 344
- Access Techniques 345
- Satellite Network Assignment Approaches 345

Further Detailed Considerations 347
- Single Channel per Carrier (SCPC) Access 347
- FDMA Access 347
- TDMA Satellite System 347
- Code Division Multiple Access 348

P.T. Thompson
University of Surrey, Guildford, Surrey, United Kingdom
e-mail: p.thompson@surrey.ac.uk

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_86, © Springer Science+Business Media New York 2013

MF-TDMA: Multiple TDMA Streams 350
DAMA with MF-TDMA 350
DVB-RCS: MF-TDMA 350
Comparison of FDMA, TDMA, and CDMA 351
Random Access Schemes 354
Selecting a Random Access Scheme 355
Multiple Access Summary 357
Notes 357
References 358
Further Reading 358

Abstract

This chapter addresses the principles involved in three key areas in satellite communications, namely, modulation (and demodulation), forward error coding, and multiple access approaches or techniques. It focuses on features and technical approaches utilized in modern-day digital satellite communications systems. Since analog approaches or techniques are today seldom used in operating satellite networks, they are not addressed in detail. Nevertheless, several references are given concerning such analog techniques should there be a historical interest in these subjects. The materials provided in this chapter are aimed at giving an appreciation of the issues involved and the performance achievable in practice. For more detailed mathematical treatments an extensive bibliography is also provided.

Key words

Aloha random access • Bit error rate (BER) • Bit rate • Block codes • Carrier to noise ratio • CDMA • Coding • Coding gain • Convolutional coding • Decoding • Demand assigned multiple access (DAMA) • Demand assignment • Demodulator • Digital video broadcast (DVB) • Encoding • Forward error correction (FEC) • Low density parity check (LDPC) coding • Medium access control (MAC) • Modem • Modulation • Multiple access techniques • Nyquist filtering • Random access • Reed–Solomon coding • Claude Shannon • Signal constellation • Spectral efficiency • Symbol rate • TDMA • Turbo coding • Viterbi algorithm or Viterbi coding • VSAT networks

Introduction

When information is to be conveyed over a satellite link it is first processed to make it suitable for impressing this information on a Radio Frequency (RF) carrier. This process is called modulation. After transmission over the satellite link the modulated RF carrier is demodulated to extract a replica of the original information. Various modulation formats are adopted in satellite communications. The choice of the modulation technique employed depends upon a range of considerations such as:

- Nature of the information to be conveyed (long streams, short bursts, etc.)
- Performance requirements

- Interference generation and tolerance
- Capability of the satellite link in terms of power and bandwidth available

To operate satellite communications systems to obtain the maximum value from the satellite resources available (power and bandwidth) it is important to exploit these resources in an efficient manner within the constraints of the system parameters. To this end, if bandwidth permits, Forward Error Correction (FEC) or coding can be employed to reduce the power requirements and therefore achieve higher capacity overall. A range of FEC approaches exist that match the specific needs of the traffic and satellite channel. These have evolved with time as the hardware developed to allow the implementation of more complex schemes. A range of these will be presented along with performance metrics and associated performance outcomes.

The third key feature addressed in this chapter is that of Multiple Access, that is, how terminals in a network communicate to each other in terms of the number of carriers that are available for utilization, their contents, and whom they are intended to reach. Again a range of approaches exist depending upon the nature of the traffic, the size of the network, and the ground terminals involved.

Modulation

Modulation is the process of impressing the wanted data on a Radio Frequency (RF) carrier which is then conveyed over the satellite link and demodulated at the receiving terminal. Thus modulation translates a baseband spectrum (in a lower frequency range) to a carrier spectrum (at a much higher frequency range).

Demodulation is the process of recovering the data at the receiver end of the link. Figure 14.1 depicts a typical transmit and receive chain with the modulation and demodulation clearly identified.

Key Aspects of Modulation

Modulation is the name given to the process of impressing the signal to be transported onto a carrier. Demodulation is the name given to extracting the wanted signal from the carrier. Thus, the process requires a modulator and a demodulator, collectively known as a modem. The input to the modulator may require some initial processing such as filtering and amplitude limiting. Digital input signals may also be processed to provide a balance signal around zero and to spread the bandwidth of the signal (with a scrambler).

Modulation of a signal onto a carrier results in dynamic changes to one or more of the following:

- The amplitude of the carrier
- The phase of the carrier
- The frequency of the carrier

Thus a sinusoidal carrier can be modulated in several ways as shown in Fig. 14.2.

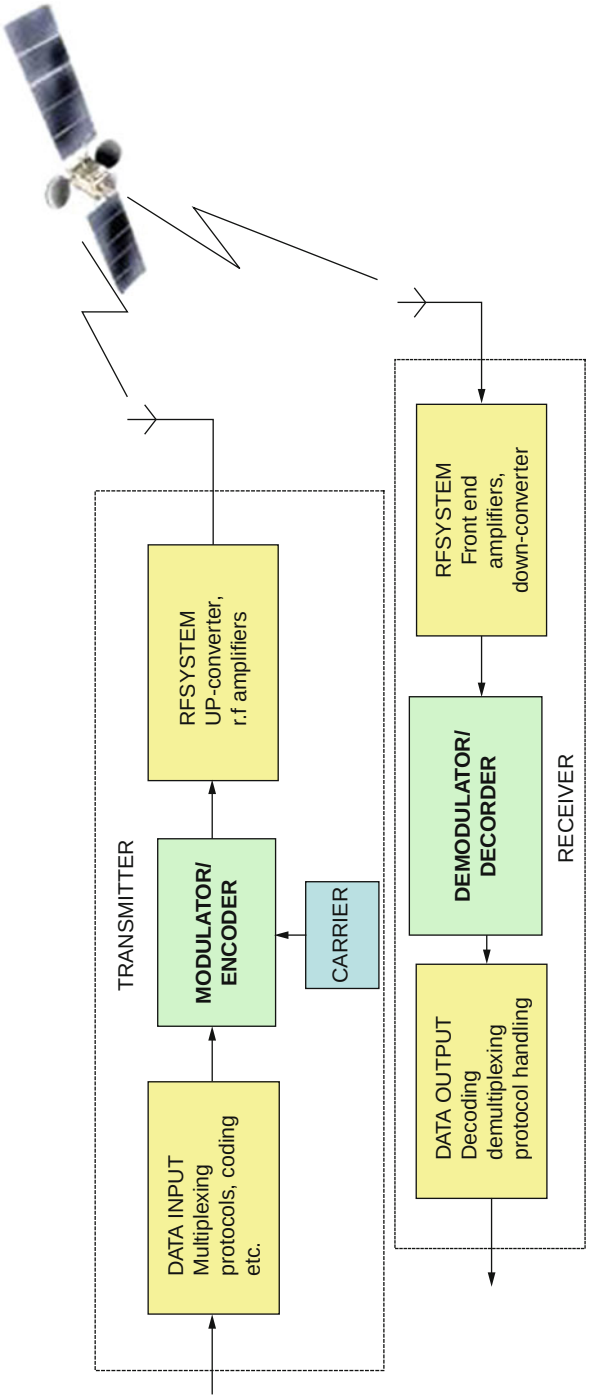


Fig. 14.1 Typical transmit and receive chain

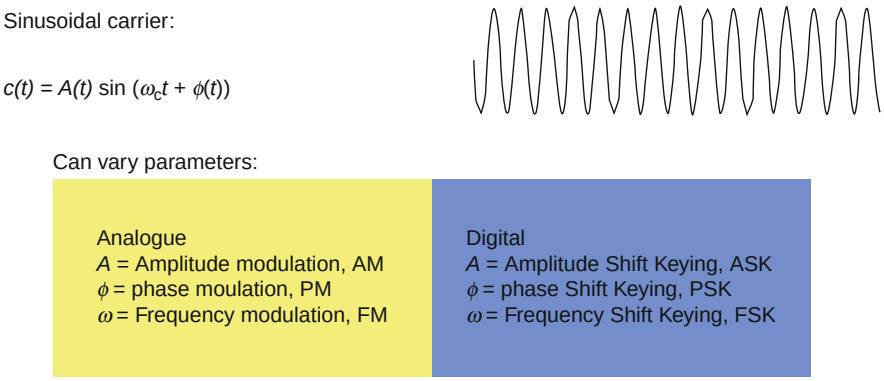


Fig. 14.2 Modulation of a sinusoidal carrier

The initial carrier is usually a CW signal with no perceptible bandwidth. The modulated carrier will have an associated bandwidth that is a factor of the bandwidth of the data signal. Two bandwidths are often considered in practice, namely, the Occupied Bandwidth and the Allocated Bandwidth. The Occupied Bandwidth is associated with the noise bandwidth in which the signal is being detected and demodulated. It is often expressed as the -10 dB bandwidth. The Allocated Bandwidth is that bandwidth which is allocated by the satellite operator and includes some extra bandwidth to avoid excessive adjacent channel interference. This bandwidth is often set at 1.4 times the symbol rate but can vary based on a number of transmission efficiency factors.

In this chapter we address the digital modulation cases.

Digital Modulation

The modulation of a carrier with a digital signal may take one of many forms as indicated below.

- Amplitude Shift Keying – ASK
- Frequency Shift Keying – FSK
- Phase Shift Keying – PSK
- Bipolar Phase Shift Keying – BPSK
- Quadrature Phase Shift Keying – QPSK
- M-ary Phase Shift Keying – MPSK
- Asymmetric Phase Shift Keying – APSK
- Minimum Shift Keying – MSK
- Quadrature Amplitude Modulation – QAM

By far the most popular is QPSK which has been employed in satellite systems for many decades while BPSK has been adopted for critical cases where interference issues are highly important. There are systems, for instance, that can drop from high-efficiency 8 Phase Shift Keying to QPSK or even BPSK modulation

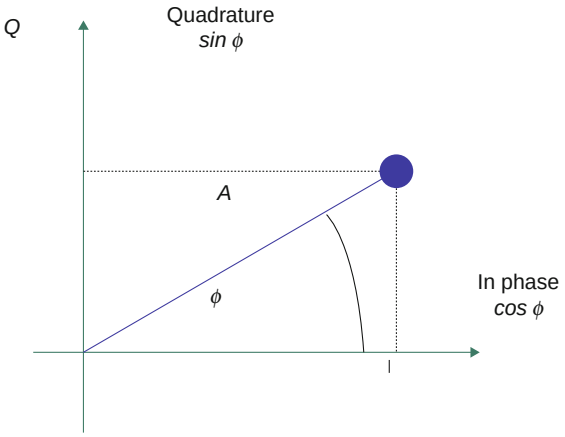


Fig. 14.3 Phasor representation of I and Q components

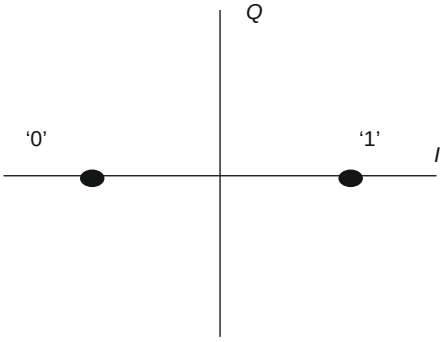


Fig. 14.4 BPSK constellation diagram

depending on the degree of interference or precipitation attenuation that might be experienced at the ground antenna system due to heavy rainfall – particularly when much higher microwave or millimeter wave frequencies are being utilized.

A sinusoid of any phase can be regarded as a combination of a cosine wave *in phase* component and a sine wave *quadrature* component which can be plotted on a phasor diagram as shown in Fig. 14.3. This so-called I/Q quadrature approach is very popular in digital modem implementations.

BPSK has just two-phase states, one representing a digital zero and the other representing a digital one. This can be depicted on a phasor diagram and is known as a *constellation diagram*. The receiver has to determine a digital zero or one by examining the phase of the carrier. The decision boundary can be plotted on the constellation diagram (for BPSK this is the "Q axis" as shown in Fig. 14.4).

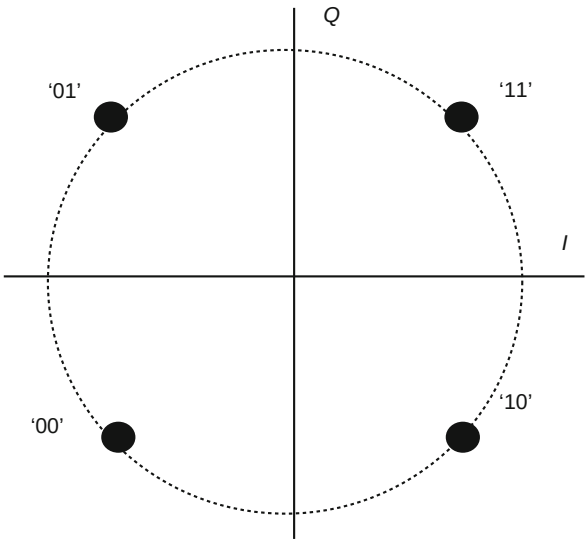


Fig. 14.5 QPSK constellation diagram

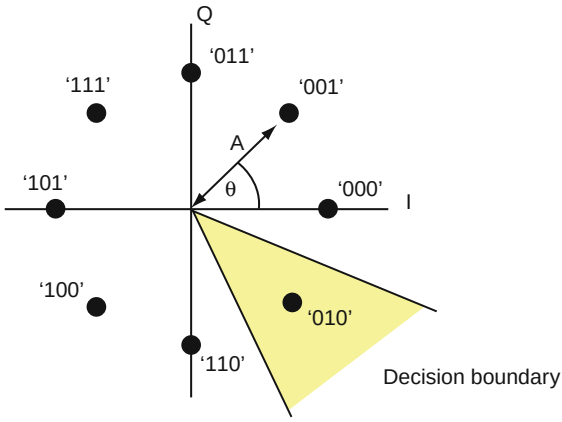


Fig. 14.6 8-PSK constellation diagram

With QPSK four-phase states represent the data. These are established with *symbols* where a symbol in this case represents 2 bits, for example, "01." The decision boundaries in this case are both the I and the Q axes as shown in Fig. 14.5. It is clear from this figure that there is more scope to misinterpret the symbols and hence QPSK will need more immunity from noise than BPSK. This is normally achieved by operating at a higher carrier to noise level (i.e., typically 3 dB higher).

For 8 PSK a symbol represents 3 bits as depicted in Fig. 14.6.

The DVB-S2 standard[1] has two special higher order modulation formats that have been developed to operate in a more optimum manner in the presence of a nonlinear

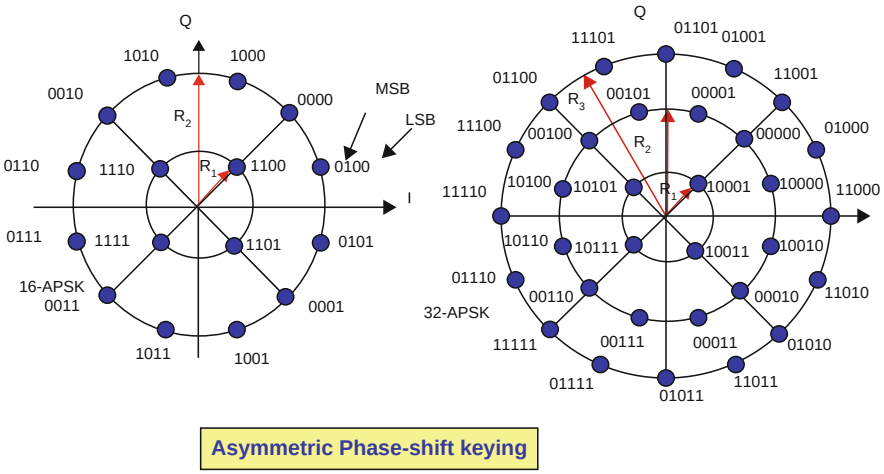


Fig. 14.7 Constellation diagrams for 16 APSK and 32 APSK

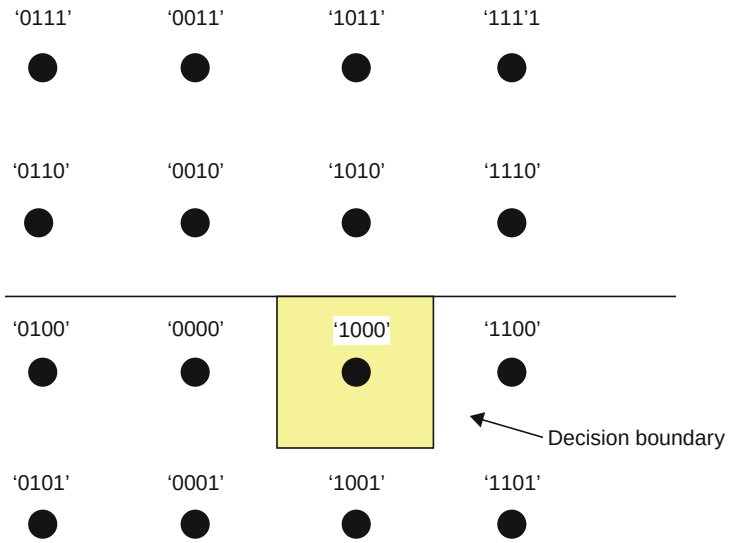


Fig. 14.8 16-QAM constellation diagram

channel (caused by operating high-power amplifiers close to saturation). These are 16 APSK and 32 APSK. Their respective constellation diagrams are given in Fig. 14.7.

An alternative to APSK is Quadrature Amplitude Modulation (QAM). Typically QAM has a rectangular constellation diagram structure as depicted in Fig. 14.8 where QAM is shown.

Consideration is now given to the issue of how to describe the performance of these different modulation formats and how to match them to communications needs and circumstances.

Performance of Modulation Schemes

The design of a satellite link is based upon the level of performance required. The performance is usually described in terms of a Bit Error Rate (BER) at which acceptable service is possible such as one error occurring in 1 million bits (often indicated as a BER probability of 1 in 10^{-6}). The characteristics of the link are then matched to this BER in terms of the Carrier to Noise (C/N) level required to achieve this performance. An alternative to C/N is E_b/N_o, which is a measure of the energy per bit to noise power spectral density ratio (the noise in 1 Hz bandwidth). The two are related as indicated below.

$$C/N = E_b/N_o \cdot \frac{f_b}{B}$$

where

f_b is the channel data rate, and B is the channel bandwidth.

In logarithmic form (dB) the relationship is:

$$C/N_{\mathrm{dB}} = 10\log_{10}(E_b/N_o) + 10\log_{10}\left(\frac{f_b}{B}\right)\mathrm{dB}$$

If E_b/N_o is expressed in dB then the above equation becomes:

$$C/N_{\mathrm{dB}} = E_b/N_{o\mathrm{dB}} + 10\log_{10}\left(\frac{f_b}{B}\right)\mathrm{dB}$$

The above analysis assumes that the noise environment of the satellite link can be described as Additive White Gaussian Noise (AWGN) that is a satellite link in which the only impairment to communication is noise with a constant spectral density and a Gaussian distribution of amplitude. This approach does not take account of fading, frequency selectivity, interference, nonlinearity, or dispersion. However, it produces an adequately simple model which is useful for gaining insight into the underlying behavior of a system.

E_b/N_o can be related to the energy per symbol per noise power spectral density (E_s/N_o) by:

$$E_b/N_o = E_s/N_o \cdot \frac{1}{\rho}$$

where E_s is the energy per symbol in joules and ρ is the nominal spectral efficiency. Spectral efficiency is defined as the number of bits per second that can be transmitted within 1 Hz. For a number of years a typical spectral efficiency in satellite communications was 1 bit/s per Hz, but today efficiencies are more typically

Table 14.1 DVB-S2 performance at the QEF threshold

DVB-S2 MOD	FEC rate	Spectral efficiency (bps/Hz)	Ideal E_s/N_o dB	IDEAL E_b/N_o dB	Including modem margin E_b/N_o dB	M
QPSK	1/4	0.490243	−2.35	0.75	1.95	4
QPSK	1/3	0.656448	−1.24	0.59	1.79	4
QPSK	2/5	0.789412	−0.30	0.73	1.93	4
QPSK	1/2	0.988858	1.00	1.05	2.25	4
QPSK	3/5	1.188304	2.23	1.48	2.68	4
QPSK	2/3	1.322253	3.10	1.89	3.09	4
QPSK	3/4	1.487473	4.03	2.31	3.51	4
QPSK	4/5	1.587196	4.68	2.67	3.87	4
QPSK	5/6	1.654663	5.18	2.99	4.19	4
QPSK	8/9	1.766451	6.20	3.73	4.93	4
QPSK	9/10	1.788612	6.42	3.89	5.09	4
8PSK	3/5	1.779991	5.50	3.00	4.20	8
8PSK	2/3	1.980636	6.62	3.65	4.85	8
8PSK	3/4	2.228124	7.91	4.43	5.63	8
8PSK	5/6	2.478562	9.35	5.41	6.61	8
8PSK	8/9	2.646012	10.69	6.46	7.66	8
8PSK	9/10	2.679207	10.98	6.70	7.90	8
16APSK	2/3	2.637201	8.97	4.76	5.96	16
16APSK	3/4	2.966728	10.21	5.49	6.69	16
16APSK	4/5	3.165623	11.03	6.03	7.23	16
16APSK	5/6	3.300184	11.61	6.42	7.62	16
16APSK	8/9	3.523143	12.89	7.42	8.62	16
16APSK	9/10	3.567342	13.13	7.61	8.81	16
32APSK	3/4	3.703295	12.73	7.04	8.24	32
32APSK	4/5	3.951571	13.64	7.67	8.87	32
32APSK	5/6	4.119540	14.28	8.13	9.33	32
32APSK	8/9	4.397854	15.69	9.26	10.46	32
32APSK	9/10	4.453027	16.05	9.56	10.76	32

in the range of 2.5 bits/s per Hz and as can be seen in Table 14.1 could range as high as 4.5 bits/s per Hz with 32 APSK digital modulation schemes and with a very low level of noise and interference. E_s/N_0 is also commonly used in the analysis of digital modulation schemes.

Alternatively they are related by:

$$E_s/N_o = E_b/N_o \,.\, \log_2(M)$$

where M is the number of modulation symbols or phase states.

The lower the operating E_b/N_o the lesser the power that is required from the transmitters or alternatively the more noise or interference that can be tolerated. Figure 14.9 shows such an effect for typical un-coded modulation.

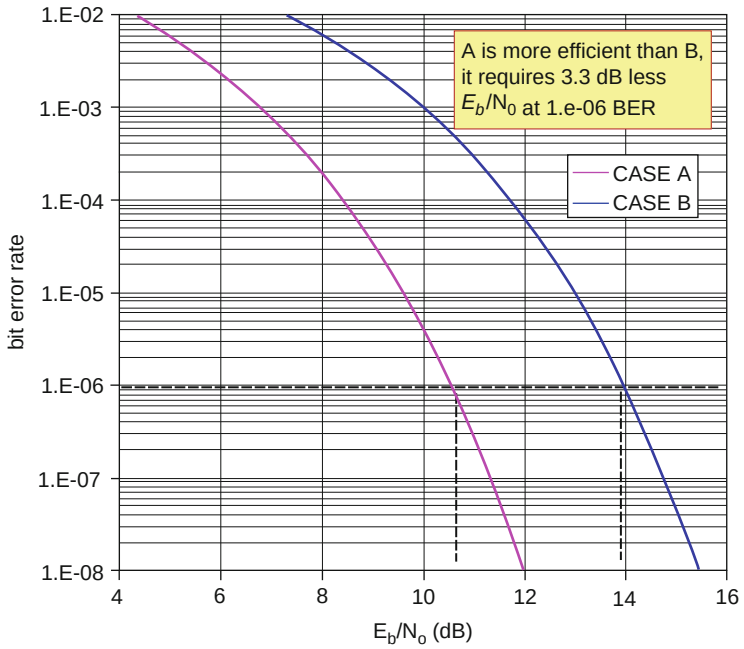


Fig. 14.9 Typical BER vs. E_b/N_o curves

The following graph (Fig. 14.10) indicates the performance of a range of modulations applicable to satellite links. The derivation of these curves is not presented here but if further details are required the reader is referred to Chap. 7 of Pahlavan and Levesque (Pahlavan and Levesque 1995) and Proakis, Salehi, and Bauch (Proakis et al. 2004).

Filtering

In practice, there is a requirement to filter the signal at the modulator and the receiver. In general, the receiver will include a "Matched filter" prior to the bit/symbol decision making process. A matched filter maximizes output signal-to-noise ratio for a particular symbol shape. The so-called Nyquist Filters are usually adopted to constrain the spectrum to a sensible bandwidth. Filtering may introduce inter-symbol interference (ISI), Nyquist filtering avoids this as much as possible.

Nyquist filters have zeros in their impulse response at multiples of Ts and are usually Raised Cosine Filters. The frequency response of such a filter is given by the following formulas:

$$H(\omega) = 1 \; if \; 0 \; < \; \omega \; < \; \frac{\pi \; (1 - \alpha)}{T_s}$$

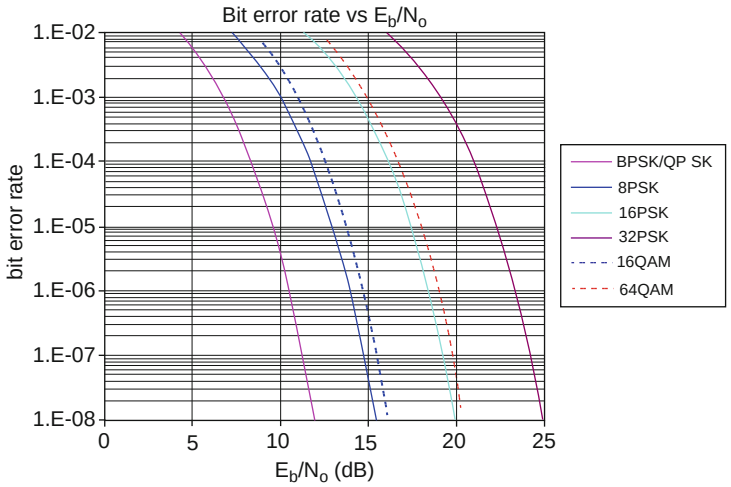


Fig. 14.10 BER vs E_b/N_o for a range of modulations appropriate to satellite links

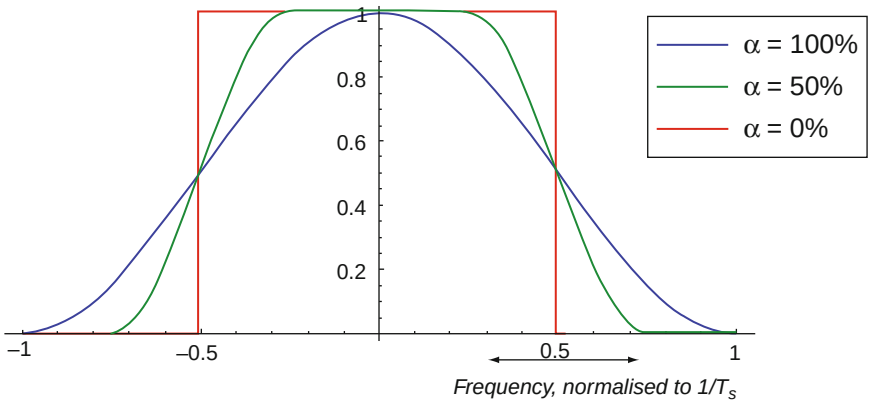


Fig. 14.11 Typical Nyquist filter shape

$$H(\omega) = 0 \; if \quad \frac{\pi\,(1+\alpha)}{T_s} \; < \; \omega$$

$$H(\omega) = \cos^2\left(\frac{T_s}{4\,\alpha}\left[\omega - \frac{\pi\,(1-\alpha)}{T_s}\right]\right) \quad if \quad \frac{\pi\,(1-\alpha)}{T_s} \; < \; \omega \; < \; \frac{\pi\,(1+\alpha)}{T_s}$$

where α is the roll-off factor which defines the filter sharpness and hence the bandwidth (Fig. 14.11).

One of the problems of filtering is the fact that the signal may no longer be a constant envelope and therefore, more susceptible to any nonlinearity in the link.

By nonlinearity we mean variations in amplitude or phase introduced by the operation of an amplifier close to its maximum output such that they do not change in a linear manner.

Modulation Summary

- Modulation provides the mechanism of carrying the required data on a RF signal.
- A bit error rate needs to be chosen as it determines resource requirements and quality of performance offered.
- The main parameters of modulation schemes are bandwidth and power efficiency for a chosen E_b/N_o (or C/N) along with appropriate filtering.
- Performance should be determined in an appropriate noise environment. Additive White Gaussian Noise (AWGN) is often used to describe such noise.
- Multi-level modulation can increase bandwidth efficiency at the cost of power efficiency.

Coding

Communications links are susceptible to noise and to achieve a high C/N (or E_b/N_o) implies a high signal power. In satellite communications, the RF power from the satellite is a key and valuable resource, consequently the power is kept to a minimum while at the same time achieving the required performance in terms of C/N.

When power is at a premium it is possible to trade power for bandwidth, provided the extra bandwidth is available. The tool for doing this trade is Forward Error Control (FEC) or "coding" as it is more commonly known. FEC operates by transmitting some redundant information in additional bits that can be used at the receiving end of the link to detect and correct errors. The more the redundant bits the better the ability to correct errors but at the price of increased bandwidth as more bits are now being sent.

A measure of the improvement obtained with FEC is the so-called coding gain. This is the difference in required C/N for the coded case compared to the un-coded case at a specified BER. An example is given in Fig. 14.12.

Error correction codes can provide error-free communication through noisy channels provided the information rate does not exceed a specific limit, that is, the channel capacity as detailed by Shannon (the so-called Shannon Limit). The Shannon Limit defines the minimum E_b/N_o that can, in theory, provide error-free transmission for a given code rate in an AWGN channel, and the theory is based on random code structure and suggests that longer code blocks achieve better performance. However, long and random structure codes are difficult to decode and practical techniques use an organized code structure to enable feasible decoding. Thus, the Shannon limit is difficult to achieve in practice, but improvements in digital technology have progressively made it easier to develop advanced codes with performance approaching the Shannon limit.

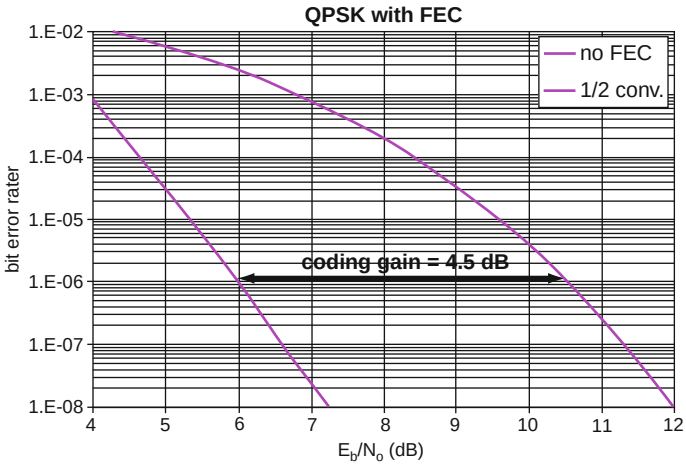


Fig. 14.12 Example of coding gain

Error correcting codes map a k-bit long data-word to an n-bit long code word (where $k < n$), and $r = k/n$ is the FEC code rate.

In practice two types of conventional FEC code approaches were adopted: the Block Codes and the Convolutional Codes. Both have advantages and disadvantages.

Block Codes

Block codes map directly a k-bit long data-word to a unique n-bit long code word. For such codes we find that the decoder complexity/memory grows exponentially with n. An important set of codes is known as the BCH cyclic block codes, which were invented by Hocquenghem and independently by Bose and Ray-Chaudhuri. The Reed–Solomon (RS) set of codes are a special set of BCH codes and are popular for outer coding.

Convolutional Codes

Convolutional codes have the advantage that optimal decoding complexity grows only linearly with the code word length, but exponentially with the encoder's memory order. In a Convolutional encoder output bits are determined by the present input and previous input bits. The Viterbi algorithm is a particularly efficient tool for decoding the received signal and outputs the most likely data-word.

Decoding Methods

There are several types of decoding methods applied to the demodulator output prior to the decoder.

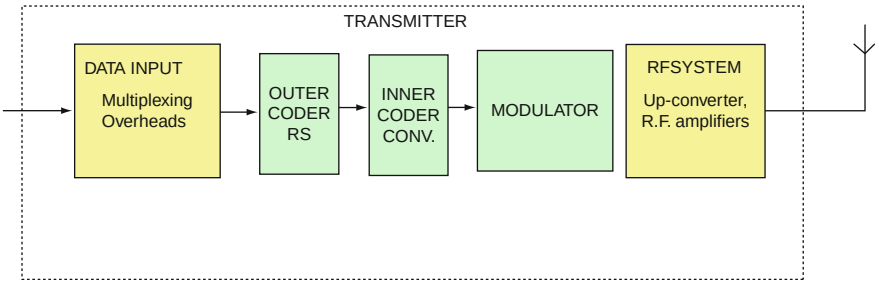


Fig. 14.13 Concatenated FEC coding

Hard Decoding

- Makes symbol decisions directly from the received signal and maps the bit decisions as input to the decoder.

Soft Decoding

- Determines symbol reliability values from the received signal using statistical approaches and maps these values as input to the decoder.
- Offers better performance but with more computations.

Serial Concatenation of Codes

There are significant advantages in adding two codes together in series, providing the additional computational complexity can be accepted. The two codes are normally called the inner code and the outer code (see Fig. 14.13). A good example employs convolutional coding (with Viterbi decoding for the inner code and a Reed–Solomon (RS) block code for the outer code). Such implementation exists for the INTELSAT IBS/IDR[2] services as well as the Digital Video Broadcast (DVB-S) standard[3]. Such an approach is useful to manage issues that arise from the fact that errors may be bunched together and therefore, adversely impact convolutional codes used alone.

A two-stage concatenated coding scheme is usually used in the following manner:

- Inner code with soft decision decoding, targeted at random errors
- Outer code with hard decision decoding, targeted at bursty errors
- Outer code corrects inner decoding errors
- Useful for achieving very low BER (e.g., 10^{-11})

Concatenated "Reed–Solomon + Convolutional" coding was the state-of-the-art solution prior to the practical implementation of Turbo codes in 1993.

Figure 14.14 presents a typical set of performance data for QPSK operation using a convolutional encoder with and without concatenated Reed–Solomon outer coding. It should be noted that typical performance is presented here. For more specific information one can refer to manufacturers' data sheets.

When FEC is employed the calculation of the various factors required in link performance assessment can be performed using the following equations.

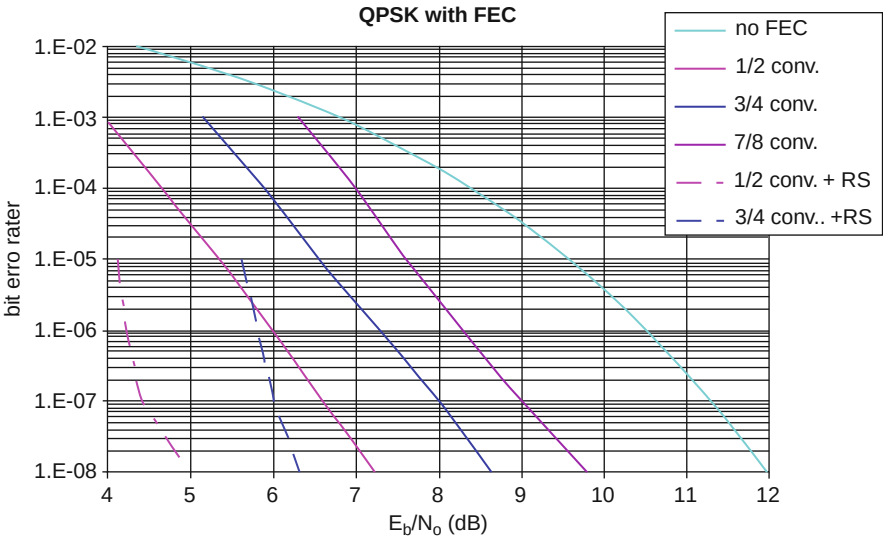


Fig. 14.14 Typical convolutional FEC and concatenated convolutional with Reed–Solomon performance

Determining the Required *C/N* from the Required E_b/N_o and Vice Versa

$$C/N = \frac{E_b}{N_o} - 10\log_{10}(1+\alpha) + 10\log_{10}(m) + 10\log_{10}(r)\ \text{dB}$$

$$\frac{E_b}{N_o} = C/N + 10\log_{10}(1+\alpha) - 10\log_{10}(m) - 10\log_{10}(r)\ \text{dB}$$

where:

$\alpha =$ The filter roll-off factor (0 to 1)

$r =$ The overall FEC rate: for example, for DVB concatenated Reed–Solomon and Viterbi ¾

$$r = 3/4 \times 188/204 = 0.6911$$

$m = \log_2(M)$ where M is the number of phase states

	M	m
BPSK	2	1
QPSK	4	2
8PSK	8	3
16APSK	16	4
32APSK	32	5

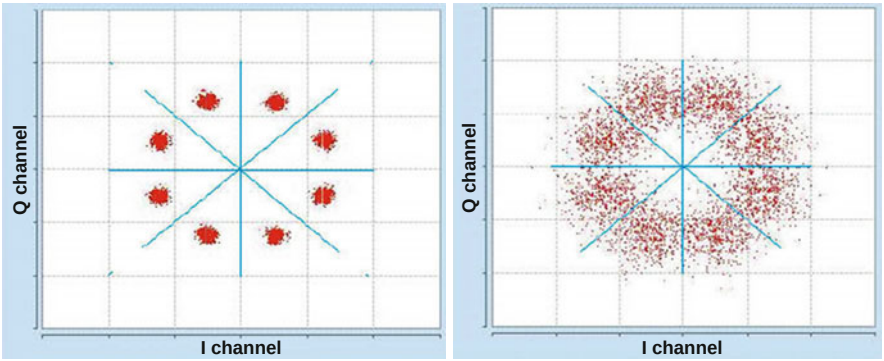


Fig. 14.15 Performance of a 8 PSK channel with additive white Gaussian noise (AWGN) and convolutional rate 2/3 FEC. 8 PSK FEC = 2/3: $E_b/N_o > 15$ dB (left), 8 PSK FEC = 2/3 $E_b/N_o = 6.8$ dB (right)

Symbol Rate

$$\text{Symbol rate} = \frac{\text{Information bit rate}}{m \times r}$$

Occupied Bandwidth

Occupied band width (B) = Symbol rate $\times$ $(1 + \alpha)$

ρ is the nominal spectral efficiency in (bit/s)/Hz.

Performance of a link may be visually observed using a plot of the received constellation diagram. Figure 14.15 depicts two such constellation diagrams with appropriate decision boundaries. The signal used in this example is an 8 PSK signal with rate 2/3 Forward Error Correction. When the E_b/N_o is high (left hand plot) it is clear that very few data points cross into the incorrect decision region. However, when the E_b/N_o is low (right hand plot) this is clearly not the case and it would be tempting to observe that many errors are occurring. This is indeed possible but the FEC takes care of these, recognizing that the constellation diagram is measured before the FEC takes effect (Fig. 14.15).

Evolution of Coding

In 1993, it was acknowledged that Turbo coding was the first coding technique that can practically approach closely to the Shannon limit. Then in l997, it was realized that Low Density Parity Check (LDPC) codes can also practically approach closely to the Shannon limit. Indeed it transpired that LDPC codes hadbeen first discovered by Prof. Robert Gray Gallager of MIT in the early 1960s, but ignored by the coding research community as the implementation was not feasible at that time!

Turbo and LDPC codes use iterative soft decoding to enhance performance as compared to conventional FEC.

Low Density Parity Check (LDPC)

Defining features of LDPC codes are:

- Linear block codes
- Sparse (low density) parity check matrix (more zeros than ones)
- Any two columns of the parity check matrix H have not more than one nonzero entry in common
- Performance close to Shannon limit

The sparseness of the parity check matrix of an LDPC code helps to guarantee a low decoding complexity.

LDPC is most effective for very large blocks and is employed in the well known DVB-S2 standard with block lengths of the order of $n = 64{,}800$ (normal block) and $n = 16{,}800$ (short block). The long-code construction is consistent with Shannon's theorem which suggests that the performance of a code improves with its length. The use also requires the service to have a reasonable degree of "delay tolerance" as the processing of such large blocks may take some time.

Thus, large code blocks are very compatible with "delay-tolerant" broadcasting such as in DVB-S2 and LDPC has lower decoding complexity than turbo codes for very long code blocks, making it possible to approach closely to the Shannon limit.

The DVB-S2 standard[1,4] uses irregular LDPC codes:

- Two code word lengths: $n = 64{,}800$ (normal block) $n = 16{,}800$ (short block)
- Ten code rates: 1/4, 1/3, 2/5, 1/2, 3/5, 2/3, 3/4, 4/5, 5/6, 8/9, and 9/10
- Helps to support variable/adaptive coding and modulation

Figure 14.16 indicates the performance of the DVB-S2 LDPC. There are several points to be aware of when considering this figure. First is the fact that it uses packet error rate rather than bit error rate and secondly it is presented in terms of E_s/N_o.

A feature of such powerful codes is the fact that the range spanning excellent performance to failed performance spans a few tenths of a dB. Hence, it is more convenient to specify the performance in a table for a single error rate.

For the DVB-S2 standard a Quasi-Error-Free (QEF) performance is defined as "less than one uncorrected error event per transmission hour at the level of a 5 Mbit/s single TV service decoder," approximately corresponding to a Transport Stream Packet Error Ratio PER $<10^{-7}$ before de-multiplexer.

Table 14.1 indicates an example of the range of parameters relating to DVB-S2 with a block length of 64,800 bits.

Turbo Codes

The Turbo encoder is constructed by the parallel concatenation of Recursive Systematic Convolutional (RSC) constituent encoders. Recursive encoders are a key component of Turbo codes. They enable a "code interleaver-gain" factor. Performance improves with code interleaver length.

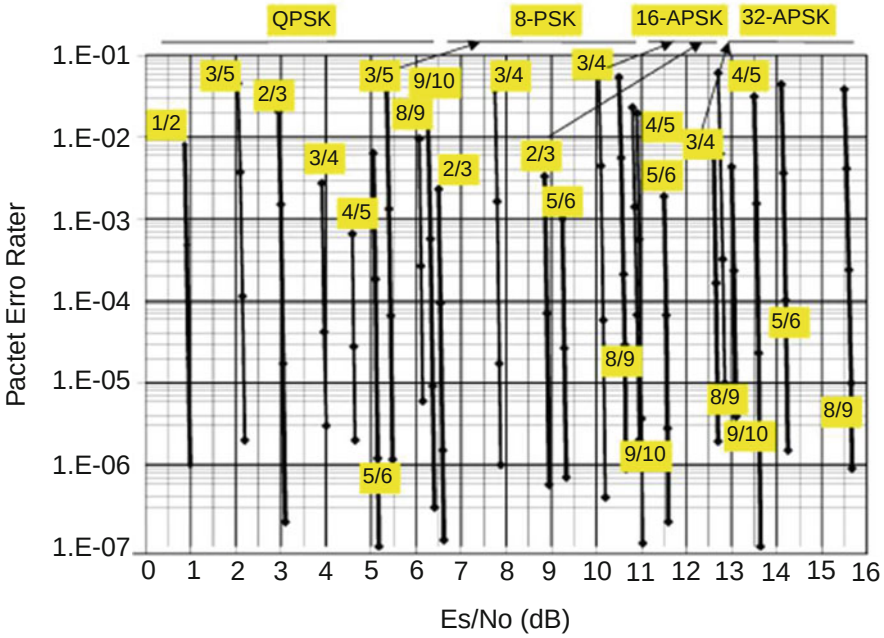


Fig. 14.16 Performance of LDPC + BCH codes over AWGN channel, N = 64,800 bits

Turbo decoding relies on constituent soft-output Trellis decoders, interconnected in a "Turbo" fashion with several iterations being performed to progressively improve the decision making process. The error performance improves with the number of turbo-loop iterations for which typically eight iterations are adequate. Parallel concatenated turbo codes typically suffer from an error-floor at BER $< 10^{-5}$, however, serial concatenated turbo codes do not suffer from this problem.

Turbo coding is used in the DVB-RCS[5,6,7] and DVB-SH standards as they tend to outperform LDPC for short code lengths.

Typical DVB-RCS turbo coded performance is given in per packet error rate or PER $= 10^{-7}$.

Table 14.2 presents an example of the E_b/N_o characteristics of the Turbo coded DVB-RCS return link.

Coding Summary

Error Control Coding mitigates the effect of channel noise under the right conditions. Forward Error Control (FEC) coding is indispensable as satellite links are very often power-limited and delay-intensive, and therefore, FEC involves a trade-off between power and bandwidth.

Table 14.2 DVB-RCS Turbo coded performance for PER $= 10^{-7}$

FEC	E_b/N_o (188 byte packets) dB	E_b/N_o (53 byte packets) dB
1/3	2.5	2.9
2/5	2.7	3.1
1/2	3.2	3.6
2/3	4.0	4.6
3/4	4.6	5.4
4/5	5.3	6.3
6/7	6.0	7.0

The Shannon Limit defines the optimum performance but computational complexity/latency determines the achievable FEC performance that can actually be achieved.

Concatenation of Convolutional encoders with Reed–Solomon block codes gives a robust implementation with reasonable performance.

LDPC and Turbo Codes are state-of-the-art FEC techniques. These techniques approach the Shannon limit in AWGN channels with practical implementation complexity. They, in fact, have been implemented in many recent satellite communications standards such as the latest DVB specifications.

Multiple Access

Introduction

Multiple Access is the mechanism by which terminals in a satellite network communicate to each other. Figure 14.17 indicates the value of using some form of multiple access. In the upper part of the figure each station sends a dedicated carrier to the stations it wishes to communicate with. This requires $N \times (N-1) = 6$ links in the case shown. In the lower part of the figure the case is depicted where a station combines traffic to all terminals (by multiplexing) such that it transmits only one carrier, the receiving station selecting the traffic destined for it. Such a configuration requires N links. If there were 100 terminals in the network then the saving in number of carriers is very significant. There are many approaches to facilitate this multiple access approach and they are the subject of this section.

As the underlying telecommunications networks evolved to digital operation and the user networks transitioned from simple circuit switched approaches (where a lot of the access functions were handled in the circuit switch) to packet transmission there was a requirement to address multiple access in much greater detail. Hence a requirement was established for some form of Medium Access Control (MAC) that covered the key aspects of multiple access in a digital era.

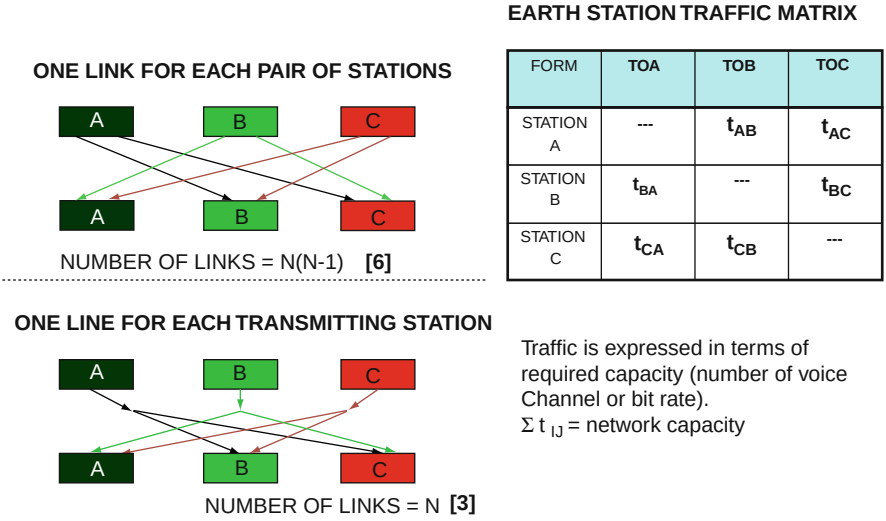


Fig. 14.17 The value of multiple access

MAC protocols can be classified in several dimensions:
Static vs. dynamic
Distributed vs. centralized
Synchronous vs. asynchronous
Popular MAC protocol classifications are:
Fixed assignment
Demand Assigned Multiple Access (DAMA) (also called contention-less access or reservation access or scheduled access)
Random access (also called contention-based access)

Access Techniques

SCPC – Single channel per carrier
FDMA – Frequency division multiple access
TDMA – Time division multiple access
CDMA – Code division multiple access
MF-TDMA – Multi-frequency time division multiple access
RA – Random access techniques for packet mode operation

Satellite Network Assignment Approaches

Basic Multiple Access approaches are depicted in the following figure (Fig. 14.18).

Fixed assignment: The frequency – time resource (i.e., RF channel) is shared between earth stations according to a scheme that does not vary with time.

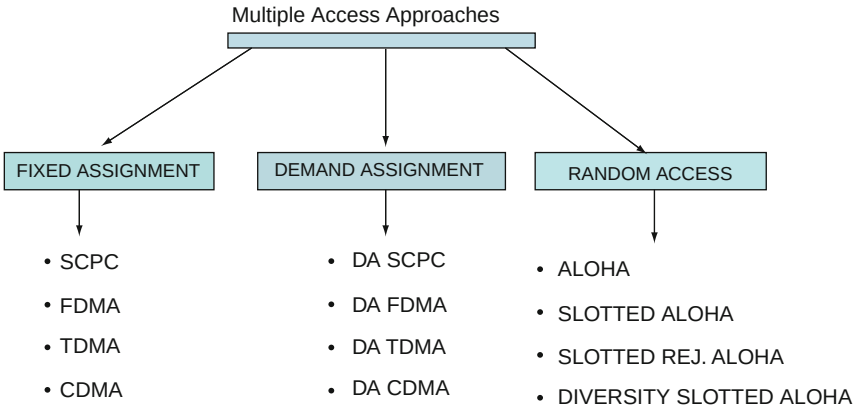


Fig. 14.18 Multiple access approaches

Demand assignment: The frequency – time resource is shared between earth stations according to the demand from individual stations. If a terminal has no traffic demand at a particular instance in time then no capacity is made available to it and others may be assigned to it (from a pool of resources). Such an arrangement is called Demand Assigned Multiple Access (DAMA) (or bandwidth on demand). DAMA is often associated with SCPC or TDMA including MF-TDMA.

Random access: In random access the stations or terminals randomly attempt to transmit a data packet. It is possible that two or more such random transmissions collide at the satellite resulting in failure to correctly receive them. This is handled by a process of re-transmitting corrupted transmissions until the message gets through correctly. This can result in quite long delays.

Some random access protocols are:

- ALOHA, the basic contention-based access scheme
- Slotted ALOHA
- Selective reject ALOHA
- DSA: Diversity Slotted ALOHA

ALOHA is the basic contention-based access scheme described above.

Slotted ALOHA is similar but transmissions are restricted to specific time slots. This has twice the throughput capability of simple ALOHA before the system becomes unstable.

In ALOHA and slotted ALOHA occasional collision among packets occurs, and on the average a packet may have to be transmitted more than once before it is received correctly. This will introduce large packet delay in satellite slotted ALOHA systems, where each round trip propagation delay is about 270 ms.

In selective reject ALOHA larger packets are subdivided into sub-packets and only damaged sub-packets are retransmitted.

DSA is a slotted ALOHA scheme in which, whenever a user generates a packet, it transmits k copies of the same packet. It is assumed that there exists some arrangement which allows a receiver to reject all but one correctly received copy of any packet. As the system may get many packets through without having to signal a retransmit event it has good throughput with lower average delays.

Further Detailed Considerations

Single Channel per Carrier (SCPC) Access

The concept of SCPC is shown in Fig. 14.19.

FDMA Access

The concept of FDMA is shown in Fig. 14.20.

Advantages

- Use of existing hardware to a greater extent than other techniques
- Network timing not required

Disadvantages

- As the number of accesses increases, intermodulation noise reduces the usable repeater output power (i.e., TWTA back-off). Hence, there is a loss of capacity relative to single carrier/transponder capacity
- The frequency allocation may be difficult to modify
- Uplink power coordination is most likely to be required

If demand assigned Access is adopted then the resource allocated is the frequency band to use.

TDMA Satellite System

In a TDMA system, each earth station transmits traffic bursts, synchronized so that they occupy ASSIGNED NON-OVERLAPPING time slots. Time slots are organized within a periodic structure called TIME FRAME.

The concept is depicted in Fig. 14.21.

Advantages

- Digital signaling provides easy interfacing with developing digital networks on the ground
- Digital circuitry has decreasing cost
- Higher throughput compared to FDMA when the number of accesses is large

Disadvantages

- Stations transmit high bit rate bursts, requiring large peak power
- Network control is required
  - Generation and distribution of burst time plans to all traffic stations
  - Protocols to establish how stations enter the network

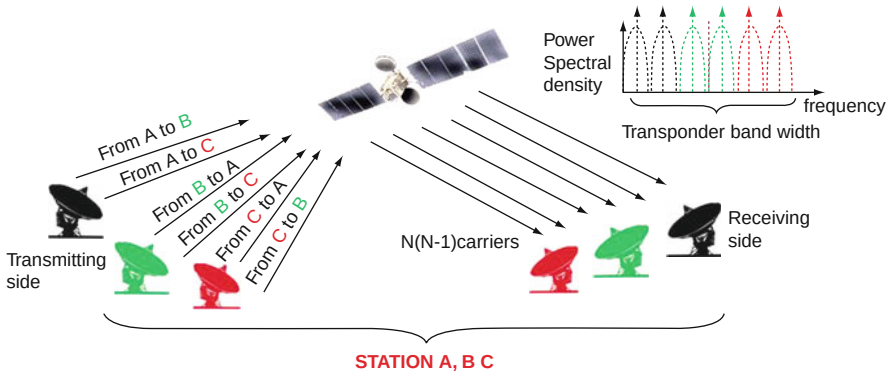


Fig. 14.19 The SCPC satellite access concept

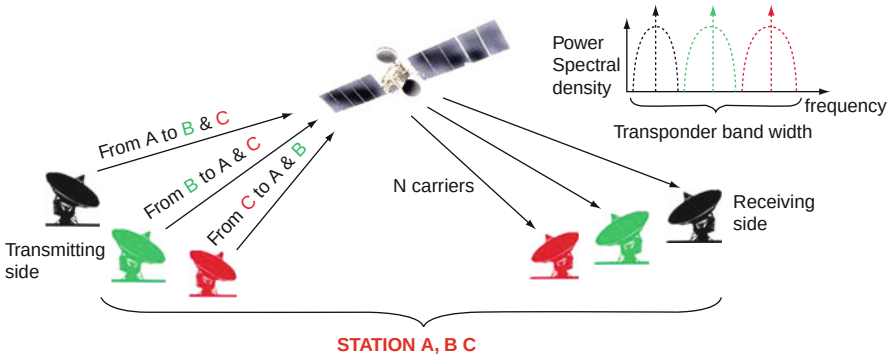


Fig. 14.20 The FDMA satellite access concept

- Provision of redundant reference stations with automatic switchover to control the traffic stations
- Means for monitoring the network are needed

Fixed and demand assigned FDMA and TDMA are shown in an alternative manner in Figs. 14.22 and 14.23 respectively.

Code Division Multiple Access

Such multiple access concepts adopt the principle of spread spectrum whereby the transmitter spreads baseband signal from bandwidth N to W by impressing the lower speed data on a much higher speed address code sequence.

W/N is the spreading factor (100 to 1,000,000), often called coding gain. The receiver de-spreads the received signal by "knowing" the proper address code. Any signal not generated with that code is not de-spread and remains as

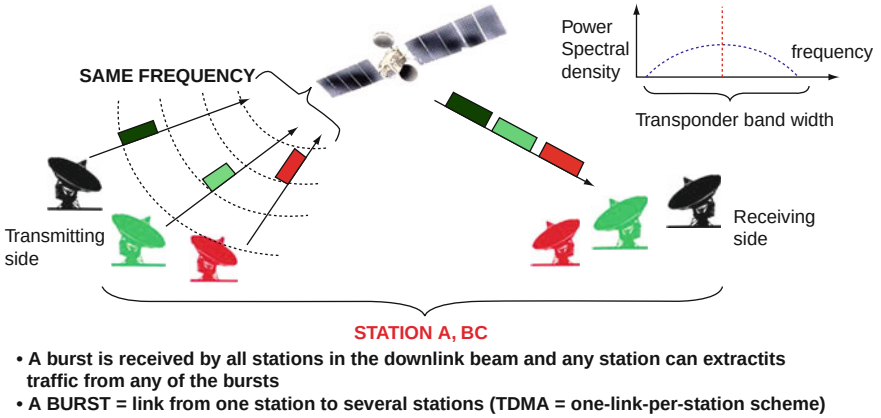


Fig. 14.21 The TDMA satellite access concept

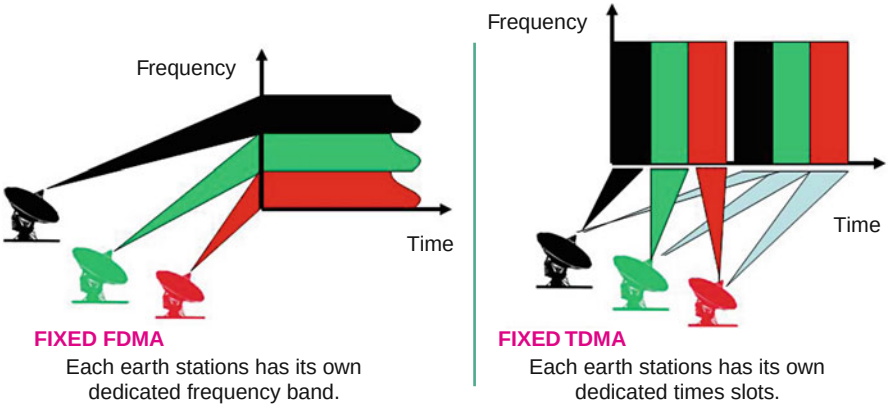


Fig. 14.22 Alternative depiction of FDMA and TDMA

wideband noise (that slightly increases the noise that the terminal has to tolerate). Thus, received signals with other address codes and jammers are spread by receiver and act as noise. Addresses are periodic binary sequences that either modulate the carrier directly (DIRECT SEQUENCE SYSTEMS) or change the frequency state of the carrier (FREQUENCY HOPPING SYSTEMS). The multiple access process is assured by operating the carriers on the same RF frequency but employing different codes for differing links. The codes are specially selected to have high de-correlation properties and are thus orthogonal codes. The concept of spreading and de-spreading is shown in Fig. 14.24 and that of CDMA in Fig. 14.25.

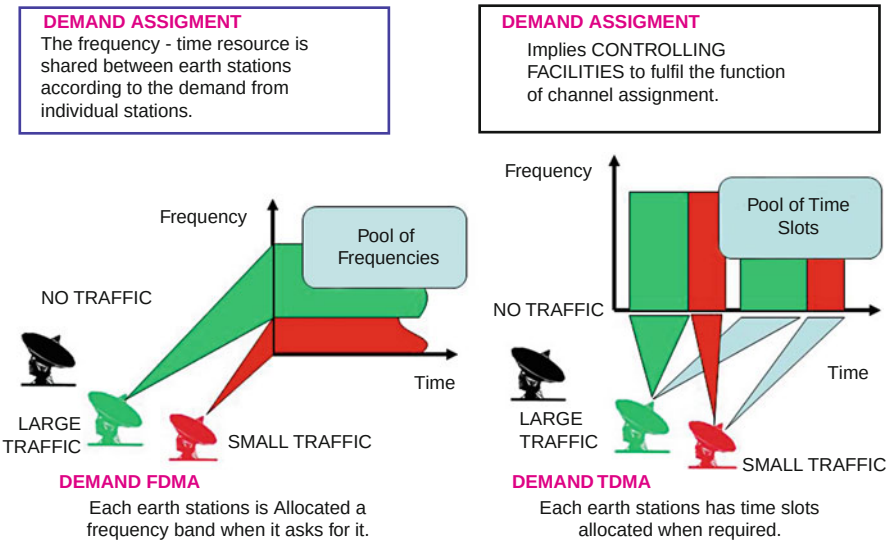


Fig. 14.23 Alternative depiction of FDMA and TDMA with DAMA

MF-TDMA: Multiple TDMA Streams

When a large number of low traffic terminals exist in a network, TDMA on the return link from the small terminals is not effective as the satellite transponder is not operated in a single carrier mode which avoids intermodulation effects. However, an effective configuration is to operate several small TDMA streams separated by frequency as depicted in Fig. 14.26. This is known as multi-frequency TDMA (MF-TDMA) and is used quite extensively in VSAT networks.

DAMA with MF-TDMA

In order to achieve a greater efficiency in small terminal satellite networks, the DAMA scheme can be used with MF-TDMA. By exchanging signaling packets, the process of call setup involves the negotiation among the Earth station and a master control station (MCS) which controls the satellite network. The signaling packets are transmitted in the synchronization area which is a fixed portion in the MF-TDMA frame. Once the connection is established, a certain amount of memory and bandwidth is allocated to the new connection.

DVB-RCS: MF-TDMA

We consider here the DVB-RCS return link with MF-TDMA and DAMA. In this standard there are 4 types of burst in the MF-TDMA:

- Traffic (TRF)

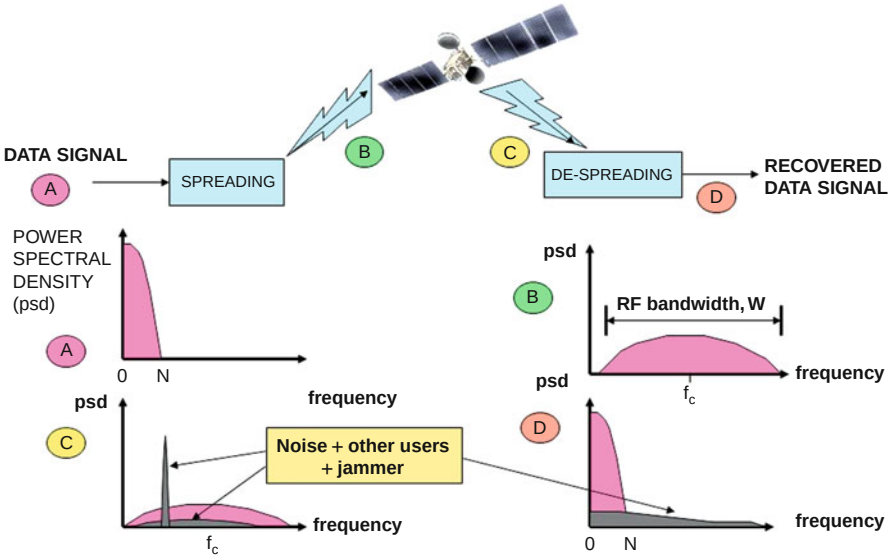


Fig. 14.24 The spread spectrum spreading and de-spreading concept

- ATM (53 bytes)
- MPEG (188 bytes)

- Acquisition (ACQ)
- Synchronization (SYNC)
- Common Signaling Channel (CSC)
  - Used by a terminal to identify itself during logon

ACQ and SYNC bursts are required for accurately positioning the terminals burst during and after logon. The concept is outlined in Fig. 14.27 where the bursts and their roles are depicted.

The DVB-RCS standard provides for the traffic bursts to change frequency as well as time slots in order to keep the spectrum loading balanced for interference advantages. Figure 14.28 shows the evolution of a five frequency DVB-RCS MF-TDMA system over time. At the start of the displayed time frequencies 3–5 are already using allocated time and frequency slots with existing traffic while frequencies 1 and 2 are just preparing to handle traffic after requesting it via the Slotted ALOHA CSC random access slot. It is of value to note that at any given frequency in the band used for MF-TDMA there is only one carrier transmitting at a given instance thereby providing a means to keep interference from the system at a low level.

Comparison of FDMA, TDMA, and CDMA

The comparison of FDMA, TDMA, and CDMA is presented in Fig. 14.30 in terms of advantages and disadvantages.

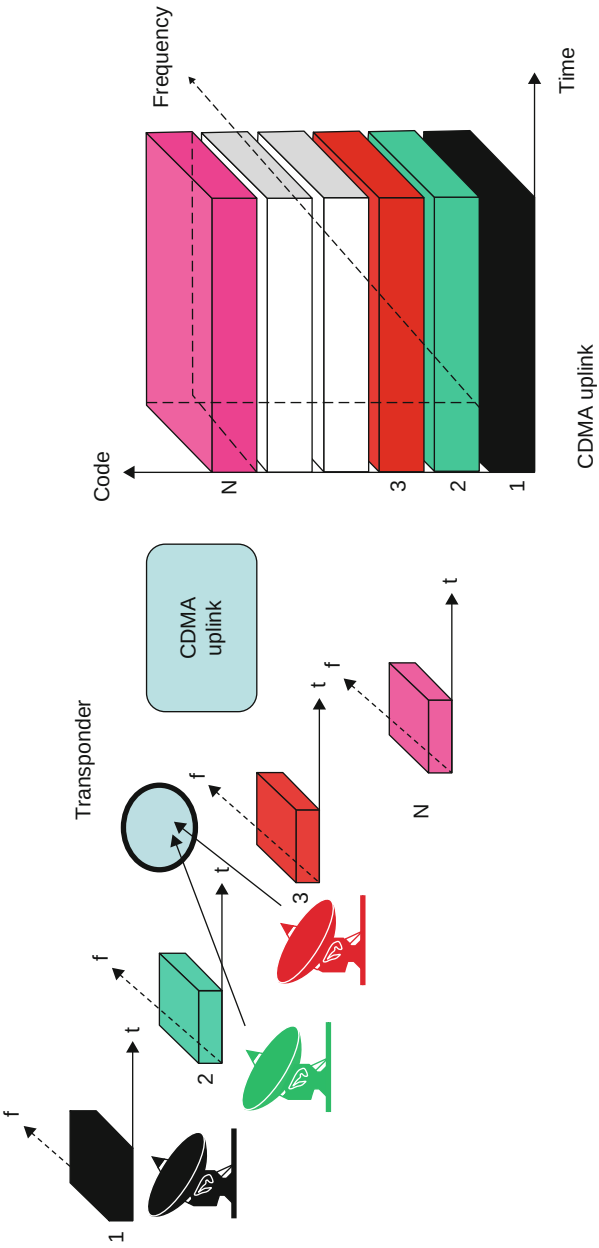


Fig. 14.25 The CDMA satellite access concept, each code being different (orthogonal)

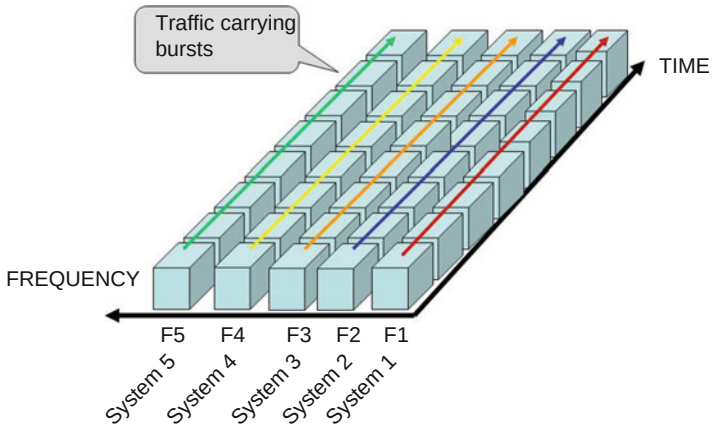


Fig. 14.26 Multiple TDMA streams (MF-TDMA)

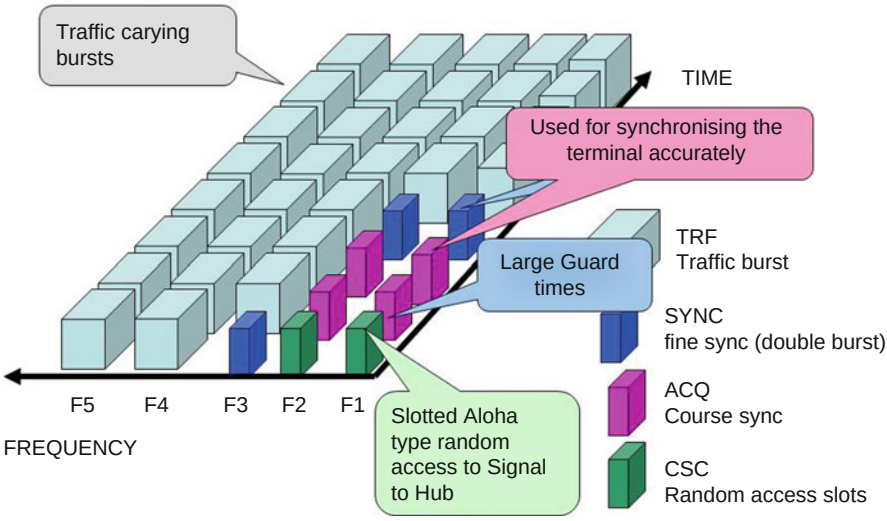


Fig. 14.27 The DVB-RCS MF-TDMA concept

The throughput of these systems is of interest and varies with the number of accesses in the system. Table 14.3 indicates in broad terms the throughput in percentage terms where 100% throughput corresponds to the capacity of the system with just one access only (Fig. 14.29). Hundred percent throughput corresponds to the capacity of the system considering one access only.

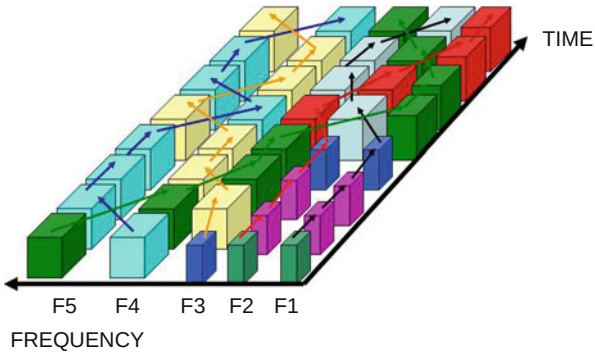


Fig. 14.28 The DVB-RCS MF-TDMA slot use over time

Table 14.3 Comparison of FDMA, TDMA, and CDMA

Type of multiple access	Advantages	Disadvantages
FDMA	Network timing not required	Intermodulation products cause degradation and poor power utilization
	Compatible to a lot of existing hardware	Uplink power control may be required
TDMA	No mutual interference between accesses	Network control required
	Uplink power control not needed	Large peak power transmission for earth station
	Maximum use of satellite transponder power, most efficient	Being digital in nature interface with analog system is expensive
CDMA	Network timing not required	Wide bandwidth per user required
	Anti-jamming capability	Strict code sync. needed

Random Access Schemes

FDMA/TDMA/CDMA fixed access schemes have been designed for circuit/stream traffic. Bursty data traffic, for example, packets, are more efficiently dealt with via random access schemes.

In random access there are no permanent assignments. In this case available resource is allocated when needed on a random basis.

The simplest system is ALOHA – where packets are randomly transmitted and if they collide with others are retransmitted with random time difference. ALOHA does not need synchronization but the maximum theoretical throughput is a mere 18%. A key impairment with such a scheme is the impact of the delay in getting information back to the transmission site that the packet was corrupted.

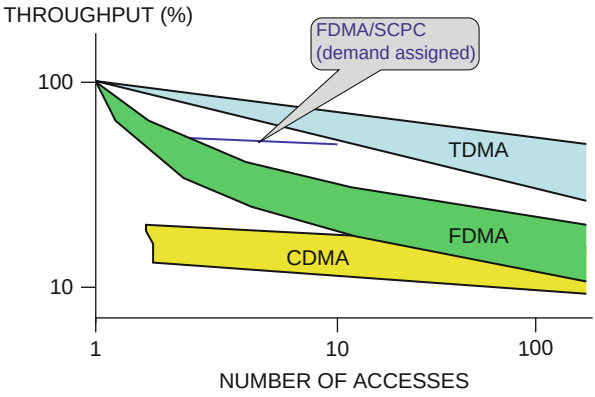


Fig. 14.29 Comparison of throughput for various access systems

SLOTTED ALOHA confines transmission to slot boundaries and needs time synchronization but the maximum throughput is increased to 36%. It is used, for instance, in the signaling channel of DVB-RCS.

As system rapidly becomes unstable as collisions build up, it is customary to operate below these maxima.

For variable length messages we need to employ more complex scheme, for example, selective reject ALOHA which breaks long packets into sub-packets and only retransmits sub-packets that collide; this provides a throughput of ≈0.37 which is independent of message length

DSA: Diversity Slotted ALOHA. In slotted Aloha, occasional collision among packets occurs, on the average a packet may have to be transmitted more than once before it is received correctly. This will introduce large packet delay in satellite slotted ALOHA systems, where each round trip propagation delay is about 270 ms. DSA is a slotted ALOHA scheme in which, whenever a user generates a packet, he transmits k copies of the same packet. It is assumed that there exists some arrangement which allows a receiver to reject all but one correctly received copy of any packet.

It has been found that multiple transmission gives better delay performance if the throughput is somewhat below its maximum. Also, if the probability that a packet fails if certain number of times or more is not to exceed a certain specified limit, then the multiple transmissions usually give greater throughput. This technique is utilized in the IP over Satellite (IPOS) the European Telecommunications Standards Institute (ETSI TS 102 354), and the Telecommunications Industry Association (TIA-1008-A).

Figure 14.30 indicates in diagram form how the various random access schemes function.

A comparison of the performance of the different approaches is given in Fig. 14.31.

Selecting a Random Access Scheme

Select RA scheme for traffic type and delay/throughput. Take care to achieve stability

- ALOHA: short bursty traffic

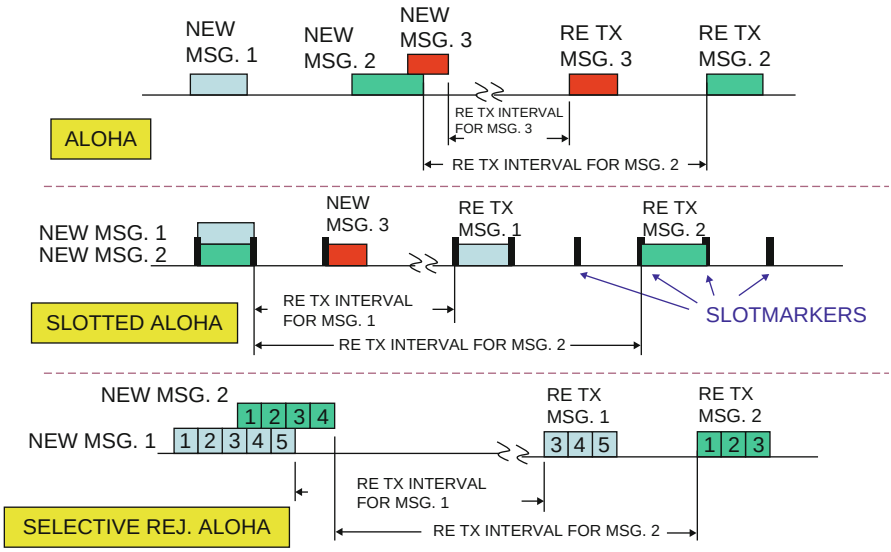


Fig. 14.30 Comparison of random access approaches

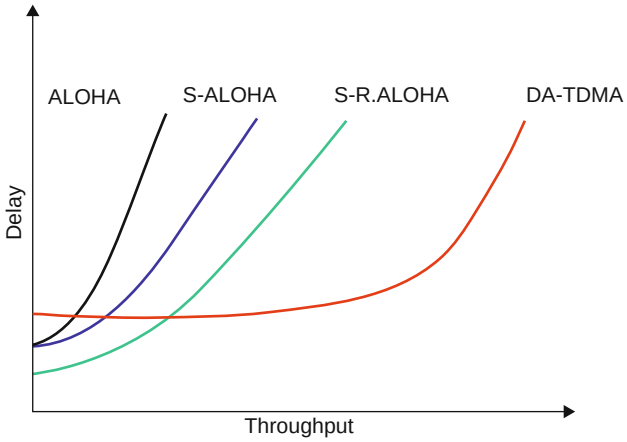


Fig. 14.31 Comparison of random access performance

- S-ALOHA: short bursty traffic – better throughput (used for signaling in DVB-RCS)
- S-R ALOHA: variable length longer messages
- DSA: Diversity Slotted Aloha (used for signaling in IPOS).

Multiple Access Summary

- This section has outlined the various access methods used in satellite communications.
- For traditional circuit switched voice and data FDMA has been commonly adopted but is less used today.
- For packet-based services TDMA and CDMA are used with the former gaining in popularity.
- Demand assignment is commonly employed with SCPC, TDMA, and CDMA which significantly improves the capacity of the system.
- Multi-frequency TDMA (MF-TDMA) is popular for low power multi-terminal uplink operation such as for DVB-RCS type service.
- Random Access methods have been outlined and compared. They are commonly used in signaling channels for MF-TDMA.
- Assignment and multiple access methods are key parts of any satellite network and are thus of critical importance.

Conclusion

This chapter has provided a background on how the information to be carried over a satellite link can be impressed on an RF carrier and then extracted at the far end. Furthermore the adoption of forward error correction (coding) has been outlined to demonstrate its value in optimizing the overall link capacity.

Connectivity between stations in a network has been addressed in terms of multiple access schemes and their functionality and efficiency. The various alternative systems used in digital satellite communications in terms of modulation and demodulation, multiple access techniques, and forward error correction and coding are discussed in terms of their individual merits and their advantages and disadvantages.

Notes

1. Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications, ETSI EN 302 307 V1.2.1
2. IBS/IDR stands for Intelsat Business Service/Intermediate Data Rate
3. Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for 11/12 GHz satellite services, ETSI EN 300 421 V1.1.
4. Digital Video Broadcasting (DVB); User guidelines for the second generation system for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications, ETSI TR 102 376 V1.1.1
5. Digital Video Broadcasting (DVB); Interaction channel for satellite distribution systems, ETSI EN 301 790 V1.5.1.
6. Digital Video Broadcasting (DVB); Interaction channel for Satellite Distribution Systems; Guidelines for the use of EN 301 790, ETSI TR 101 790 V1.4.1.
7. Digital Video Broadcasting (DVB); Interaction channel for Satellite Distribution Systems; Guidelines for the use of EN 301 790 in mobile scenarios, ETSI TR 102 768 V1.1.1.

References

K. Pahlavan, A.H. Levesque, *Wireless Information Networks* (Wiley, Hoboken, 1995a). ISBN 0-471-10607-0
J.G. Proakis, M. Salehi, G. Bauch, *Contemporary Communications Systems* (Brooks/Cole, Pacific Grove, CA, 2004). ISBN 0-534-40617-3

Further Reading

A. Burr, *Modulation and Coding* (Prentice Hall, Upper Saddle River, NJ, 2001)
I. Glover, P. Grant, *Digital Communications*, 2nd edn. (Prentice Hall, Upper Saddle River, NJ, 1998)
S. Haykin, M. Moher, C. Systems, *Communication Systems*, 5th edn. (Wiley, New York, 2010) (International Student Edition)
S. Lin, D. Costello, *Error Control Coding*, 2nd edn. (Prentice Hall, Upper Saddle River, NJ, 2004)
T. Moon, *Error Correction Coding* (Wiley, New York, 2005)
K. Pahlavan, A. Levesque, *Wireless Information Networks* (Wiley, New York, 1995)
J. Proaki, *Digital Communications*, 3rd edn. (McGraw Hill, New York, 1995)
M. Reza Soleymani, Y. Gao, U. Vilaipornsawai, *Turbo Coding for Satellite and Wireless Communications* (Kluwer, Netherlands, 2002)
B. Sklar, *Digital Communications*, 2nd edn. (Prentice Hall, Upper Saddle River, NJ, 2001)
P. Sweeney, *Error Control Coding* (Wiley, Chichester, 2002)

Satellite Transmission, Reception and On-Board Processing Signaling and Switching

15

Bruno Perrot

Contents

Introduction 360
Services Applications 361
Physical Transport Layer Services 363
Network Layer Services 363
Satellite Implementation Issues 364
Payload Architectures 364
Channelized Transponder Payload Implementation 365
De-modulation-Re-modulation (De-mod-Re-mod) Payload Implementation 369
Implementation Summary 371
Channelized Transponder Processor 371
De-mod-Re-mod Processor 371

Abstract

This chapter explains the technology that makes onboard processing function as well as explores the new and important applications that communication payloads, based on onboard processing techniques, can effectively support. Further it assesses the pros and cons associated with employing this technology in terms of performance, complexity, reliability, and cost. Satellite systems providing fixed and mobile services are evolving from bent-pipe payloads to more and more enhanced satellites with more and more capabilities and "intelligence." Thus one has seen the evolution of satellite capabilities to be able to achieve more and more functionally in space. We started with so-called non-intelligent or bent-pipe satellites and then moved quickly to more flexible multi-points-type

B. Perrot
SES, Chateau de Betzdorf, L-6815 Betzdorf, Luxembourg
e-mail: bruno.perrot@ses.com

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_16, © Springer Science+Business Media New York 2013

satellite services. Next there was the transition to more enhanced satellites with onboard switching, and then most recently there have been design innovations to bring true "intelligence to space." This has been seen in the move toward highly capable satellites with increasingly "intelligent forms" of onboard processing (OBP).

This evolution involves moving from more efficient beam switching to actual processing of signals to enhance signal and remove attenuation affecting the uplink and thus partially overcome rain attenuation. The addition of so-called intelligent functions to the satellite that were once found only in terrestrial signaling and switching systems allows satellites to become more efficient and versatile.

In particular this transition will allow the design and deployment of:

- Multi-beam RF-IF switched transponder satellites (i.e., the ability to provide effective "beam switching" among satellite beams) This allows satellites to provide Physical Transport Layer Network Services that were once restricted to advanced terrestrial networks.
- And eventually there will be an evolution to advanced packet switched ("Data Switched" Asynchronous Transmit Mode (ATM or ATM like services). This will allow onboard processed multi-beam satellite systems that provide specific and an increased array of network-level services.

Keywords

Asynchronous transfer mode (ATM) switching • Baseband processing • Bent-pipe satellites • Bit error rate • Intelligent satellite • Intermediate frequency (IF) • Multi-beam antennas • Onboard processing (OBP) • Radio frequency • Solid state amplifiers (SSAs) • Traveling wave tube amplifier (TWTA)

Introduction

Onboard switching systems are designed to make more efficient use of a satellite communication network, especially those that employ multi-beam technology that entails onboard switching to interconnect uplink and downlink beams with a high degree of efficiency. Onboard processing can also be used to reduce bit error rates for the uplink and downlink transmissions and to allow satellites to be optimized to provide a wider range of service applications more efficiently.

The added cost and complexity of including such a capability on a satellite, however, has for a number of years served to argue against investing in this technology. Industry players and satellite operators must, in effect, look at the "opportunity cost" represented by adding onboard capability to a satellite in terms of the extra mass, volume, and engineering and manufacturing costs. The cost, mass, volume, and reliability considerations associated with an OBP payload rule out the option of simply adding more conventional throughput capability that would have little incremental cost. The service flexibility and throughput efficiencies, the

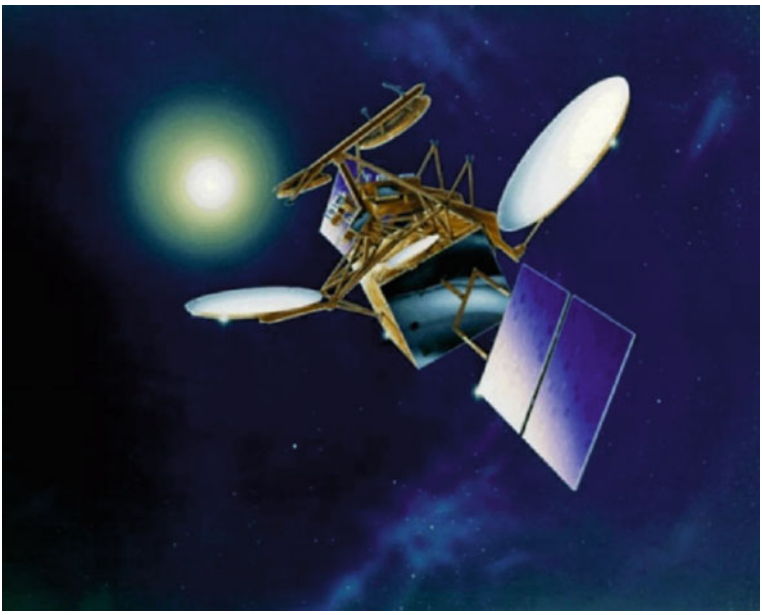


Fig. 15.1 The NASA developed Advanced Communications Technology Satellite (ACTS) (Graphics courtesy of NASA)

improved bit error rate capabilities, and the greater functionality of an "intelligent spacecraft" on the other hand have over time served to move telecommunications satellites – particularly those providing fixed and mobile services – in this direction over time. In earlier sections it has been noted that experiments to demonstrate onboard processing have been in progress for some time. Most recently there have been the experimental satellite systems of Japan, including the WINDS and QZSSS spacecraft. It was the ACTS satellite deployed by NASA over a decade ago that first demonstrated onboard processing for space communications and ATM networking (Fig. 15.1).

Services Applications

In a historical context, satellites have served to provide a connection between physical nodes of a network and not perform any of the higher level communications functions. Thus telecommunications satellites have in the past not participated as active nodes within networks and processed where or how information is routed. This is to say that the active signaling and processing of information at base band or intermediate frequencies were all accomplished within terrestrial switching centers. Satellites, in contrast, were simply employed to transmit higher frequency RF

signals as essentially "cables in the sky." They worked only at the physical layer of the seven-layer OSI Model as provided below.

OSI model levels – used in ATM switching according to X.200 recommendation	
The seven layers	Function
#7: *Application*	Network process to application: actual content such as e-mails, video images, voice, data
#6: *Presentation*	Data representation, encryption and decryption, convert machine dependent data to machine independent data: provides for such functions as encryption or data conversion
#5: *Session*	Inter-host communication: starts and stops sessions and creates the correct order
#4: *Transport*	End-to-end connections, reliability and flow control: ensures that the entire and complete message is delivered
#3: *Network*	Path determination and logical addressing: routes information to a particular location based on network address
#2: *Data link*	Physical addressing: routes data packets from node to node based on station addresses and the actual transmission mode
#1: *Physical*	Media, signal and binary transmission: provides the physical channel to connect nodes in a network

Note that the layers 1–3 are named "Media layers" while layers 4–7 are "Host layers."

As satellites have become more complex and as multi-beam antenna systems are used to switch waveforms between a growing number of uplinked and downlinked beams, spacecraft have been designed to become, in effect much "smarter." This added processing power and "intelligence" have allowed satellites to accomplish more complicated functions and provide for functionality above the physical level of the Open System Interconnection model (OSI model). Since the processing power was being added in any event, the thought evolved as to how one might add even more signaling capability to the satellite to accomplish even more complex tasks at ever higher levels within the seven-layer model of ATM-based communications. The same logic has also been applied to the more streamlined architecture of TCP/IP protocol as is used within the Internet.

The most important thought in this regard has been the idea that additional processing power on the satellite could serve to demodulate information (i.e., waveforms) back down to the base band or intermediate frequencies.

This could allow improvements to the signal, could reduce bit errors in the uplinking and downlinking of signals or otherwise assist with digital service applications. The basic thought has been that "smart satellites" could ultimately help allow services sent via telecommunications satellite to be more competitive with terrestrial fiber optics and digital switching and signaling systems on the ground. It could allow satellites to deliver advanced services directly to mobile users or small office or home office (SoHo) users without the need of operating through terrestrial ATM or other modern digital switches.

In short, satellites with onboard switches can play a more active role within the digital networks of the future. Two kinds of services can thus be supported.

Physical Transport Layer Services

This type of service is the most basic. Physical Transport Layer Satellite Services are considered to be systems that are utilized to provide a physical layer interconnection between two network nodes, such as a fiber or terrestrial link would be used. This level of service provides node interconnections that are distance independent and readily available for remote service providers.

The figure below illustrates such a network application (Fig. 15.2). Note that for these services the satellite is not an active network-switching node in the manner of an ATM switch node or similar terrestrial network nodes.

Statistical multiplexing and buffering are kept on the ground and service revenue increases due to their utilization are therefore realized by the network operators and users, not the satellite operator.

It is envisioned that Physical Layer Network Satellite Services can be implemented using a Channelized Transponder Payload (see next section for details) configured by ground control of the payload switching configuration. Network users obtain satellite resources via a terrestrial control center that controls user access and system configuration. This demands that an additional earth-space-earth hop is required for user to control center communication. Thus Physical Layer Network Satellite Services (PLNSS) are appropriate for applications that have set-up, reconfiguration, and tear-down requirements compatible with fractional second timelines or longer.

These types of PLNSS operations therefore appear appropriate for leased line, dynamic assignment, and replenishment operational applications. Representative applications are listed in Table 15.1 at the end of this chapter. These applications also imply the use of satellite resources by a variety of networks, since channelized transponder implementations of physical layer satellite resources are independent of the waveform and relatively independent of the network protocols they are suited to varied applications.

Physical Layer Satellite Services implemented with a Channelized Transponder allow for modulation and protocol changes between users and over time. The satellite operator is therefore involved in a more traditional capacity and the necessary channel switching can be used to optimize channel capacity leasing opportunities.

Network Layer Services

Network Layer Satellite Services are services where the satellite is an active switching node within the network as illustrated in Fig. 15.3.

A Network Layer Satellite Network implementation provides network operations on the ground and on the satellite. The satellite operator therefore can benefit from the statistical channel loading and variable service-type multiplexing. Implementation of Network Layer Satellite Services requires the implementation of more complex De-modulation-Re-modulation (De-mod-Re-mod) Payloads (see next section

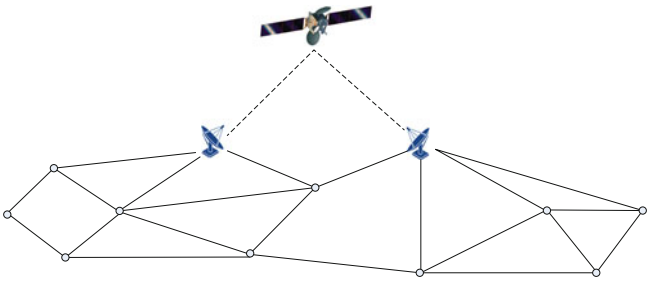


Fig. 15.2 Physical transport layer services

for details). This approach implies that all users must adhere to a fixed set of modulation and protocol formats compatible with the satellite. Changes between users and transition in modulation or protocol formats over time are difficult if not impossible to achieve. External network users are constrained by the satellite network implementation. The satellite operator is therefore effectively the network operator.

Satellite Implementation Issues

The design, engineering, and manufacture of a satellite with onboard processing thus leads to a number of key decisions involving the payload architecture, the antenna design, and the type of amplification sections that are utilized.

Table 15.1 below summarizes the advantages and disadvantages of each configuration and compares four different processor architectures:

Payload Architectures

Channelized Transponder payloads filter received signals within a switchable bandwidth and connect the filtered signal to one or more downlink beams.

In contrast, De-modulator-Re-modulator payloads receive signals with specific uplink modulation formats, demodulate the received packets, process the packet addresses, route the received packets to the destination beams, and re-modulate the packets with the specific downlink modulation format.

Physical Layer Network Satellite Services can be implemented with either Channelized or De-mod-Re-mod payload implementation. Since the Channelized Transponder implementation is simpler, the normal case would be for a Channelized Transponder payload to provide Physical Layer Services.

A De-mod-Re-mod payload could also be used for Physical Layer Services in a manner analogous to leased line services implemented by a switched network or Virtual Private Network (VPN) operations. However, Network Layer Services, as used here, require the implementation of a De-mod-Re-mod payload. Baseband packet processing is required to implement the network node switching and routing functions.

Table 15.1 Summary of implementation issues

Processor type	Applications	Advantages	Disadvantages
Channelized transponder processor			
Analog	• Video conference • News gathering • Interactive video forward • Two-way file transfer • Software downloads • VPN (larger bandwidth applications)	• Switches arbitrary signals • Switching flexibility increased compared to mechanical switches • Minimum satellite power	• No of cross-points drives complexity • Limited by minimum channel bandwidth • Variable bandwidth • Mass, volume • Multiple carrier D/L
Digital	• Interactive audio • Interactive video return link • Games • Data downloads • Video conference • Digital messaging • Internet user • Video streaming • E-mail, fax (smaller bandwidth applications)	• Switches arbitrary signals • Full cross-point connectivity possible • Small switching channel bandwidth possible • Variable bandwidth possible	• Power, mass, volume, thermal • Multiple carrier D/L
De-mod-Re-mod processor			
Asymmetric	• Interactive video forward • Software downloads • Data downloads (asymmetric BW applications)	• Full cross-point connectivity possible • TDM or packet routing possible to optimize throughput	• Power, mass, volume, thermal • Switches specific signals • Single carrier D/L
Symmetric	• Interactive audio • Games • Data downloads • Video conference • Two-way file transfer • Digital messaging • Internet user • Video streaming • E-mail, fax • T1 virtual private network (VPN) (less than T1 applications	• Full cross-point connectivity possible • TDM or packet routing possible to optimize throughput	• Power, mass, volume, thermal • Multiple demodulators drives power requirements • Switches specific signals • Single carrier D/L

Channelized Transponder Payload Implementation

The figure below summarizes the functional operation of a channelized transponder processor that assumes a digital implementation of the channelized transponder filtering and switching function (Fig. 15.4). Any signal received within bandwidth A1–Z1 is down-converted to an intermediate frequency or baseband and digitally sampled. These samples are digitally filtered, stored, and routes to the switch port

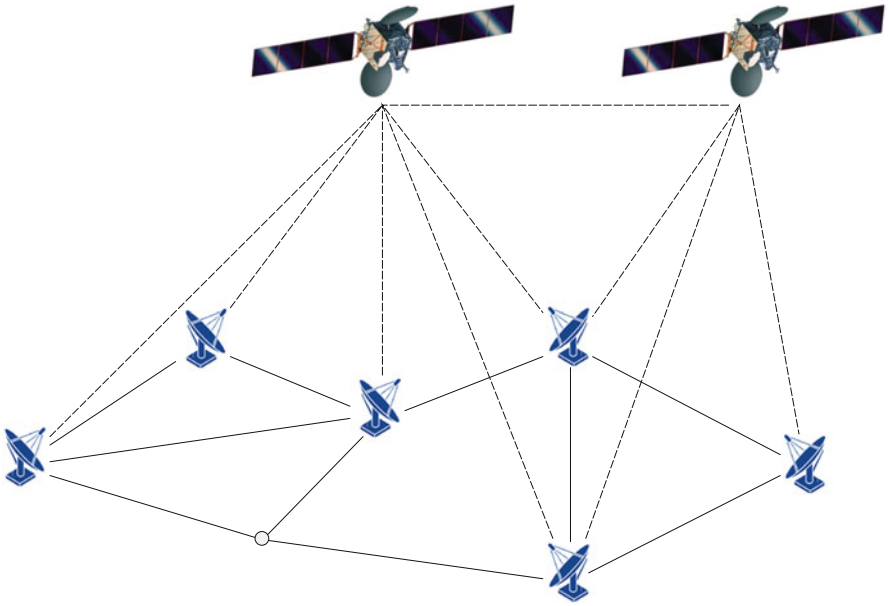


Fig. 15.3 Network layer services

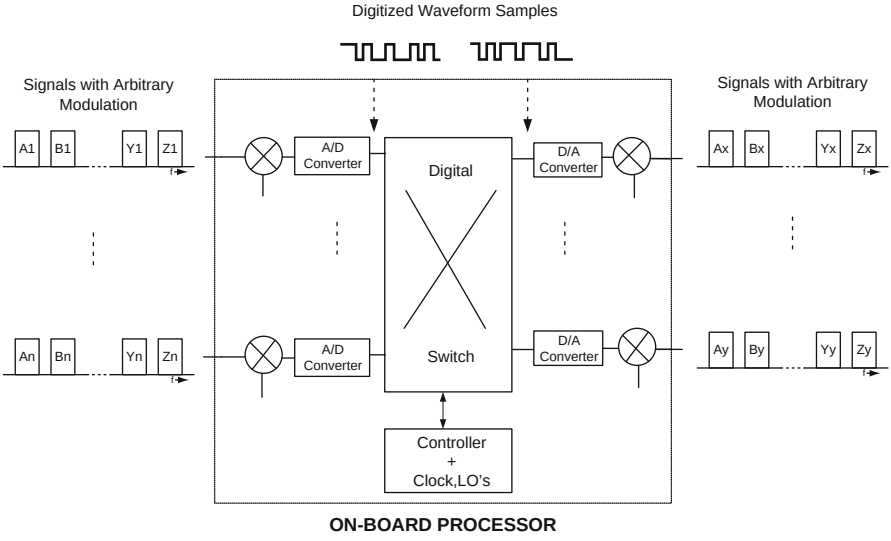


Fig. 15.4 Channelized transponder onboard processor architecture

corresponding to the desired downlink beam. This routing may be accomplished by a simple readdressing of the stored digital samples within a common output buffer memory or by a more traditional digital switch implementation.

The figure below, however, illustrates a conceptual block diagram for a Channelized Transponder payload hosting the processor described previously. Multiple uplink beams are formed by multiple feed apertures or phased array antennas. Multiple antenna apertures may be necessary due to feed packing, spillover, and beam shaping requirements. Current technologies support the implementation of received phased arrays for certain applications.

Signals received in each beam are amplified and down-converted before filtering and switching by the processor. Multiple processors may be utilized to minimize power dissipation and allow variable filter bandwidths. The output of the processor is a digitally routed and reconstructed version of the received waveform. Each output is up-converted to the appropriate downlink frequency and amplified. Similar to traditional transponders, Channelized Transponder payload amplifiers must be operated in back-off mode if multiple signals are present within the channel to avoid waveform degradation due to nonlinear effects. The resulting payload power efficiency is a significant consideration for Channeled Transponder payload applications.

Figure 15.5 also includes a multiple beam transmit antenna farm with a set of Low Noise Amplifiers (LNAs) and Traveling Wave Tube Amplifiers (TWTAs).

The achievable power efficiency for solid state power amplifiers (SSPAs) is not currently high enough to be optimal to support transmit phased arrays. In future years this may be the case. Thus currently tube-based TWTA implementations are most likely indicated for the most cost-effective design, even though both options might be viable for lower powered systems. The technology in this area is in transition and the cost and performance equations are thus constantly changing. Additional system requirements that favor phased arrays such as beam reconfiguration or steering could mandate solid state amplifiers at least for some designs.

Channelized Transponder processors and payloads can be implemented with currently available technologies. Receive multiple beam antennas can be implemented as passive arrays or active arrays that allow reconfiguration. Multiple beam systems require use of numerous low noise amplifiers (LNA) in receivers and down-converters. This generally requires one per beam. Technologies that enable power efficient implementation, interconnection, and packing of numerous components are required to support growth to a large number of beams. Channelized processor using surface acoustic wave (SAW) analog technology and digital technology has been developed for narrowband Mobile Satellite Systems (MSS). These MSS technologies continue to advance. Larger bandwidth channelized processors are being developed utilizing the analog SAW technologies.

Current digital technologies can be used to implement channelized processors that take care of wider bandwidths. Digital processors often can more easily implement multiple or variable filter bandwidth than analog processors. Digital processors, however, require additional satellite power to achieve that degree of operational flexibility.

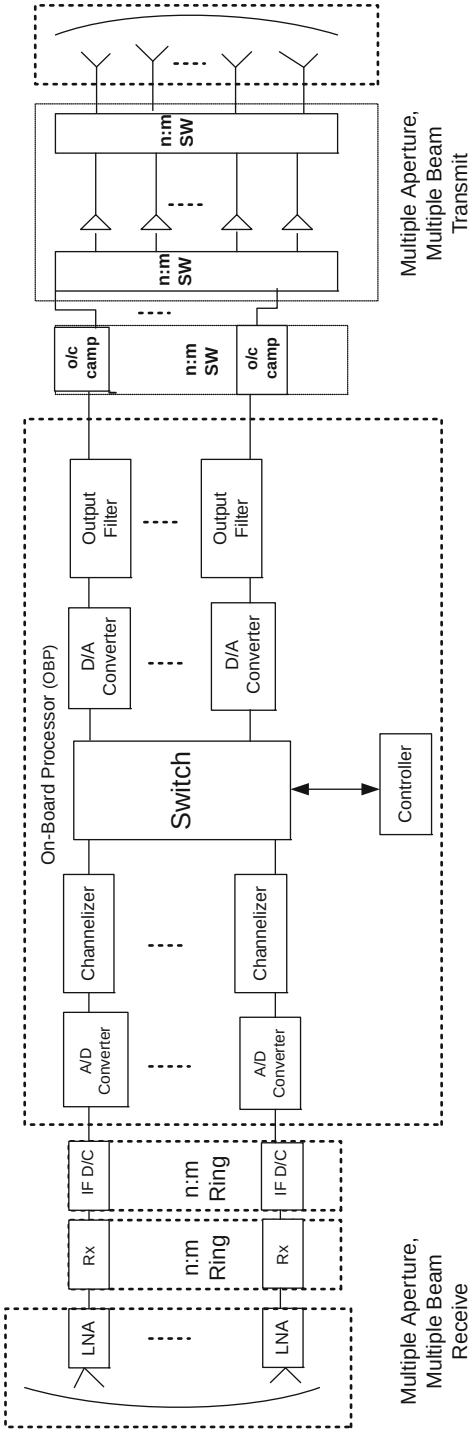


Fig. 15.5 Channelized transponder payload architecture

De-modulation-Re-modulation (De-mod-Re-mod) Payload Implementation

The figure below summarizes the functional operation of a digital implementation of a De-mod-Re-mod processor (Fig. 15.6). Any signal received in channel A1–Z1 (See chart below) is down-converted to an intermediate frequency or baseband and digitally filtered, demodulated, and error correction coded. The demodulated packets are stored and routed to the switch port corresponding to the desired downlink beam. This routing may be accomplished by memory or by traditional digital switch implementation as discussed with respect to the channelized transponder functional description.

A conceptual block diagram for a De-mod-Re-mod payload which could utilize the processor, described above is shown below. Just as is the case for the channelized transponder, multiple uplink beams are formed by multiple feed systems working to conventional parabolic reflectors or phased array antennas. In both cases the signals received in each beam are amplified and down-converted before filtering, demodulation, decoding, and switching within the processor.

The output of the processor is an encoded, re-modulated waveform. Each output is up-converted to the appropriate downlink frequency and amplified. In contrast to traditional transponders, De-mod-Re-mod payload amplifiers can be operated at saturation since the data from multiple uplink signals are combined at baseband by the switch and modulated onto one transmitted carrier. However, this power saving may be more than offset by the power required by the filtering and demodulation.

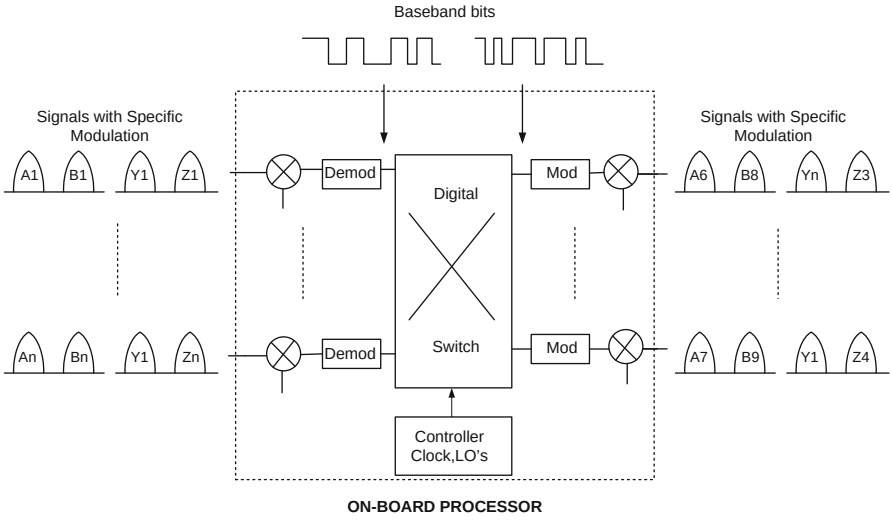


Fig. 15.6 De-mod-Re-mod onboard processor architecture

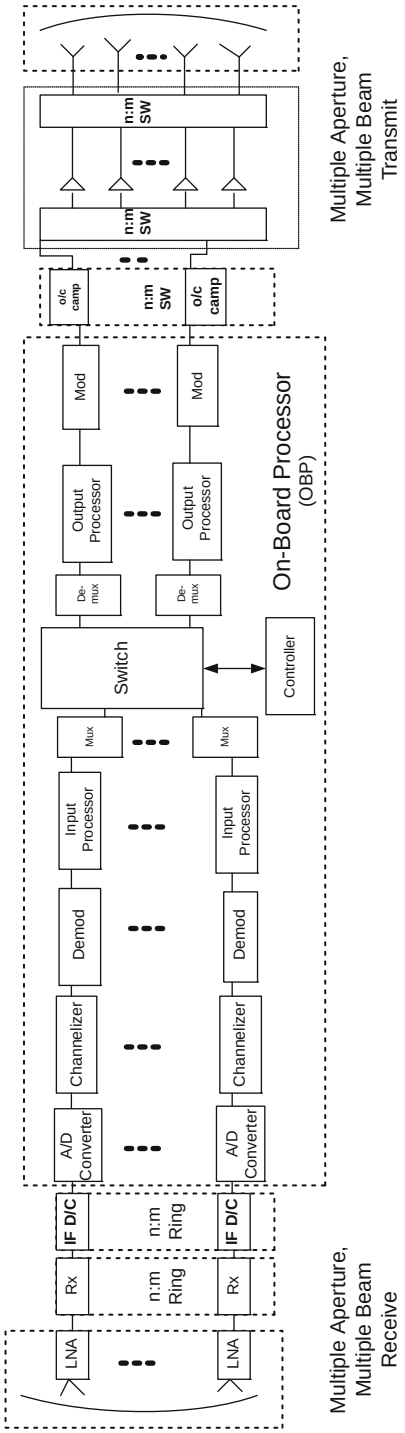


Fig. 15.7 De-mod-Re-mod payload conceptual block diagram

This is true for the same reason that TWTAs are advantageous for a channelized transponder payload.

The figure above also includes a multiple beam transmit antenna farm with TWTAs. De-mod-Re-mod processors and payloads implementation are currently limited to available space qualified technologies (Fig. 15.7). The power required for many proposed Ka-band multi-beam, multi-channels payloads is driven by the power required for the IF/baseband processors, de-modulators, and re-modulators. Due to important hardware quantities, technology enhancements that allow significant reduction in power of these components will be required before this type of payload come to widespread use. The state of all the other payload technology is similar to that described for channelized transponder payloads.

Implementation Summary

Channelized Transponder Processor

The channelized transponder payload can be implemented with analog filtering and switching or digital filtering and switching. The applications suited to these implementations potentially can be allocated based on bandwidth. Analog implementations may be more appropriate for channel bandwidths greater than 1 or 2 MHz. Digital implementations are generally more adequate for channel bandwidths below 1 or 2 MHz or applications where multiple or variable bandwidth are required. Processor power and thermal dissipation definitely are a spacecraft implementation dominant trade parameter.

De-mod-Re-mod Processor

Two De-mod-Re-mod implementations are discussed. Both are quite viable options. The asymmetric implementation includes fewer uplink beams and demodulators than downlink beams and modulators. Applications for this architecture could most likely include data distribution networks. The symmetric implementation includes an equal number of uplink beams and demodulators and downlink beams and modulators. Applications for this architecture include advanced Internet applications or networks that require symmetric data services. Again the processor power will be a dominant spacecraft implementation trade-off consideration.

Conclusion

- Onboard switching system makes efficient use of a satellite communication network having onboard multi-beam technology.
- Channelized Transponder payload is adequate to Physical Transport Layer Services as a result of its simplicity.

- A Demodulation-Remodulation Processor payload is more adapted to the Network Layer Services thanks to its flexibility and reconfigurability.
- Processor power and thermal dissipation are among the most dominant trade-off parameters for payload implementation on the spacecraft.
- Evaluation of advanced satellite network architectures must be based on solid business plan before advanced architectures are accepted and implemented.

Satellite Communications Antenna Concepts and Engineering

16

Takashi Iida

Contents

Introduction 374
Fundamental Parameters 375
Antenna Pattern 375
Beamwidth 375
Background Knowledge 375
Radiation Direction 378
Radiation Power 378
Gain 379
Reversibility of Antenna Characteristics 380
Effective Area 380
Polarization 381
Basic Antennas 383
Linear Wire Antenna 383
Horn Antenna 383
Reflector Antenna 385
Helical Antenna 387
Microstrip Antenna 387
Array Antennas for Scanning and Hopping Beams 390
Function of Array Antenna 390
Directivity of Array Antenna 391
Gain of Array Antenna 391
Phased Array Antenna 391
Multibeam Antennas with Multiple Feed Systems 392
Function of Multibeam Antenna 392
Type of Multibeam Antenna 394

T. Iida
Tokyo Metropolitan University, 1-23-4 Mejirodai, Hachioji, Tokyo, 193-0833, Japan
e-mail: QZF04134@nifty.com

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_18, © Springer Science+Business Media New York 2013

Antennas for Optical Communications Systems 394
References 396

Abstract

The most critical component of a communication satellite is its antenna system. The purpose of this chapter is to show how the antenna works by describing the basic concepts related to satellite antenna pattern, side lobe, gain, and polarization so that satellite antenna systems are designed and engineered to meet specific requirements. First, the fundamental parameters such as antenna pattern, beamwidth, radiation power, gain, and polarization are introduced. Second, basic antenna such as linear wire antenna, horn antenna, reflector antenna, and microstrip antenna is described. Third, array antenna for scanning and hopping beams is described for its function, gain, and phased array. In the fourth, multibeam antenna is described in terms of its function and type. Finally, an antenna for optical communications system is introduced briefly.

Keywords

Antenna gain • Antenna pattern • Array antenna • Cassegrain antenna • Circular polarization • Effective area • Half-power beamwidth • Helical antenna • Horn antenna • Linear polarization • Main beam • Microstrip antenna • Multibeam antenna • Parabolic antenna • Phased array antenna • Polarization • Reflector antenna • Side lobe

Introduction

The most critical component of a communication satellite is its antenna system and supporting electronics. Without this communications subsystem the satellite cannot fulfill its mission. The design of this system ultimately determines the communications capacity of the satellite. The satellite antennas and communications also determine the types of ground antennas or user devices that can access the satellite and the telecommunications services that can be provided. This chapter explains the fundamental concepts on which satellite antenna systems are designed and engineered to meet specific requirements. As spacecraft platforms have matured to allow the deployment of larger and higher gain satellite antennas as well as to point them with greater accuracy, it has been possible to launch more and more capable space communications systems. The engineering fundamentals are presented in this chapter, and the next chapter explains the evolution of these more sophisticated antennas and the resulting capability to launch higher capacity satellites that can work to ever smaller ground antennas, including handheld satellite phones. After reading these two chapters, it is useful to also read the chapters on ground antenna systems that connect with the satellite antennas.

Fundamental Parameters

Antenna Pattern

The key to a satellite antenna is the pattern, or the concentrated, beam that is emitted.

Figure 16.1 shows a typical pattern of antenna. Figure 16.1a shows a three-dimensional image and Fig. 16.1b indicates a cutting edge of the antenna pattern at an angle of "ϕ" whose vertical axis shows radiation strength at the horizontal axis of "θ." The direction of "z" shows a main beam and the directionality of antenna. The lobe other than the main beam is called a side lobe. The side lobe should be suppressed as much as possible because it may cause interference to or from other satellites and/or terrestrial communication system. In the case of an aperture antenna such as a parabolic antenna, if the aperture is illuminated by uniform amplitude and phase, the maximum gain is obtained for the given aperture and the gain of the first side lobe is decreased by 13 dB from the gain of the main beam. This is to say that the first side lobe is 20 times less powerful than the main beam.

Beamwidth

The beamwidth is obtained from antenna pattern. Usually a half-power beamwidth is used, which is the angle of beamwidth at 3 dB decreased gain from the peak gain as shown in Fig. 16.1. The half-power beamwidth, $\theta_{1/2}$ of usual reflector type antenna such as parabolic antenna is given approximately as

$$\theta_{1/2} = \alpha \frac{\lambda}{d} \quad (\text{deg.}) \tag{16.1}$$

where d is diameter of antenna and value of 65–70 is used as α.

Background Knowledge

Coordinate System

As noted earlier, a polar coordinate system (R, θ, ϕ) is used in the antenna technology usually. Figure 16.2 shows a polar coordinate system.

Solid Angle

When the area on a spherical body is S, the solid angle is defined by

$$\omega = \frac{S}{R^2} \tag{16.2}$$

where R is radius of sphere and whose unit is "steradian."

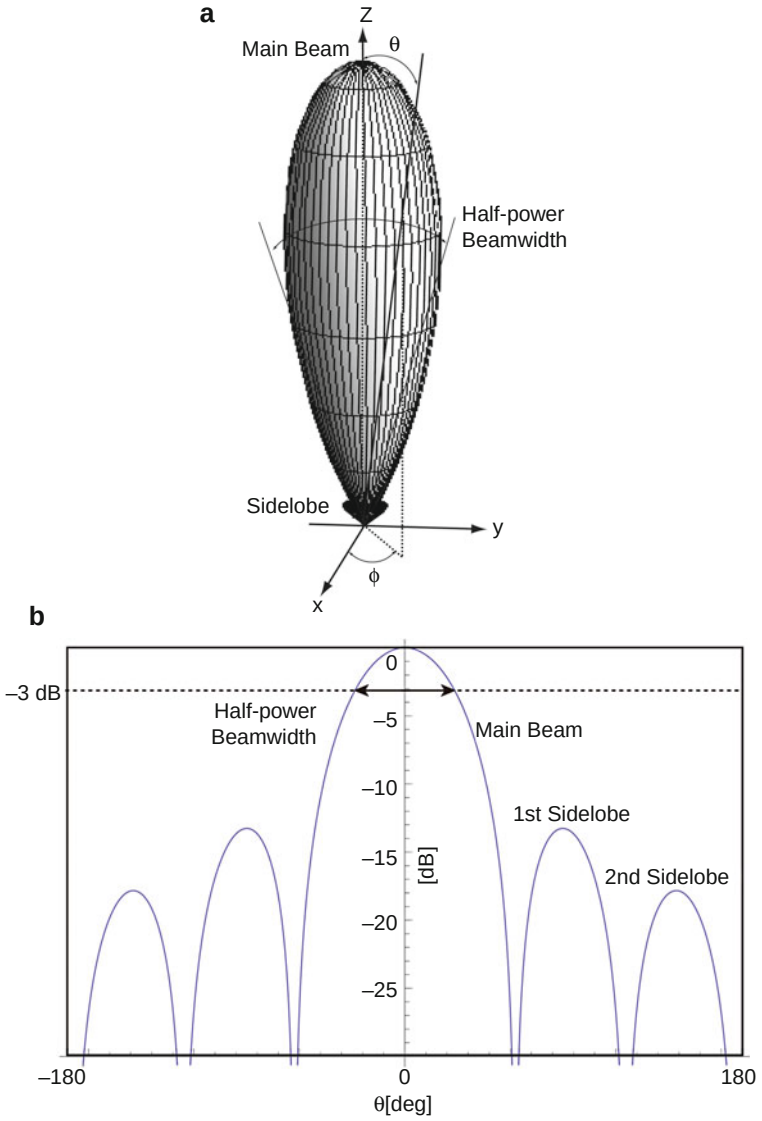


Fig. 16.1 Typical antenna pattern. (**a**) Three dimensional image (**b**) Two dimensional image

The relationship between a unit area and a unit solid angle is shown as follows: The unit area means $S = 1$ in Fig. 16.3. This is indicated by solid angle as

$$\omega = \frac{1}{R^2} \tag{16.3}$$

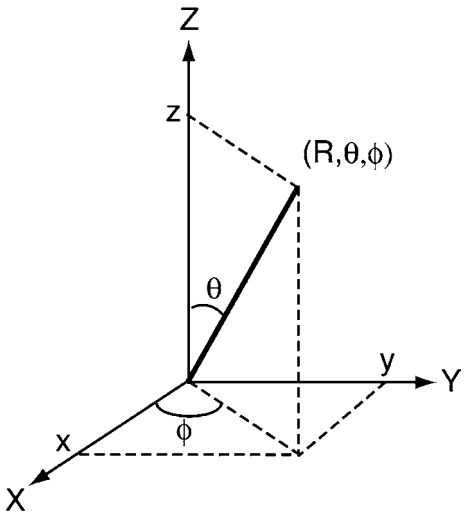


Fig. 16.2 Polar coordinate system

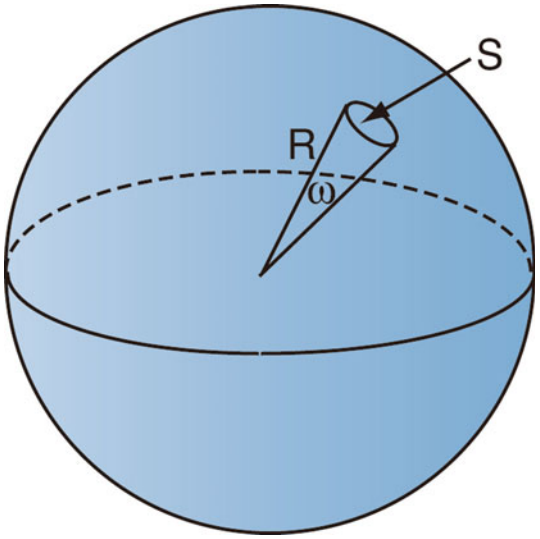


Fig. 16.3 Solid angle

This equation means the transformation of unit area to unit solid angle. The unit solid angle means $\omega = 1$ (steradian), that is, since

$$1 = \frac{S}{R^2} \tag{16.4}$$

the area per unit solid angle occupied on the surface of the sphere is given by

$$S = R^2 \tag{16.5}$$

The solid angle of the entire sphere is, since the area of the sphere is $S = 4\pi R^2$,

$$\omega = \frac{4\pi R^2}{R^2} = 4\pi \tag{16.6}$$

Radiation Direction

The electromagnetic field at the point $(R,\ \theta,\ \phi)$ sufficiently far radiated by an arbitrary antenna positioned at the origin of the polar coordinate system in Fig.16.2 is given by

$$\left.\begin{aligned} E(R,\theta,\phi) &= \frac{e^{-jkR}}{R} U(\theta,\phi) \\ k &= \frac{2\pi}{\lambda} \end{aligned}\right\} \tag{16.7}$$

where λ: wavelength and R: distance.

$U(\theta,\phi)$ is called the direction function and is determined only by θ and ϕ independently of distance, that is determined only by direction. It includes all the characteristics about the direction. Cutting $U(\theta,\phi)$ by a plane and observing it is called antenna pattern.

(Equation 16.7) is calculated from the famous Maxwell's equations that are the key basis for understanding electromagnetic transmissions. As for detailed solution from the Maxwell's equations, see (Balanis 1997).

Radiation Power

The power flow on the surface of sphere with the sufficiently large radius of R is perpendicular to the spherical surface and directed to the outside. The power per unit area on the sphere, that is, power density is given by

$$P(R,\theta,\phi) = \frac{(E(R,\theta,\phi))^2}{Z_0} = \frac{(U(\theta,\phi))^2}{Z_0 R^2} \tag{16.8}$$

where Z_0 is performance characteristics of media which is called inherent impedance and is given in the vacuum by

$$Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} \cong 120\pi = 367.6 \quad [\Omega] \tag{16.9}$$

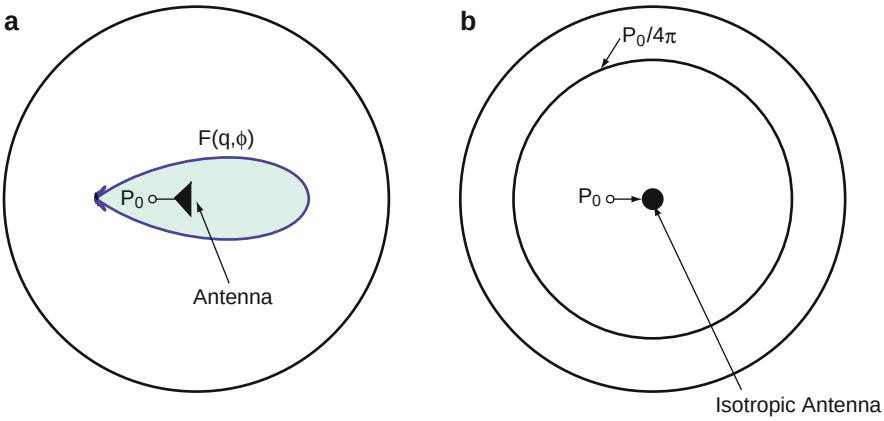


Fig. 16.4 Antenna gain. (**a**) Directive antenna (**b**) Isotropic antenna

The power per unit solid angle, $F(\theta, \phi)$, is given by

$$F(\theta, \phi) = \frac{|U(\theta, \phi)|^2}{Z_0} \tag{16.10}$$

where $F(\theta, \phi)$ indicates the power strength to the direction of (θ, ϕ).

Gain

The antenna gain is defined as "a ratio of power per unit solid angle radiated to the arbitrary direction from the antenna to the power per unit solid angle radiated from the isotropic antenna derived by the same power as one of the antenna," where an isotropic antenna means an ideal nondirective antenna which radiates uniform strength of electromagnetic field to every angle. The isotropic antenna is used as a reference of antenna gain. Given the power density $F(\theta, \phi)$ radiated from an antenna to the direction (θ, ϕ), gain $G(\theta, \phi)$ is given by

$$G(\theta, \phi) = \frac{F(\theta, \phi)}{P_0/4\pi} = 4\pi \frac{F(\theta, \phi)}{P_0} \tag{16.11}$$

where P_0 means a supplied power.

This can be considered as follows: When the power P_0 is supplied to the input terminal of antenna as shown in Fig. 16.4a, $F(\theta, \phi)$ means the power per unit solid angle of the radio wave radiated to the direction (θ, ϕ). Meanwhile, when the power P_0 is supplied to the input terminal of isotropic antenna with no-loss as shown in Fig. 16.4b, all the power P_0 is radiated to the space because the isotropic antenna is no loss, that is, energy per unit solid angle is converted to $P_0/4\pi$, since the energy P_0 is radiated to every solid angles and this radiation is uniform for every direction.

Therefore, according to the definition of gain mentioned above, (16.11) is a right expression. The expression of gain in (16.11) includes loss at antenna and it is also called power gain. The value of $G(\theta, \phi)$ at the maximum radiation direction (θ, ϕ) is called absolute gain and expressed in decibel [dBi], where the suffix "i" means isotropic antenna.

Reversibility of Antenna Characteristics

An antenna has the same characteristics of gain and pattern for using it as transmitting antenna as for using it as receiving antenna and the reversibility is held.

Effective Area

Supposing that the power of incident electromagnetic wave per unit area is P and the effective area of antenna is A_e as shown in Fig. 16.5, the maximum power, W, taken from a receiving antenna is given by

$$W = PA_e \tag{16.12}$$

The effective area A_e is given by using antenna gain G,

$$A_e = \frac{\lambda^2}{4\pi} G \tag{16.13}$$

where λ is wave length.

Examples of effective area of simple antenna are shown in Table 16.1. In addition, on the aperture antenna such as a horn antenna, ratio of actual aperture area A to effective area A_e,

$$\eta = \frac{A_e}{A} \tag{16.14}$$

is called the gain coefficient or aperture effectiveness. This is often also referred to as the efficiency of the antenna. From (Eq. 16.13) and (16.14), the antenna gain G having actual aperture area A is given by

$$G = \frac{4\pi}{\lambda^2} \eta A \tag{16.15}$$

If the actual aperture is a circle of diameter d, the antenna gain G can be given as

$$G = \left(\frac{\pi d}{\lambda}\right)^2 \eta \tag{16.16}$$

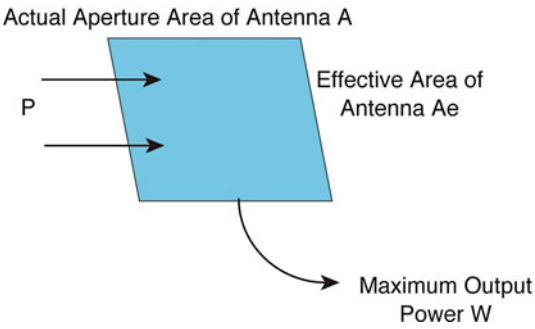


Fig. 16.5 Effective area of antenna

Table 16.1 Example of effective area of antenna

Antenna	Absolute gain	Effective area
Omnidirectional	0 dBi	$0.0796\lambda^2$
Half-wave dipole	2.15 dBi	$0.13\lambda^2$

In the case of a parabolic antenna that is popular for satellite communication, the aperture effectiveness or efficiency is typically in the range of a low of 0.5 to a high of 0.7 or 0.75.

Polarization

The electric field and magnetic field of electromagnetic wave are directed toward a specific direction. This directional characteristic is called polarization. In the field of satellite communications either linear or circular polarization is used to expand the effective use of the limited spectrum that is made available for satellite communications. By distinguishing between polarized signals by filters made for this purpose one can reuse the same frequencies. Thus one can use horizontally and vertically "separated" signals to reuse the same spectrum or one can distinguish between right-hand and left-hand circularly polarized signals.

Linear polarization involves discrimination between "wanted" and "unwanted" vertical and horizontal signals based on perpendicular coordinates. In this case, the electric field or wave form perpendicular to the ground is called vertical polarization. In contrast the electric field or wave form, horizontal with the ground is called horizontal polarization. If a wave of horizontal polarization is combined with equal strength wave of vertical polarization by 90° phase difference, the combined wave has circular polarization whose electric field is rotated. Figure 16.6 shows the circular polarization. The circular polarization consists of right-hand circular polarization and left-hand circular polarization depending its direction of rotation. As shown in Fig. 16.6b, the right-hand circular polarization is defined as that the rotation direction of electric field is clockwise looking from the opposite to the

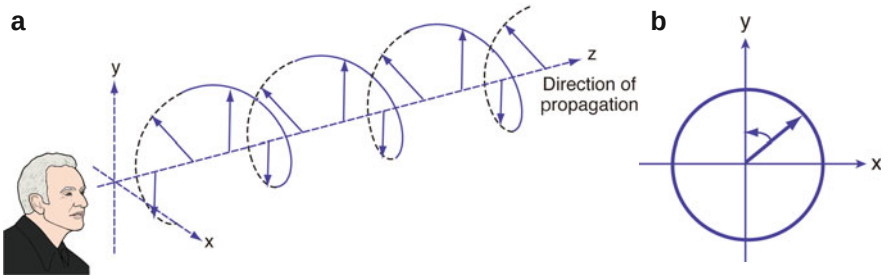


Fig. 16.6 Circular polarization. (**a**) Propagation of circular polarized wave plane (Left circular polarization) (**b**) Circulation of polarized looking from −z axis (Right circular polarization)

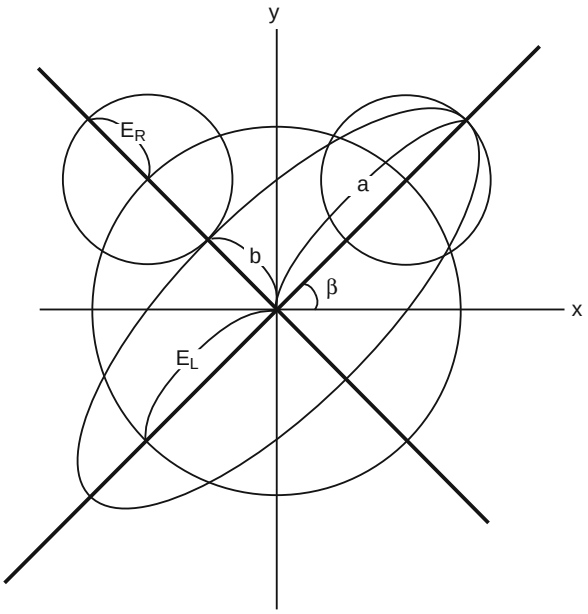


Fig. 16.7 Elliptical polarization

wave forward direction (−z), the left-hand circular polarization is defined as being vice versa. Figure 16.6a shows the left-hand circular polarization.

Generally speaking, since it is difficult to adjust the strength of electromagnetic waves to be equal or to adjust the phase to be 90°, the circular polarized wave has elliptical polarization. As shown in Fig. 16.7, the elliptical polarized wave can be decompose into the component E_R of right circular polarization and component E_L of left circular polarization. These components can be used to define the following equations.

Ellipticity:

$$r = \frac{a}{b} = \frac{E_L + E_R}{E_L - E_R} = 20\log_{10}\left|\frac{E_L + E_R}{E_L - E_R}\right| \quad [dB] \tag{16.17}$$

Circular polarization ratio:

$$\rho = \frac{E_L}{E_R} \tag{16.18}$$

Cross polarization discrimination:

$$XPD = 20\log_{10}|\rho| \tag{16.19}$$

Where the relationship between r and ρ can be given as follows:

$$\left.\begin{aligned} r &= \frac{\rho+1}{\rho-1} \\ \rho &= \frac{r+1}{r-1} \end{aligned}\right\} \tag{16.20}$$

The ellipticity is also called axial ratio. The angle of major axis tilted from reference axis (X axis) is called tilt angle β as shown in Fig. 16.7.

Basic Antennas

Linear Wire Antenna

A dipole is a representative of a linear wire antenna and its structure is shown in Fig. 16.8. It is called a half-wavelength dipole whose length is $\lambda/2$. Its electric field pattern, E, is given by

$$E = E_0 \cos((\pi/2)\cos\theta)/\sin\theta \tag{16.21}$$

where E_0 is the maximum strength of electric field. Figure 16.9 shows the antenna radiation power pattern, E^2, of half-wavelength dipole displayed in three-dimensional indication. Its gain, G_d, is given by

$$G_d = \frac{1}{\int_0^{\pi/2} \frac{\cos^2\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta} d\theta} = 1.64 \quad (= 2.15 \quad dB\,) \tag{16.22}$$

The half-power beam width is 78°. It is omnidirectional pattern whose gain is equal in all directions when it is taken of slice of the pattern with a plane normal to the dipole.

Horn Antenna

A horn antenna is often used as a primary emission element of a reflector type antenna. In addition, it is used as an antenna itself when wide beam width is

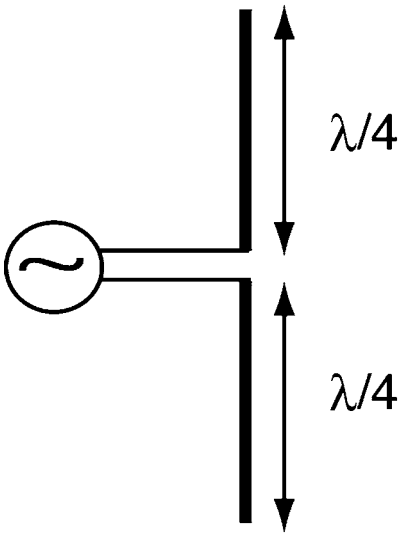


Fig. 16.8 Structure of dipole antenna

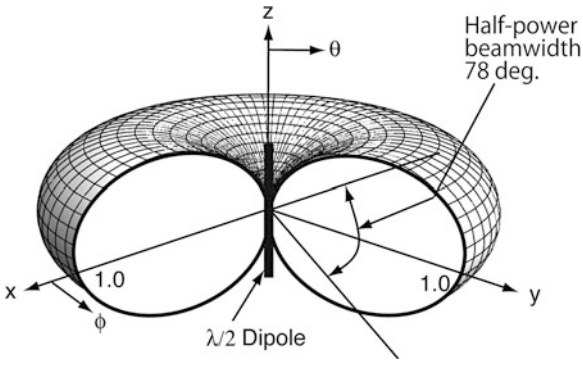


Fig. 16.9 Antenna pattern of half-wavelength dipole (3 dimensionl image)

necessary. The horn antenna is often used as satellite-borne antenna since it has a proper beamwidth for global coverage that intended to look at whole earth from a geostationary satellite whose look up angle is about 18°.

The horn antenna is categorized roughly into two kinds of pyramid horn that widened rectangular wave guide and a conic horn that widened circular wave guide as shown in Fig. 16.10. Since the theoretical gain of a horn antenna coincides well to actually measured gain and it is strong structurally, it is often used as a reference antenna for measuring the gain of various antenna in the microwave frequency.

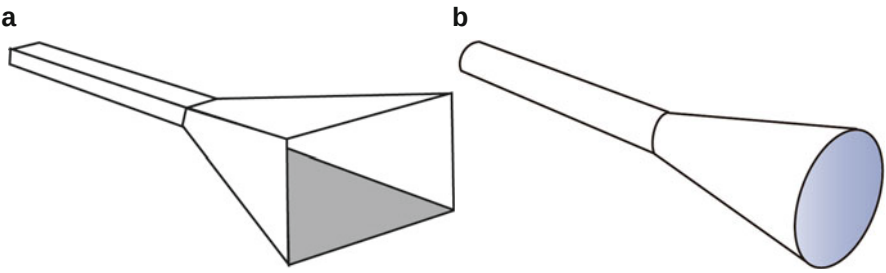


Fig. 16.10 Horn antenna. (**a**) Pyramid horn antenna (**b**) Conic horn antenna

Reflector Antenna

The representatives of reflector antenna are:

- Parabolic antenna
- Cassegrain antenna
- Offset parabolic antenna
- Offset Cassegrain antenna

The reflector type antenna is often used as satellite communication as well as broadcast satellite receive antenna. A reflector antenna realizes high gain and low side lobe by converting the spherical wave radiated by a primary radiator to the plane wave. It is for these reasons that reflector antennas have become the predominant type of antenna used in satellite communications.

Parabolic Antenna

Structure of a parabolic antenna consists of a parabolic reflecting surface and a primary feed at a focus as shown in Fig. 16.11. Since the parabolic surface is a paraboloid of rotation, sum $(l_1 + l_2)$ of the distance l_1 from a focus to the reflector and distance l_2 from the plane normal to the antenna axis to reflecting point is constant, the spherical wave radiated at a primary feed put at focus is converted to the plane wave at the reflector. A horn antenna is used for the primary emission device. The gain of parabolic antenna has been previously given by (16.16).

Cassegrain Antenna

A Cassegrain antenna is a dual reflector antenna which consists of a parabolic surface as a main reflector and a hyperboloid of revolution as a sub-reflector as shown in Fig. 16.12. Among two foci of a sub-reflector, one accords with the phase center of primary feed and another accords with a focus of the main reflector. The sub-reflector works as a converter of spherical waves for primary feed and main reflector and the main reflector is a converter between spherical waves and plane waves.

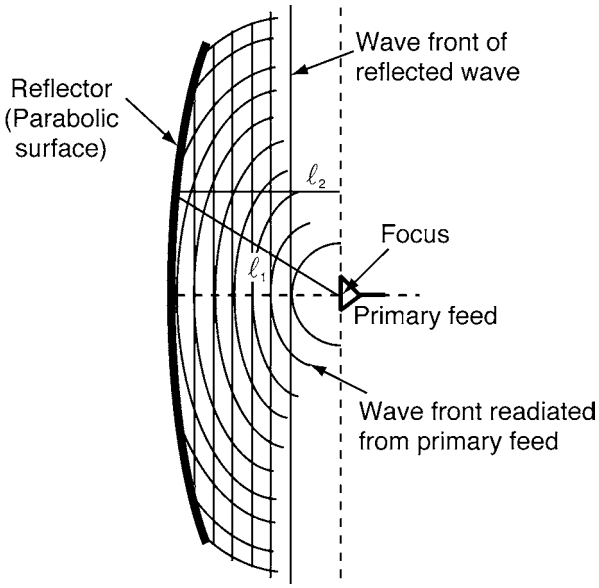


Fig. 16.11 Principle of parabolic antenna

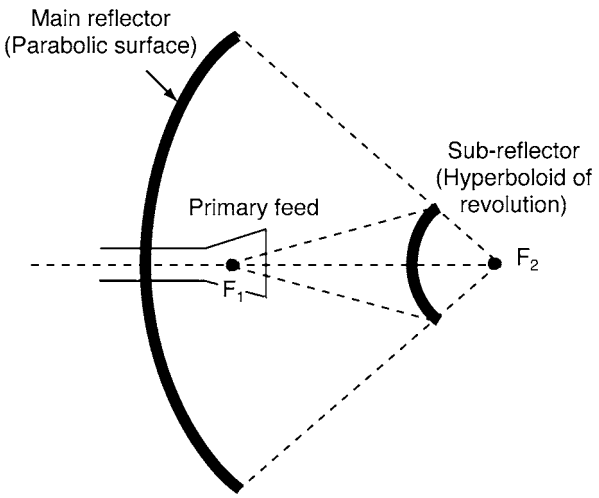


Fig. 16.12 Cassegrain antenna

The characteristics of Cassegrain antenna that is different from a parabolic antenna is that it can accommodate a low noise amplifier at the back of the main reflector. Also there is a margin to shorten the length of wave guide feed and thus to decrease the transmission loss. A Cassegrain antenna is used as an earth station antenna for satellite communications widely.

Offset Parabolic Antenna

Both parabolic antennas and Cassegrain antennas have a main reflector with a symmetrically rotated parabolic surface. This leads to the need for props supporting the primary feed and sub-reflector and these must be positioned directly in front of the main reflector. This causes emission characteristic deterioration because of blocking the incoming and outgoing electric waves, increase of side lobes, and decrease of gain. In order to avoid these obstacles, an antenna that sets the primary feed and sub-reflector outside of the aperture is an offset parabolic antenna. This is accomplished by using only a part of a parabolic surface as a reflector.

An offset parabolic antenna is shown in Fig. 16.13. A low side lobe is possible to be established for this antenna because of its design architecture.

Helical Antenna

Figure 16.14 shows a helical antenna whose structure has a wounded conductor (helix conductor) in front of a reflector and a feeding point on the conductor. As for this antenna, the electromagnetic wave is radiated in an axis direction when circumference length is 0.75–1.33 wavelength and pitch of spiral is 0.1–0.5 wavelength. When the reel number and the full length are increased, the antenna gain increases. In addition, when circumference length is small in comparison with a wavelength and the full length is around a wavelength, the electromagnetic wave is radiated in the direction perpendicular to the axis (i.e., an axis mode of operation).

Microstrip Antenna

A microstrip antenna has recently received a good deal of attention primarily in terms of user antennas. This approach has been applied to antennas for automobiles, receiving antenna of broadcast satellites and aircraft borne antenna.

Characteristics

The characteristics of the microstrip antenna are:

- Thin structure
- Light weight
- Simple structure
- Easy productivity
- Easy accumulation with semiconductor circuits
- Comparatively high gain for a simple antenna (about 7 dBi)
  This antenna has the possibility to become widely used.

The gain of a microstrip antenna generally increases with the decrease of the dielectric constant. A deficiency of microstrip antennas, however, is the narrow frequency bandwidth it affords. So currently research is being directed at increasing the effective frequency bandwidth by various methods.

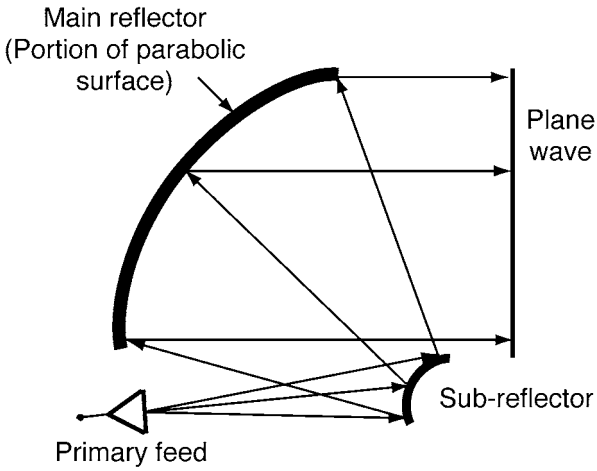


Fig. 16.13 Offset parabola antenna

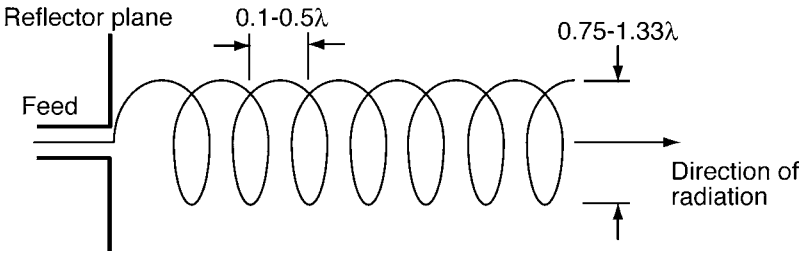


Fig. 16.14 Helical antenna

Rectangular Microstrip Antenna

It is generally considered that the resonance device formed as a rectangular-shaped open-type plane circuit on a thin dielectric substrate has low Q of resonance due to a loss of emission as shown in Fig. 16.15. The antenna that used this emission loss positively is a microstrip antenna. Structure of resonance device is suggested with a circle, a triangle, and a pentagon other than the rectangular mentioned above in various ways.

The working principle of the antenna is shown in Fig. 16.15. The electromagnetic wave is emitted by leaking of an electric field formed at the border of the microstrip antenna. For example, radiation to the front direction of antenna is conducted by the leaking of an electric field at right and left side of an antenna element. The leaking electric field at the top and bottom side of the antenna does not contribute the radiation due to drowning out each other.

A resonance frequency is approximately given by

$$f = \frac{c}{2(d + t/2)\sqrt{\varepsilon_r}} \tag{16.23}$$

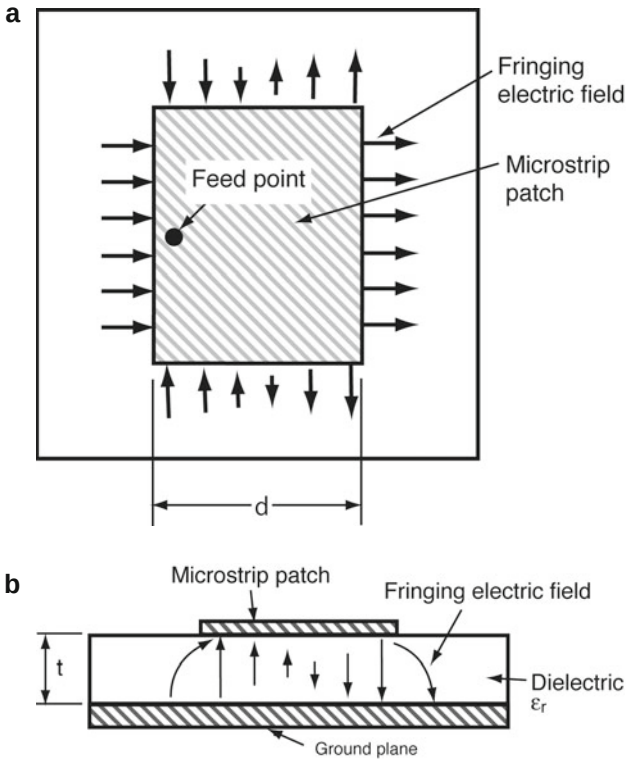


Fig. 16.15 Rectangular microstrip antenna. (**a**) Front view (**b**) Side view

where c is velocity of light, d is given in Fig. 16.15a, t is the thickness of the substrate, ε_r is the dielectric constant of substrate.

Circular Microstrip Antenna

The structure of the circular microstrip antenna is shown in Fig. 16.16. The resonance frequency is given approximately by

$$f = \frac{1.841c}{2\pi\{a + (t/\pi)2\ln 2\}\sqrt{\varepsilon}} \tag{16.24}$$

where c: light velocity, a: see Fig. 16.16, t: thickness of substrate, ε_r: dielectric constant of substrate.

The maximum gain is obtained at the front direction and a single direction pattern. As for a feeding system of microstrip antenna, there are microstrip feeding, pin feeding, and their combination (Iida 2000).

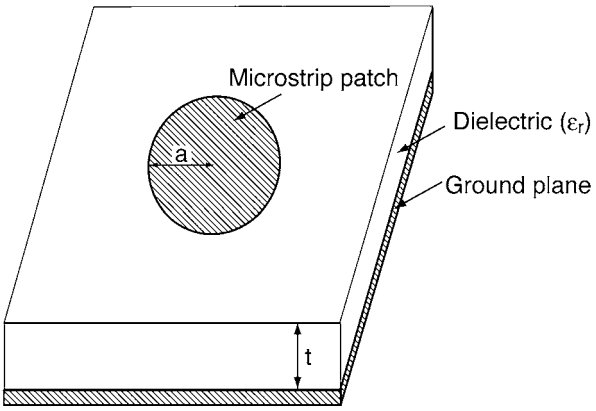


Fig. 16.16 Circular microstrip antenna

Array Antennas for Scanning and Hopping Beams

Function of Array Antenna

The antenna system which deploys an "array" of the same type of antennas is called an array antenna. In this case, an arranged antenna in the array is called an antenna element. The array antenna can have various kinds of antenna functions that cannot be conducted by a single emission element. The performance of an array antenna system depends on the kind of antenna elements, the arrangement method and the way radiation is accomplished. The directivity of an array antenna is important. The particular method that is created to optimize directivity performance is called the directivity composition. When the combined directivity of the antenna elements are set for the array, the combined directivity is generally given by

$$\begin{aligned}(\text{Combined directivity}) = &(\text{Directivity of antenna element}) \\ &\times (\text{Arrangement directivity of omni-directional antenna})\end{aligned} \tag{16.25}$$

This is one of the biggest functions of an array antenna. The array antenna with the following function is obtained by changing drive amplitude and phase of each antenna element of an array antenna.

- To get desired directivity
- To change width of main beam of emission directivity
- To suppress side lobes and to control their level
- To specify zero points of the emission directivity
- To get a desired gain

Another characteristic of an array antenna is possible to scan the main beam of emission directivity and do so three-dimensionally. Fixing an arrangement of an antenna element, the main beam can be pointed to an arbitrary direction of space by changing driving phase of each antenna element. This is called a phased array antenna.

Directivity of Array Antenna

In an array antenna, all of the elements are excited simultaneously by dividing the feeding power. In Fig. 16.17, the directivity of array antenna $D(\theta, \phi)$ is given when N antenna elements arranged by equal space of d are derived by amplitude I_n and phase ϕ_n,

$$D(\theta, \phi) = g(\theta, \phi) \sum_{n=1}^{N} I_n e^{j\{\phi_n + (n-1)kd \sin\theta(\cos\phi + \sin\phi)\}} \tag{16.26}$$

where $k = 2\pi/\lambda$, $g(\theta, \phi)$ is directivity of a single antenna element. This $g(\theta, \phi)$ depends on the antenna element. But the term of Σ is same, namely the term of Σ indicates the characteristics of array antenna. This is called array factor.

Gain of Array Antenna

It is often used that many antenna elements are arranged to increase a gain of an array antenna. Supposing that there is no interaction at all between each antenna elements, the gain of an array antenna, G_n, is n times of gain of an antenna element, G_e, in the case where the number of the antenna elements is n.

$$\frac{G_n}{G_e} = n \tag{16.27}$$

In other words, the gain increases in proportion to the number of antenna elements. However, actually, since there is mutual combination between each antenna elements, the gain does not become n times. But (Eq. 16.27) is used as an aim of a gain of array antenna well.

Phased Array Antenna

Function of Phased Array Antenna

A phased array antenna is the antenna which changes feeding phase of each antenna element electronically, and can scan the main beam of emission directivity three-dimensionally. In mobile satellite communications, it is used as an antenna of a vehicular side on the ground. The main beam is changed its direction according

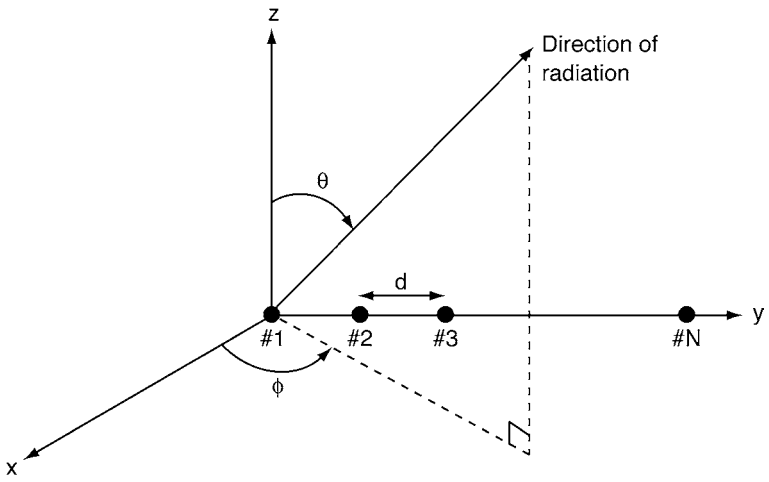


Fig. 16.17 Coordinate system of array antenna

to the direction of satellite to prevent the link from disconnecting when a position of a vehicle changed. It is also used as an antenna of the geostationary satellite (a relay satellite) side for inter-satellite communications.

Configuration of Phased Array Antenna

A configuration of phased array antenna is shown in Fig. 16.18. Phase shifting device is installed at every antenna element, and it is controlled from its outside. The operation of the phased array antenna is explained for the reception of signal.

The equiphase plane (wave front) of incident wave from a certain direction at the Nth (#N) antenna element propagates a distance $(N-1)d\sin\theta$ from the wave front to the first (#1) antenna element. The phase of signal received at #N element advances $k(N-1)d\sin\theta$ during this propagation, where $k = 2\pi/\lambda_g$, λ_g is wavelengths in a feeding circuit. The signal at the #N element is delayed by a phase shifter. Namely, the phase shifter is set to $-k(N-1)d\sin\theta$ so that the phase difference with the signal received by the #1 element is zero. By doing such a processing for all of each antenna element, and incident signal can be strengthened by calculating the optimum output.

Multibeam Antennas with Multiple Feed Systems

Function of Multibeam Antenna

A multibeam antenna is an antenna that has plural beams, plural input, and output terminals and to be able to transmit plural independent information as shown in Fig. 16.19. It enables increase of a gain by a spot of the beam, frequency reuse by

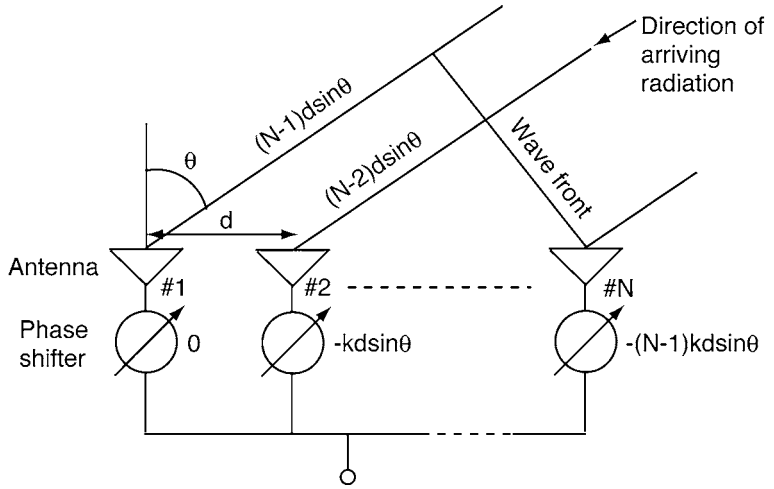


Fig. 16.18 Configuration of phased array antenna

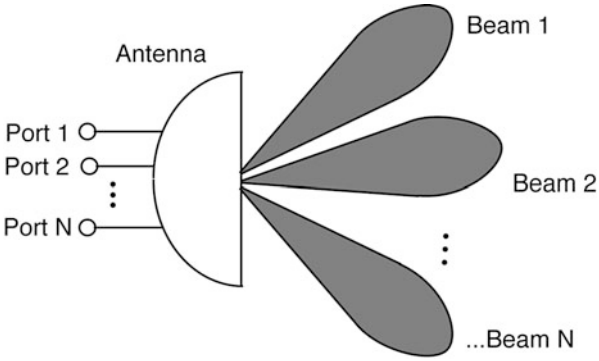


Fig. 16.19 Concept of multibeam antenna

the space division of the beam. Let's consider why a multibeam antenna is necessary for a mobile satellite communication, although it can also be quite effectively used for fixed satellite communications as well.

- Since it is difficult for a mobile vehicle to install a large-sized antenna and/or high-powered transmitter generally, an antenna of a big gain is necessary at the satellite side, but the aperture area A_e must be big to raise gain G from the equation indicated between gain and aperture area, (16.13). As for the antenna, the bigger the aperture area A_e is, the narrower the beam width and it becomes spot beams from relationship between aperture diameter d and beam width Eq. (16.16). In the case of mobile satellite communications one needs not only high-powered spot beams to communicate with small user terminals, but also a lot of them in order to cover a wide service area. Thus, the solution is a very large aperture antenna with a multibeam feed system.

- A demand of many spot beams is satisfied by preparing for many single spot beam antennas. But many large antenna cannot be embarked on a satellite due to constraint of both weight and space, thus the multibeam antenna which can emit an independent spot beam of a plural number from an antenna is necessary.

Type of Multibeam Antenna

The multibeam antenna is categorized into the following:
- Reflector type
- Array type
- Reflector + array type

Reflector Type
In this case, multiple primary feeds (usually, horn antennas) are installed in the neighborhood of the reflector's focus. This reflector-type antenna architecture allows high performance for few beams and with a simple structure. But when the number of primary feeds increases, a gap between the primary focus grows too wide, and the performance characteristics deteriorates. For this type of satellite antenna and feed system, offset parabolic antenna and offset Cassegrain antenna are employed.

Array Type
An array antenna has a characteristic to be able to operate with a high degree of directivity. This is accomplished by changing the placement and phase of each element.

Reflector + Array Type
This is a multibeam antenna that deploys a conventional parabolic reflector but then uses an array type multibeam antenna at a focus as primary feed. There is not deterioration of each beam with this type of antenna and this type of design can serve to make the feeding circuit loss small. This can also be the most cost-effective solution as well.

Antennas for Optical Communications Systems

It is necessary to perform acquisition, tracking, and pointing (ATP) to establish optical satellite communications (ATR 1995). The ATP optics includes an optical antenna (a telescope), an acquisition sensor, but also a rough coarse tracking sensor, a precise fine tracking sensor, and the mechanical elements are two-axis gimbals to control the antenna pointing direction drive device, and a fast steering mirror actuator. These sensors and actuators form a dynamic system to assure link stability in transmitting and receiving a very narrow optical beam.

Figure 16.20 shows a Cassegrain telescope as a typical optical antenna. Usually, a primary mirror is constructed made within paraboloid shape and a secondary

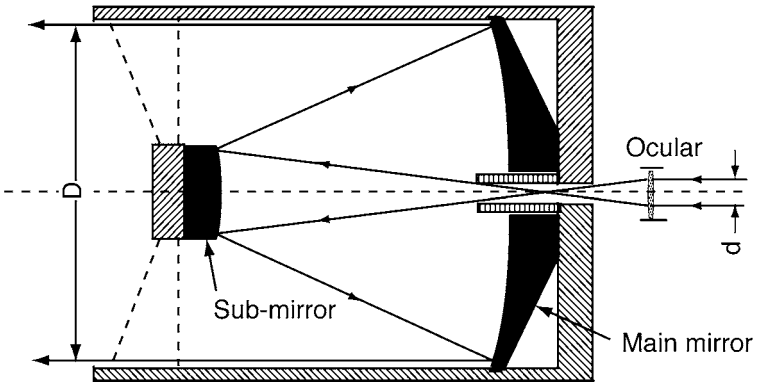


Fig. 16.20 Optical cassegrain antenna

mirror with hyperboloid shape, and the laser light is lead to an internal optical system from through a hole which is made in the center of the primary mirror. If the aperture diameter of the main primary mirror (antenna) is D and, the diameter of the internal beam is d, and the magnification is defined as $M = D/d$, the precision requirement for the internal optical components and their alignment accuracy can be relaxed at the internal optics, because the angle deviation at the part of antenna aperture is magnified by M. However, the precision of lower than 1 μm is still necessary for the optical antenna to be used for inter-satellite links. Therefore, a material with low coefficient of thermal expansion coefficient, such as invar (iron-nickel alloy) or zerodur (glass ceramic composite material) has to be used.

The bigger the aperture diameter of antenna is, the bigger the antenna gain will be, but it is necessary to be careful tracking accuracy accordingly because an antenna pattern becomes sharp. For example, a tracking accuracy of less than 1 μrad is necessary so that an antenna gain of 30 cm in diameter is made available. In addition, the transmitting laser beam can be approximated as a Gaussian beam that has a flat wavefront and a Gaussian amplitude distribution. This Gauss beam degrades its energy/power because the outskirts of the beam are truncated by the finite primary mirror size or a body tube and the beam center is obstructed by the secondary mirror. This truncation and obstruction will limit the aperture efficiency of the antenna. In addition, wavefront error due to optical aberration or mirror surface inaccuracies will cause another degradation which is expressed with the quantity of the "Strehl ratio." It is necessary to increase maintain the wavefront error less than 1/10 wavelength to keep the Strehl ratio better than −1.7 dB.

Conclusion

It is clear from the above descriptions of satellite antennas that there are a wide range of antenna types that can be used in satellite communications. Despite this diversity of antenna designs, parabolic reflectors with different types of feed systems, including phased array feed systems, are the most common satellite antenna systems today. This is because of the ability to achieve higher capacity,

higher gain beams, lower powered side lobes that create interference to other satellites and ground based communications systems. Broadcast systems generally need more power to work to small dishes. Mobile satellite systems have evolved higher and higher spot beams using multibeam antenna technology. Fixed satellite service satellites have employed similar satellite antenna technology as well. In the next chapter, the implementation of the antenna technology by satellite communications around the world will be addressed in more detail.

Acknowledgment The author thanks Dr. Joseph N. Pelton of ISU, Dr. Masato Tanaka of NICT, and Dr. Yoshinori Arimoto of NICT for their support to organize this chapter.

References

ATR, *Optical Inter-satellite Communication* (Ohmsha, Tokyo, 1995). in Japanese
C.A. Balanis, *Antenna Theory – Analisis and Design* (Wiley, New york, 1997)
T. Iida, *Satellite Communications – System and Its Design Technology* (Ohmsha/IOS Press, Tokyo/Amesterdam, 2000)

Satellite Antenna Systems Design and Implementation Around the World

17

Takashi Iida and Joseph N. Pelton

Contents

Introduction 398
Design Factors Driving the Development of Satellite Communications Antenna Systems 399
The Need to Effectively Achieve More Useable Spectrum 399
Techniques for Improvement of Satellite Throughput 403
Differences in Satellite Antenna Designs for GEO, MEO, and LEO Orbits 403
Technical and Economic Challenges in Designing Satellite Antennas 408
The Evolution of Satellite Antenna and Communications Systems 409
Phase One: The Earliest Phase of Satellite Communications with Omni Antennas 410
Phase Two: Three Axis Stabilization and Higher Gain Satellite Antennas 410
Phase 3: The Creation of Higher Gain Parabolic Reflector on Communications Satellites 411
Phase 4: The Advent of Body-Stabilized Communications Satellites 412
Phase 5: Service Diversification and Alternative Satellite Antenna Design 413
Future Satellite Antenna Technology 416
Phased Array Antenna for the Quazi-Zenith Satellite System (QZSS) 419
Large Deployable Antennas for Mobile Services 420
The Antennas of Terrestar and Light Squared 421
Optical Communications Systems 422
Optical Antenna of OICETS Satellite 423

T. Iida (✉)
Tokyo Metropolitan University, 1-23-4 Mejirodai, Hachioji, Tokyo, 193-0833, Japan
e-mail: QZF04134@nifty.com

J.N. Pelton
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: joepelton@verizon.net

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_19, © Springer Science+Business Media New York 2013

European Optical Antennas 425
Improved Large-Scale Satellite Systems with Large Reflectors with Phased Array Feeds and Nano Satellite Arrays 425
Notes 426
References 426

Abstract

In this chapter, the objective is to discuss the practical implementation of various types of satellite antenna designs over time and to indicate the current state of the art and future trends to develop even higher gain satellite antennas with greater efficiencies in terms of frequency reuse or higher capacity FSS or MSS type satellite systems. Although there continue to be smaller satellites that are launched for communications purposes, the antenna designs utilize the same technologies and concepts that are employed in larger scale satellites.

The evolution of antennas for satellite communications has generally conformed to the following historical pattern:

Low gain omni and squinted-beam antennas

- Increased gain types of satellite antennas (horn type and helix antennas)
- Parabolic reflectors (including multi-beam antennas with multiple feed systems)
- Deployable antennas (particularly for achieving more highly focused beams and support much high-gain multi-beam antennas
- Phased array feed and Phased array antennas
- Scanning and hopping beams
- Optical communications systems (initially for Intersatellite links and interplanetary communications, but this type of technology might possibly be used for Earth to Space systems in the future as well).

Examples of many of these types of satellite antennas will be presented in the following chapter. But first the factors that have led engineers to design improved and higher performance antennas will be discussed and examined.

Keywords

Deployable Antenna • High-Gain Antenna • Horn Antenna • Isotropic Antenna • Mobile Satellite Services (MSS) • MSS with Ancillary Terrestrial Component (ATC) • Multi-Beam Antennas • Multi-Feed Systems • Off-Set Feed Antenna • Omni Antenna • Orbits-Geo • Meo and Leo • Parabolic Reflector • Path Loss • Phased Array Antenna • Polarization • Three Axis Body and Spin Stabilized Spacecraft

Introduction

In the preceding chapter the basic concepts related to satellite antenna patterns, interfering sidelobe transmissions, gain, linear and circular polarization, as well as different types of antennas with improved efficiency of performance. This

discussion explained how satellite antenna designs have evolved to be more effective in producing higher gain, more effective re-use of available frequencies, etc. In this chapter the objective is to discuss the practical implementation of these different types of satellite antenna designs over time and to indicate the current state of the art and future trends to develop even higher gain satellite antennas with greater efficiencies in terms of frequency reuse or higher capacity FSS (Fixed Satellite Services) or MSS (Mobile Satellite Services) type satellite systems.

Design Factors Driving the Development of Satellite Communications Antenna Systems

There are many factors that drive the development of new and improved satellite antennas. These include the need to reuse frequency bands because of limited spectrum allocations, the need to have antennas that can operate at higher frequencies with higher bandwidth, and the desire to deploy higher gain antennas while minimizing the required mass and volume.

The Need to Effectively Achieve More Useable Spectrum

Perhaps the prime design factor that has driven the research and development to design improved satellite antennas has been the need to make more efficient use of the available allocated spectrum for satellite communications. Today, fiber optic networks within closed cables have access to almost unlimited spectrum within the optical frequencies that have very broad spectra available. Satellites, which must transmit their signals within the open environment and largely within the UHF, microwave, and millimeter wave rf bands have much more limited spectra to utilize. Thus, there is a continuing need to develop satellite antenna technology that can use the available spectrum ever more efficiently. This primarily means finding ways to reuse the same spectrum many times over where ever possible.

Frequency Reuse Concepts

As just noted the radio frequencies available for satellite communications are very limited resources. Effective utilization of this sparse resource, namely the band allocated to various satellite services, continues to be very critical to meet every rising service needs. There are fortunately now a number of frequency reuse technologies made available via antenna-related technologies that allows multiple types of frequency reuse in many types of communications satellites. Frequency reuse means to use the same frequency repeatedly. Reuse of the same frequency, however, usually causes interference to others trying to reuse the frequencies as well. If the same frequency is used in geographic areas that are sufficiently removed from one another then such reuse becomes technically possible. Another method, separate from spatial separation, is to use different types of polarization techniques to separate the signals. This is, in a way, similar to the techniques used to separate "wanted and unwanted" incoming light such is accomplished by wearing polarized sun glasses. The prime methods of frequency reuse are thus either to separate antenna

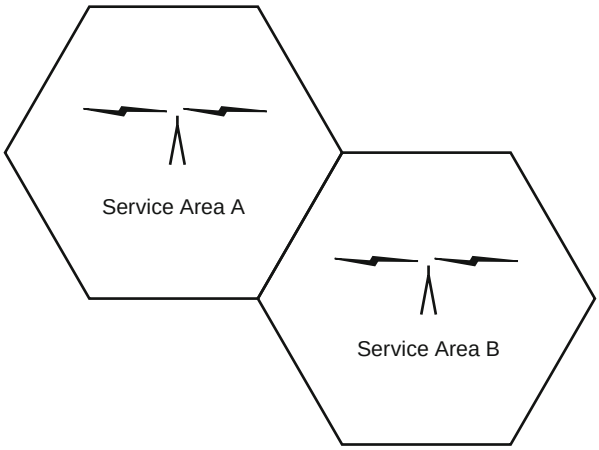


Fig. 17.1 Frequency reuse patterns using geographic separation of cells

beams spatially (i.e., to transmit narrow spot beams that use the same frequencies to transmit signals to different geographic areas that are widely separate from each other) or to utilize orthogonal polarization of electromagnetic waves.

If a certain service area A and another service area B are located separately in the case of a terrestrial cellular communication system as shown in Fig. 17.1, the same frequency can be used over again.

The number of times that the same frequency can be effectively utilized can be increased with improved digital technology. This can be accomplished by dividing the cells into ever smaller service areas. In the case of satellite communications, the service area cannot be decreased to areas as small as the cells used in the terrestrial communication systems. This is because the service area is determined by antenna beam width of satellite. Since the satellite is in Earth orbit well away from the ground, the satellite beams spread much further than a terrestrial cellular beam.

However, the number of frequency reuses can be increased by squeezing the satellite beam width as small as possible. In this case, CIR (Carrier to Interference Ratio), namely a ratio of interference power to carrier power, is a criterion for determining the degree of frequency reuse that can be reasonably obtained. If CIR is maintained more than 10 dB (i.e., the wanted signal is 10 times "clearer" than the unwanted signal), practical use can be possible, especially in the case of digital communication that are more tolerant of interference levels.

The second method that can be used is, of course, the use of polarization. In this respect, there are two options – linear or circular polarization. In the case of linear or orthogonal polarization, the polarizer creates one signal that is irradiates in a horizontal plane and the other signal is irradiated in the vertical plane so the two signal can be clearly differentiated from one another when received. The other polarization alternative is the case of right hand and left hand circular polarization where the two signals are distinguished from one another by either rotating in right hand direction or rotating in the opposite left hand circular direction. CIR discrimination, from 27 to 35 dB, can be obtained by using polarization. Orthogonal discrimination is less expensive but does not provide quite the same

CIR discrimination as circular polarization. The ratio of a component of the polarization to its opposite one is called XPD (Cross Polarization Discrimination), and it is usually expressed in decibel (dB). The degradation of XPD due to rain fall must be considered in devising a communication link design (Iida 2000). In practical application, XPD must be within a regular value, namely, the direction of antenna and polarization must be adjusted. This might be as much as 27 dB for a VSAT (Very Small Aperture Terminal) and perhaps more than 35 dB for the other larger and higher performance earth stations.

The Migration from Lower Frequencies Bands to Higher Frequencies VHF, UHF, SHF, and Now EHF

The spectra used for communication satellites have persistently migrated upward many orders of magnitudes in terms of the type of radio frequencies used for services over the past half century. Frequencies have increased from Very-High Frequencies (VHF), to Ultra-High Frequencies (UHF), to Super-High Frequencies (SHF), and now to Extremely High Frequencies (EHF). The amount of spectra that is available in the Very-High Frequency (VHF) and Ultra-High Frequency (UHF) bands is only a few Mega Hertz. This is simply due to the laws of physics. There is less bandwidth available at these lower frequencies. Further, there are also many other competing uses for these bands from terrestrial services such as mobile communication via cellular networks as well as for radio and television. This competing use leads to problems of interference.

The solution for satellite communications has been to migrate upward to the higher frequency allocations in the microwave and millimeter wave bands. In the Super High Frequency (SHF) band, that is, from 3 Giga Hertz (GHz) to 30 Giga Hertz (GHz) and the Extremely High Frequency band, that is, from 30 to 300 GHz, there are many broadband frequency allocations of "rf" spectra for satellite communications. There is a 500 MHz allocation in the C band (6 GHz uplink and 4 GHz downlink) for Fixed Satellite Services (FSS), plus another allocation in X-band for defense related communications (i.e., 8/7 GHz), plus another 500 MHz in the Ku band (14/12 GHz). Further, there is a very wide allocation of 1,000 MHz allocation in the Ka band (30/20 GHz). There is also a major allocation for the downlinking of direct broadcast satellite services at 18 GHz. For the future, there are also significant additional allocations in the Q, V, and W bands at 38, 48 and 60 GHz. For Mobile Satellite Services, there are smaller spectrum allocations in the 800 MHz band, plus allocations around in the 1,600/1,700 MHz bands, in the 2,000/2,100 MHz range and then another allocation in the 2,500 MHz band. In addition, there is the ability to use the aforementioned FSS allocations as feeder links to uplink signals from MSS major earth station feeder networks. A detail listing of the bands that are utilized for FSS, MSS, and BSS services are provided earlier in this Handbook.

The question that might spring to mind is why not just go to the wide frequency bands at the highest frequencies in the EHF bands if there are very wide bands of spectra available there? The problem is there are technical, operational, and financial reasons that argue against such a migration. The lower frequencies require lower cost and easier to manufacture equipment. Further transmissions are more

tolerant of various types of interference and obstacles blocking the transmission path since the longer wave lengths do not have to be direct line of sight to be received. Also at the higher frequencies rain attenuation and other atmospheric attenuation factors make it difficult to send signals when there are adverse weather conditions.

The migration from lower frequencies to higher frequencies has many specific difficulties associated with transitioning to the higher bandwidth allocations. There is the advantage of broader spectrum allocations on one hand but all the other factors in such a transition are adverse to moving up to higher spectra. This is to say that ground and satellite antenna systems and associated electronics are more difficult and expensive to build, rain attenuation can block signals, and any physical interference (such as can be anticipated in mobile communications satellite systems will tend to block the signal).

In addition to all these considerations, there is the issue of required signal power at higher frequencies. The received signal power "C" that a satellite antenna collects through free-space propagation is given by using antenna gain formula as follows (Demers et al. 2009):

$$C = P_T G_T G_R \left(\frac{\lambda}{4\pi d}\right)^2$$

Formula for received signal power C

C is the received carrier power, P_T represents transmitted power, G_T represents antenna gain of the transmitter, G_R represents the antenna gain of the receiver, λ represents the wave length, and d represents the distance between the transmitter and the receiver. The antenna gain G is given by:

$$G = \eta \left(\frac{\pi D}{\lambda}\right)^2$$

Formula for antenna gain

In this formula, G represents antenna gain, η stands for efficiency, D represents the aperture of antenna, and λ represents wavelength.

As can be seen from these equation, the shorter the wave length and thus the higher the frequency, the larger the need for the received signal power, assuming the antenna aperture is constant. Additional power leads to additional design complexity and added cost to the satellite as well as increased launch costs.

In summary, if higher frequencies are to be used for satellite communications, there are the following challenges that must be faced:

- There is the need to develop totally new devices to operate at the very demanding frequencies and increasing small wavelengths.
- There accordingly tends to be an overall increase of cost in general for ground and space systems.
- At sufficiently high frequencies there will also be an increase in rain attenuation and other atmosphere degradations in signal strength.
- Due to rain attenuation and other such atmospheric phenomena, there will also be a need for higher link margin to maintain quality and reliability of service. The proportionate transmit power is also a factor in this regard.

- There is a considerable technical challenge of not only designing but manufacturing low loss antenna systems. Particular challenges include:
  - Necessity of increasing the surface accuracy of antenna
  - The need to have very small mesh in case of mesh structure that is exactly conformal to the parabolic shape to direct the signal to and from the feed system

It needs to be noted that these various challenges increase as one moves to higher frequencies not in a linear fashion, but rather exponentially. Despite these difficulties a great deal of progress has been made through R&D and experimental satellite programs so that operational satellites in the Ka band are now being implemented. Research and development over the past 10 years has now made it possible to design functional fine mesh deployable antenna for commercial use up through to even the Ka band (Demers et al. 2009).

Techniques for Improvement of Satellite Throughput

In light of the difficulties of going to the higher frequencies, a good deal of attention has been focused on ways to use the lower frequencies more efficiently for satellite communication. This literally means increasing the throughput of digital bits per available hertz within the lower band frequency allocations. Not long ago, a typical communications satellite would transmit about 1 bit per hertz. With improved modems and encoding (i.e., coder and decoders known as codecs), it is possible to achieve transmission efficiencies on the order of 2–2.5 bits per hertz. This is equivalent to expand the spectrum allocation by 2–2.5 times. In the future, capabilities in the 4–5 bits per hertz range might be achieved. As always there are tradeoffs and technical difficulties to be overcome. Advanced encoding such as turbo-coding only works well in a relative low noise or interference environment. The use of intensive encoding such as 8 bit or 16 bit encoding can be successful in a low noise and essentially clear sky condition. Thus some advanced encoding systems can be stair stepped down to lower throughput efficiencies in the case of heavy rain and higher levels of rain attenuation.[1]

Differences in Satellite Antenna Designs for GEO, MEO, and LEO Orbits

The path loss characteristic of communications satellites is one of the most critical issues. The spreading out of the signal from the time, it leaves the satellite until it reaches the ground station or user terminal on an aircraft of ship decreases the satellite's telecommunications performance and throughput capabilities. Since a low earth orbit (LEO) satellite can be up to 40 times closer to Earth than a GEO satellite, the relative performance in terms of the flux density can be up to 1,600 times (40×40) if the satellite antennas and power of the GEO and LEO satellite were to be exactly the same. The path loss is of course due the spreading circle of the signal strength. Since the area of a circle is πr^2 then the spread of the signal over the transmission path must be calculated by squaring the height of the orbit.

The varying look angles or masking angles for Leo satellites in a global constellation, however, must be taken into account. When this is consider and the relative maximum transmission distances between a GEO satellite at its longest transmission path is compared to the longest transmission of a LEO satellite, then the relative lengths of the path distances shrinks to about 25 to 1. This means the relative advantage of a LEO satellite to GEO in terms of path loss shrinks to perhaps 625 times (25 × 25). This is a complicated way of saying that a LEO satellite can have on the order of a 28–30 dB advantage over a GEO satellite due to path loss consideration.

As always there are a number of other things to consider here. When it is recognized that three satellites give global coverage from a GEO orbit but it takes on the order of 50 or so LEO satellites to give complete global coverage at all times, then the seeming relative advantage begins to disappear. This type of comparison is also based on the satellite power and antenna systems' gain being exactly the same. It is possible of course to increase the power and add high-gain antennas add a relative 25 dB of transmission gain to GEO satellites. By making this engineering change to a GEO satellite, by adding power and high-gain antennas, the path loss disadvantage can be compensated for especially now that very large high-gain fine mesh deployable antennas can be deployed on GEO satellites. The case of the MEO satellite and its relative path loss is, of course, somewhere in the middle between GEO and LEO satellites. Thus for a MEO satellite, it is a simple matter of calculating the height of the constellation's orbit. Then one again computes the path loss that come from the spreading circle of the signal strength that comes from a medium orbit satellite. The technical specifics of these considerations are provided below.

GEO Systems

The free-space propagation loss L (represented in decibels), is given in the formula below. In this case L is equivalent to P_T/C. The calculation for C (which is the value for received signal power) was provided earlier. For GEO systems $GT = 1$ and $GR = 1$.

$$L = \frac{(4\pi d)^2}{\lambda^2} \text{ or in a more convenient form } L(\text{dB}) = 32.44 + 20\log f(\text{MHz}) + 20\log d(\text{km}).$$

Formula for determining free-space propagation loss (L): Since the distance between a GEO satellite and an Earth station is around 38,000 km, the propagation loss at 20 GHz in the Ka band is about 210 dB. Thus the requirement of link budget is severe. This can be overcome as noted above by adding high power to the GEO satellite and extremely high-gain antennas many meters in diameter.

An example of a GEO orbit satellite communications system is the WINDS experimental satellite is shown in Fig. 17.2.

In the case of the WINDS satellite, a 5 m aperture earth station antenna is needed for the highest broadband speeds of about 1.2 Gbit/s communication, a 2.4 m class can support 622 Mbit/s speeds, while a 1.2 m class antenna allows 155 Mbit/s throughput, and the smallest 45 cm class antennas can still support 1.5–6 Mbit/s speeds (Fig. 17.3).

A transportable class WINDS 2.4 m aperture antenna is shown in Fig. 17.4.

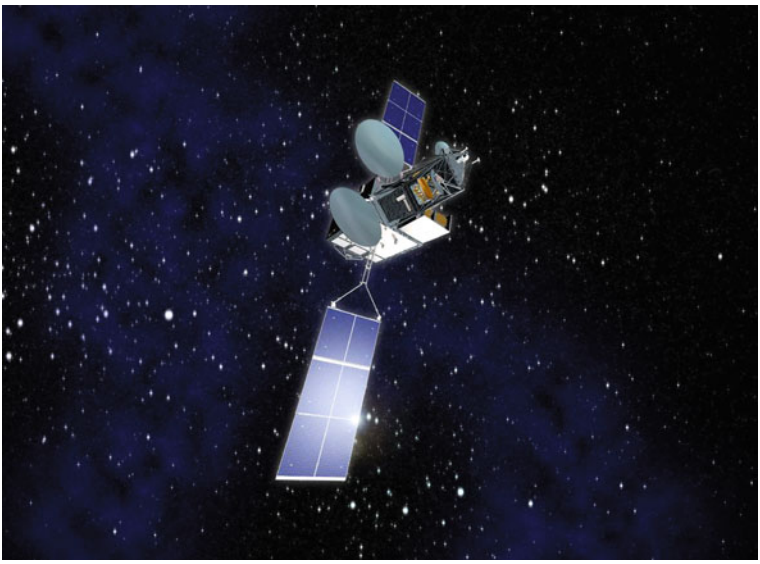


Fig. 17.2 The Ka-band WINDS satellite with conventional and phased array antennas (Graphic courtesy of JAXA)

In addition, GEO orbit systems when utilized for mobile satellite service (i.e., MSS communications) lead two additional considerations that add to the need to larger satellite antennas than is the case for a Fixed Satellite Service (FSS) satellite like WINDS. First of all, lower frequencies are used for mobile communications because radio waves with longer wavelengths are more tolerant of obstacles being in the way such as trees, the roof of a car, or truck, etc. But lower wave lengths mean the antennas must be larger to focus the longer wavelength signals more accurately. Also the receiving antennas that consumers are using must be quite small since they are intended to be completely mobile. Thus, these user transceivers cannot be effectively equipped with a tracking capability. To accommodate the need for small user devices and thus a low gain antenna, the satellite antenna in space, especially from Geo orbit, must be much, much larger to compensate.

In fact, the GEO satellite antennas for mobile communications today can be anywhere from 12 to 20 m in diameter (or in the range of 40–65 ft). In the case of the Engineering Test Satellite-8 (ETS-8) of Japan, the deployable downlink mobile satellite antenna was 17 m by 19 m in size (see Fig. 17.5 to see a graphic of the ETS-8 Satellite and its large deployable antennas for uplinking and downlinking in the MSS bands).[2]

Since the ETS-8 GEO orbit antenna system is so high gain, this satellite is able to work with an S-band hand-held terminal as shown in Fig. 17.6. Its size is 58 mm (W) × 170 mm (D) × 37.5 mm (H) and weight is 266 g (or about a half a pound) without battery. The terminals designed for use with the latest MSS systems such as Thuraya, Light Squared, Terrestar, and Inmarsat (as also pictured in Fig. 17.19 have handsets that are comparable or even smaller in size and include (Fig. 17.6).

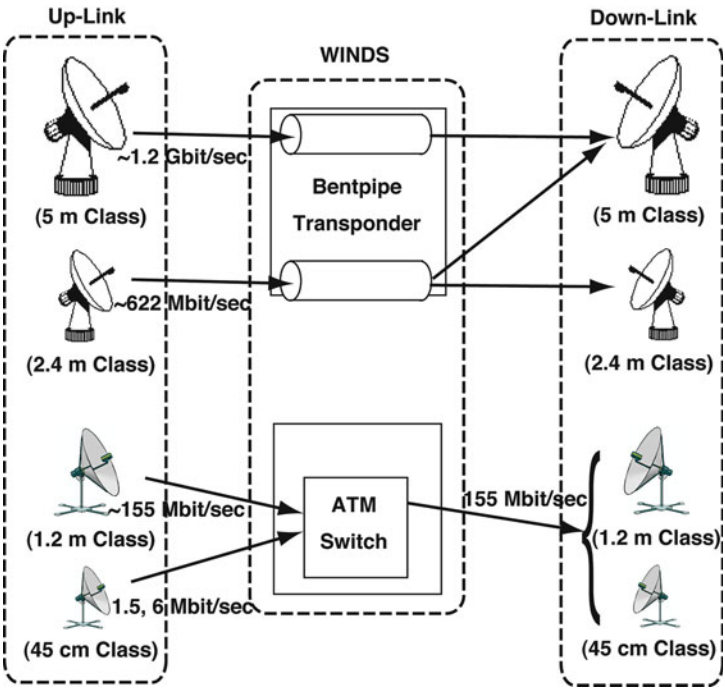


Fig 17.3 The various classes of earth stations required for broadband services (Graphic courtesy of NICT of Japan)

MEO Systems

As for the MEO (Medium Earth Orbit) systems, the distance between a satellite and earth stations is usually in the range of 8,000–20,000 km (5,000–12,500 miles). These MEO orbits are typically chosen to avoid the most intense radiation of the Van Allen Belts. Thus LEO and MEO constellations carefully designed for deployment either below or above the most intense radiation these belts emit.

The propagation loss for MEO orbits is in the range of 189–201 dB in the Ku-band frequency (14 GHz). This propagation loss is smaller than the case of GEO satellite. But it is still very large. Therefore, antennas of both satellite and ground station often tend to have tracking capability. It is important that the MEO satellite antennas, its onboard electronics and its solar cell power array employ protective and shielding against the Van Allen radiation belts that extends in various zones to an altitude beyond 10,000 km. The following point must be considered carefully in terms of designing an antenna for MEO systems:

- Designing an antenna of MEO satellite must consider carefully the optimum orbit. This means both assessing levels of radiation from the Van Allen belts and the increased path loss that higher orbits entail.



Fig. 17.4 A transportable 2.4 m Ka-band satellite that supports 622 Mbit/s (Graphic courtesy of NICT of Japan)

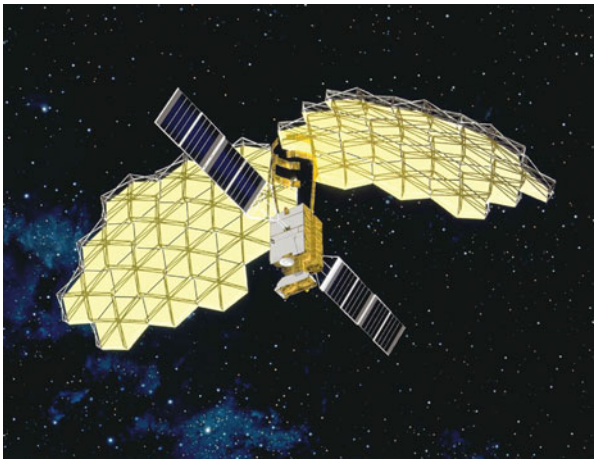


Fig. 17.5 The ETS-8 satellite with large deployable antennas for mobile satellite service (Graphic courtesy of JAXA)

- In addition to the antenna system design, one must also give special consideration to the design and protection of the computer control system plus all of the satellite electronics devices. It also means applying silica coating for the solar cells to prolong their life against the hazards of Van Belt radiation.



Fig. 17.6 (**a**) The hand-held terminal designed for the ETS-8 satellite. (**b**) the new ISat phone pro that operates with the Inmarsat 4 satellites (Graphics courtesy of NICT of Japan and Inmarsat)

LEO Constellations

LEO satellite constellations often are deployed at altitude from about 650–1,200 km (i.e., from about 400–750 miles). The frequency bands utilized for MSS services are typically in the UHF, the L band, or S band. In the case of LEO constellations, propagation path losses are on the order of 170–180 dB. The much smaller path loss in LEO orbits tends to make it easier to design satellite handset for users that do not require any tracking. An example of typical LEO system is shown in Table 17.1 below.

As shown in Table 17.1 the intersatellite link of Iridium system is a unique feature. This capability reduces the number of gateway stations, while Globalstar and Orbcomm need many gateway stations in the world because there are some many satellites in the constellation to be tracked and commanded. Iridium and Globalstar can provide both voice and data communication, while Orbcomm can provide only data communication with store-forward communication. Since the service area depends on both the inclination and altitude of a constellation orbit, communications using Globalstar or Orbcomm will typically suffer from a bigger propagation loss than one using Iridium.

Technical and Economic Challenges in Designing Satellite Antennas

The demands of designing efficient satellite antenna systems to meet the special needs of effective telecommunications from GEO, MEO, and LEO orbits have driven the satellite industry forward over the past 50 years. Overriding the demands to meet the requirements unique to the particular orbits has been the prime desire to design satellite antennas and associated communications electronics systems that were higher in overall performance. This has meant finding ways to design very stable platforms to all the precise pointing of high-gain antennas. It has meant finding ways to build larger and more capable antennas for ever higher frequency

Table 17.1 Typical LEO systems and their characteristics

Name of system		Iridium	Globalstar	Orbcomm
Number of satellites		66	48	34
Altitude		780 km	1,100 km	785 km
Inclination		90°	52°	45°
Intersatellite link (ISL)		Yes	–	–
Store-forward com		–	–	Yes (Business to Business (B2B) data service)
Voice com		Yes	Yes	No
Data com		Yes	Yes	Yes
Latitude of service coverage		Global	±60°	±50°
Frequency band	User link	L band	L band (up), S band (down)	VHF band
	Feeder link	Ka band	C band	VHF band
	ISL	Ka band	–	–
Type of satellite borne antenna		Phased array	Phased array	Cross Yagi
Type of antenna of hand-held terminal		No-tracking and built-in extendable antenna	No-tracking and built-in extendable antenna	Whip

bands that could be constructed of lighter and more cost-effective materials. It has also meant finding ways to design "deployable antennas" that could not only be very high-gain but could be launched within the constraints of the fairings of available rocket launch systems. At the same time, there has been a drive to increase reliability and to support longer lived satellites while driving down the cost of manufacturing and testing of these systems. In all the results are impressive. The most advanced satellite antennas of today when compared to the omni antennas of the very first experimental communications satellites are on the order of a million more capable when examined in terms of gain efficiency and lifetime.

The Evolution of Satellite Antenna and Communications Systems

The following history of the evolution of satellite antenna systems for telecommunications services can be summarized as having the following main elements:

- Design of spacecraft platforms that can stabilize more sophisticated antenna systems and allow more accurate pointing of narrower transmission beams
- Design of higher gain antenna systems
- Design of lighter weight and less massive high-gain antennas
- Design of more reliable antenna systems in terms of test, deployment, and operation
- Design of systems to support intersatellite links
- Design of antenna systems that can operate at ever high frequencies, that is, up to Ka band and beyond into the EHF frequencies and overcome the effects of environmental attenuation

- Design of improved feed systems that can allow the creation of a very large number of spot beams that can be interconnected via on-switching and board processing
- Design of satellite antenna and electronic systems that can work interactively with terrestrial telecommunications systems and/or support multiservice satellite capabilities from a multipurpose bus
- Design antenna systems to support a wide range of telecommunications and data relay requirements for a wide range of application satellites including remote sensing, satellite navigation, satellite meteorology, as well as scientific satellites
- Design satellite telecommunications systems capable of supporting interplanetary communications

Phase One: The Earliest Phase of Satellite Communications with Omni Antennas

One of the first of the world's artificial telecommunications satellites such as the 1962 low earth orbit Telstar, pictured below, contained an omni antenna. This omni antenna was deployed on the Telstar because there was no stabilization or pointing system that could accurately point this satellite toward Earth as would be required by a higher gain antenna. Most of the irradiated power from this small satellite was thus uselessly transmitted into outer space. Any attempt to significantly increase the telecommunications throughput of a communications satellite would clearly need to find a way to concentrate the irradiated energy toward Earth rather than sending electronic signal in every direction with equal effect (Fig. 17.7).

By the time of the launch of the Early Bird satellite (i.e., the Intelsat I (F-1)), the Hughes Aircraft designers, led by Dr. Harold Rosen, concluded they could "squint" the antenna beam so that a signal could be somewhat directed toward Earth while circling in geosynchronous orbit. This was more efficient than a full omni-beam antenna but still a great deal of the energy was lost into outer space (Fig. 17.8).

Phase Two: Three Axis Stabilization and Higher Gain Satellite Antennas

The next step in the evolution of satellite communications antenna performance came by deploying a "despun antenna" that could be constantly pointed toward the Earth. This was achieved by developing a "spinning" spacecraft with sufficient rotational speed and angular momentum to maintain a constant vertical stabilization just like a spinning top. The critical aspect of the design was to have the antenna and electronics inside the spacecraft to spin in the reverse direction from the outside spacecraft body that contained the power and fuel for the firing of jets. Since the rotational speed of the spacecraft on the outside could be exactly matched to the interior antenna, the effect was to create a system that could continuously point to the Earth in all three axes.

The horn antenna system on the Intelsat 3 satellite, pictured below spun at 60 revolutions a minute and the spacecraft matched this speed in to opposite direction. The result was a reasonably high-gain antenna constantly and steadily illuminating the

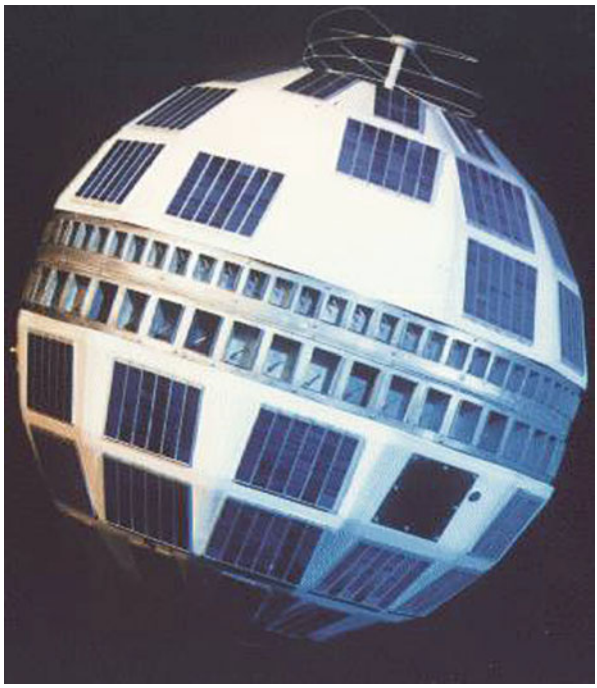


Fig. 17.7 The Telstar satellite designed by Bell labs and launched by NASA (Photo courtesy of NASA)

earth. The combination of higher power and higher gain antenna system allowed this satellite to achieve a capacity of 1,200 duplex voice circuits plus two color television channels. This capacity of this satellite when launched in the late 1960s in effect duplicated the entire world's capacity for overseas communications at that time. The bearing in the Intelsat 3 that allowed the spinning antenna to operate in this mode unfortunately froze in the first Intelsat III satellite to be launched successfully. Refined engineering of the diameter of the bearing and a new lubricating system allowed the remaining satellites in this series and hundreds of three-axis spinners that followed to operate successfully. In fact, it was this type of satellite that allowed transoceanic services across the Atlantic, Pacific, and Indian Oceans to create the first truly global satellite communications system to be established in mid year 1969 and allowed global television coverage of the first Moon landing in July 1969 (Fig. 17.9).

Phase 3: The Creation of Higher Gain Parabolic Reflector on Communications Satellites

Now, that 3-axis spinning satellite communications systems had been demonstrated to work in geosynchronous orbit, the next step was to upgrade from horn antennas to parabolic reflector antennas on commercial satellites. The Intelsat 4 and 4A satellites, that were manufactured by the Hughes Aircraft Company, featured such parabolic antennas in the 1970s (Fig. 17.10).

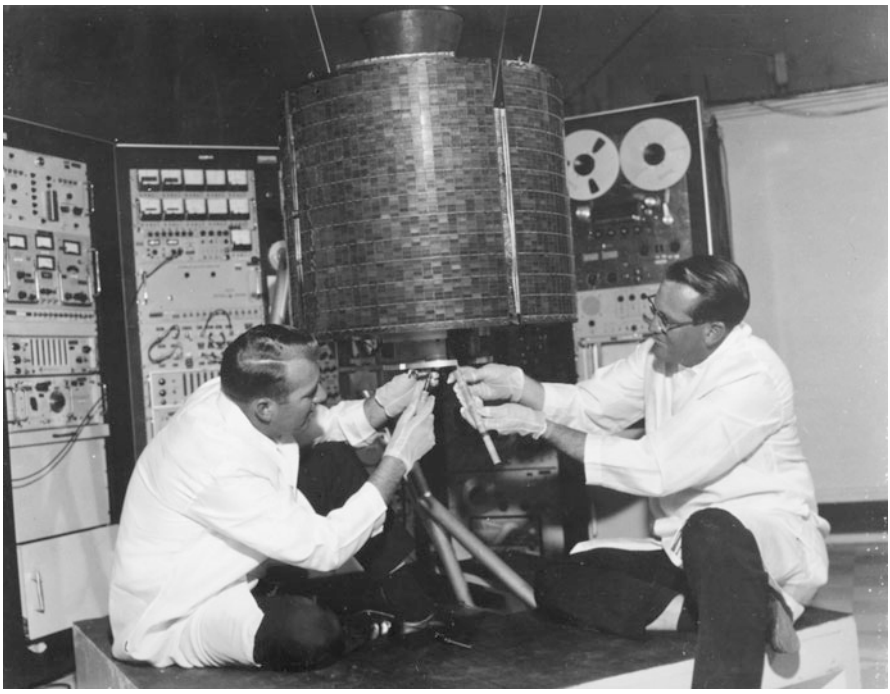


Fig. 17.8 The early bird satellite launched by Intelsat in 1965 improved performance via a squinted-beam omni antenna (Graphics courtesy of Comsat legacy foundation)

Once again the addition of more power (as a result of much larger solar arrays and drop-down circular panels) plus significantly higher gain parabolic antennas allowed another a significant boost in communications capacity to support trans-oceanic communications.

Phase 4: The Advent of Body-Stabilized Communications Satellites

The "despun" satellite worked well for pointing 1 or even 2 m parabolic satellite antennas toward Earth, but if one wanted to deploy even higher gain antennas that could create very tightly pointed beams to targeted areas, this design had limitations. Further, the spinning design with the solar cells on the outside of the drum meant the solar cells were being illuminated by the sun only about 40% of the time.

The answer that was developed at the Jet Propulsion Lab for interplanetary exploration missions was to replace the spacecraft spinning like a top and instead put momentum wheels that spun much faster inside the spacecraft body to create a platform that revolved around the Earth that remained stable in all three axes. These momentum or inertial wheels because they rotated at speeds up to 5,000 rpm could be much smaller than the overall spacecraft since mass times velocity determines momentum.



Fig. 17.9 Intelsat 3 Satellite Manufactured by TRW with "Despun" horn antenna (Graphics courtesy of Comsat legacy foundation)

This three axis stabilized spacecraft with solar arrays extended from the box-like body offered a number of advantages. It could be pointed with an accuracy of 0.5° or less and thus host very large antenna systems. The resulting spot beams could be held to their optimum location with much less variation in gain. The extended solar arrays could "see" the sun, except during eclipse periods, all the time and the arrays could be tilted to get maximum exposure. In practice, this design has also proved quite reliable and magnetically suspended momentum wheels have essentially no friction and thus no mechanically wear out or lubrication issues. This design for communications satellites was commercially deployed in the 1980s with the Intelsat V and other spacecraft (Fig. 17.11).

Phase 5: Service Diversification and Alternative Satellite Antenna Design

The last 30 years has been a time of diversification with many new types of satellites being designed for fixed, mobile, broadcast, and defense related services. These different services operated in many different frequency bands. This required different types of satellite antennas to meet system users' needs. These various requirements led to the creation of different types of antennas on the satellites – that generally grew bigger. But on the ground or on the oceans or in airspace the user antennas and receive only terminals have grown smaller and smaller.

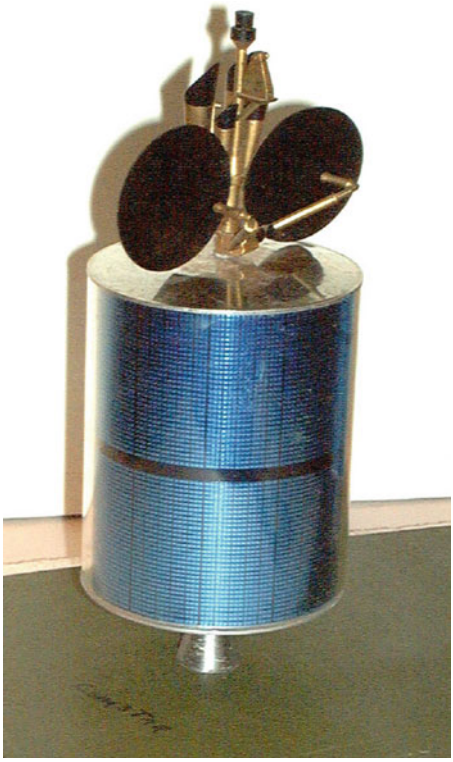


Fig. 17.10 The Intelsat 4 with high-gain despun antennas (Graphics courtesy of Comsat Legacy Foundation)

NASA's Application Test Satellite 5 (ATS-5) demonstrated some 35 years ago the feasibility of deployable antennas for broadcast and mobile service applications. This technology has evolved further and further as deployable antennas, as demonstrated by the ETS-8 experimental satellite of Japan pictured earlier, and today's operational mobile satellite systems with truly huge satellite antennas such as now demonstrated by Inmarsat 4, Thuraya, Terrestar, and Lightening Squared now demonstrate.

Also leading the evolution toward large space based satellite antennas was the Communications Test Satellite – a joint project of NASA and the Canadian Space Agency. This satellite also demonstrated new mobile satellite service capabilities and the feasibility of very small aperture terminals (VSATs) in remote locations such as the Amazon. A series of experimental vehicles in Europe (known as the Experimental Communications Satellites (ECS) series and Maritime Experimental Communications Satellites (MAREC) plus the Japanese Experimental Test Satellites helped to advance communications satellite antenna design. Innovative operational programs such as the Marisat satellite design to provide communications services to the U.S. Navy and commercial maritime shipping and the maritime packages on some of the later Intelsat V satellites also demonstrated not only a host of new satellite applications but a wide range of new satellite antenna designs including the idea of having more than one communications payload and sets of antennas on one satellite.



Fig. 17.11 The Intelsat V designed by space systems/ Loral featured 3 axis body stabilization (Graphics courtesy of Comsat Legacy Foundation)

In general, three types of satellites and antenna systems evolved. (1) There are broadcast satellites for BSS service have smaller antennas for broader coverage. These spacecraft are equipped with very large power systems so that they could operate with very small receiver dishes. (2) Mobile satellites for MSS service have evolved toward spacecraft with very large antennas because they operated at the lower frequencies and new to work to very small mobile antenna units including hand-held satellite telephone units. This has led to these types of spacecraft operating with a large aperture multi-beam antenna that can create a large number of beams which can be interconnected via onboard switching. (3) Finally Fixed Satellite Service (FSS) spacecraft also have moved toward higher power and larger antennas to support not only Very Small Aperture Antennas (VSAAs) but even Ultra-Small Aperture Antennas (USAAs) especially as these services migrated to Ku band and even to Ka band. These space craft also retain smaller aperture antennas for global and zonal coverage.

The most demanding designs for broadcasting and mobile satellite systems often benefit the design of FSS satellites that do not have the most stringent requirements. Today, there are satellites that tend to be evolving toward "multi-purpose" busses that can meet a variety of requirements if simply equipped with the right antennas to service the right spectrum bands that can be fitted to the right class of spacecraft body and power system suited to the various service needs. Just as automobiles are designed

to be built on various sized platforms, the satellite industry has evolved in the same direction. Thus one takes a solar power array system and supporting battery system, takes an appropriate spacecraft body and stabilization system and then fits it with the various types of antennas need to meet one or perhaps several missions. Those this mean that the evolution of satellite technology is essentially complete and most types of antennas need to meet future needs already invented? The answer is clearly no. The next generation of satellite antenna technology is currently being invented.

Future Satellite Antenna Technology

One of the keys to the future may well be phased array satellite antennas or phased array feed systems that can create a very large number of beams off of a multi-beam antenna reflector. This type of technology is currently being developed on experimental satellites and phased array antennas have actually been used on the Iridium mobile satellite system.

Phased Array Satellite Antennas of WINDS Satellite

An example of a Ka-band active phased array antenna (APAA) is the one that was designed for the broadband Internet satellite WINDS (Wideband Internetworking Demonstration Satellite). This satellite operates in the 28 and 18 GHz bands. The WINDS satellite has been shown previously shown in Fig. 17.2. The WINDS satellite performs as a very broad band Ka-band satellite with high-speed transmission capability of in the gigabits/s range. This satellite can operate to a single location at a maximum speed of 1.2 Gigabit/s. The system can operate as a bent pipe or it can use onboard ATM (Asynchronous Transfer Mode) switching, and multi-beam antennas with a high-speed scanning capability (Yajima et al. 2007).

The APAA can scan independently each two electronic beams of transmit and receive and covers almost whole regional area of the satellite's outlook as shown in Fig. 17.12. The antenna control method enables consecutive wave modes and SS-TDMA (Satellite Switched Time Division Multiple Access) mode to reach various sites (Yajima et al. 2007).

The Active Phased Array Antenna pictured below in Fig. 17.13 consists of 128 antenna elements that can be flexibly used to create beams for satellite transmission. Each element consists of a high density RF module.

The main performance characteristics of the WINDS APAA are shown in Table 17.2. The block diagram of APAA is shown in Fig. 17.14 below.

The existence of such a large number of antenna elements allows the flexible creation of many beams, not through conventional physical beam forming off of a multi-beam antenna, but through the creation of beams through electronic processing.

The WINDS satellite represents state of the art capabilities for a broad band phased array satellite antenna on a geo satellite.

Phased array satellite antennas have been used on commercial and defense communications satellites with lesser active elements and at lower orbits. The Iridium satellite constellation, launched in 1997 and 1998 operates in low

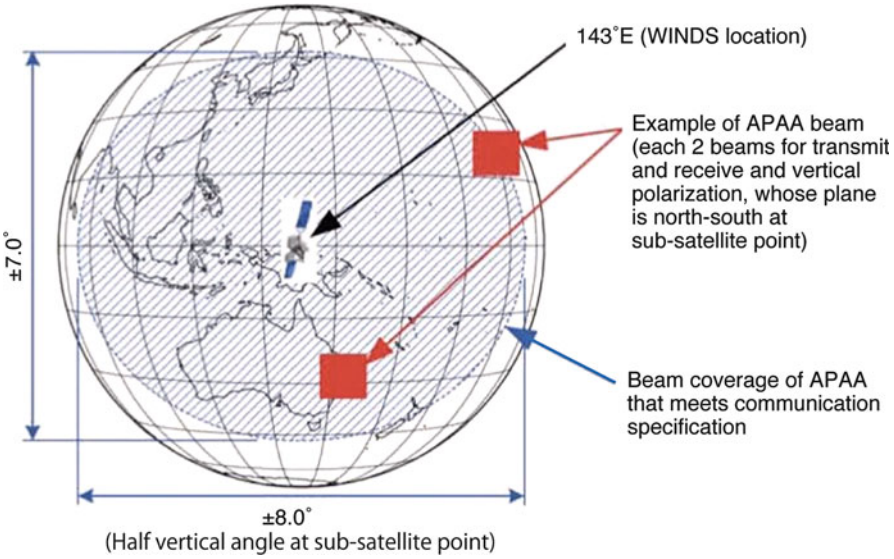


Fig. 17.12 Active phased array beam forming capability of the WINDS satellite (Graphic courtesy of the NICT of Japan)

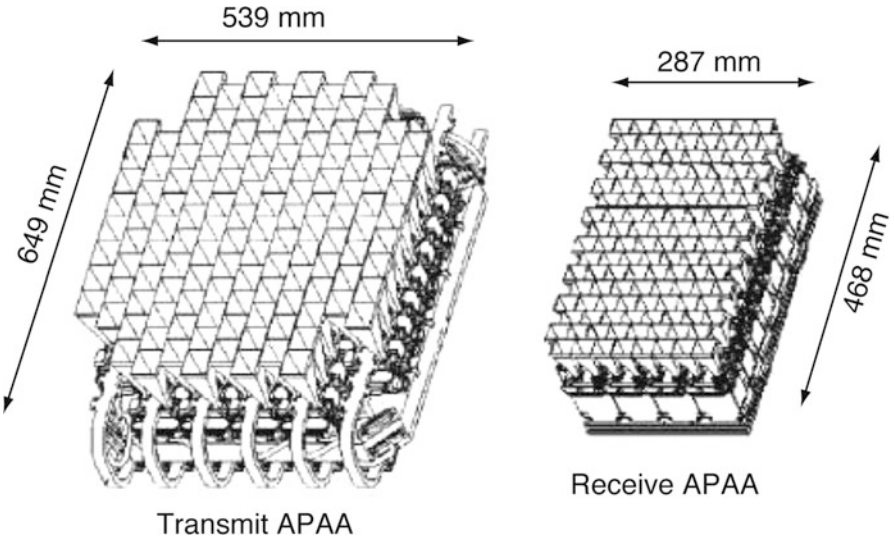


Fig. 17.13 Phase array elements in the WINDS satellite design (Graphics courtesy of JAXA)

Earth orbit and its mobile services are provided in the L band. In order to achieve the 48 cell coverage required to generate the 48 beams projected by each Iridium satellite the design called for three separate panels that were able to simultaneously transmit or receive 16 beams.

Table 17.2 Performance of WINDS active phased array antenna (APAA)

		APAA	
Item	Unit	Transmit antenna	Receive antenna
Style of antenna		Directly emitted phased array antenna	Directly emitted phased array antenna
Size of antenna	mm	1,510 × 990 × 1,530	
Weight of antenna	kg	183	
Aperture of array	mm	649 × 539	287 × 468
Frequency band	GHz	18	28
Frequency bandwidth	GHz	1.1	
Number of elements		128	128
Polarization		Linear polarization	
Scan range of beam	deg	Within ellipse of (θ, ϕ) (see Fig. 17.12)	
		Long axis: $\theta \leq 8$, Short axis: $\theta \leq 7$, $\phi = 0 \sim 360$	
Effective isotropic irradiated power (EIRP)	dBW	≥54.6 per carrier	
		≥52.1 per 2 carrier	
G/T	dB/K		≥7.1
Number of bit of phase shifters	bit	5	5
Operation mode		SS/TDMA mode	
		Continuous mode	
Timing of beam scanning	ms	2 (SS/TDMA mode)	
Consumption power	W	≤750	

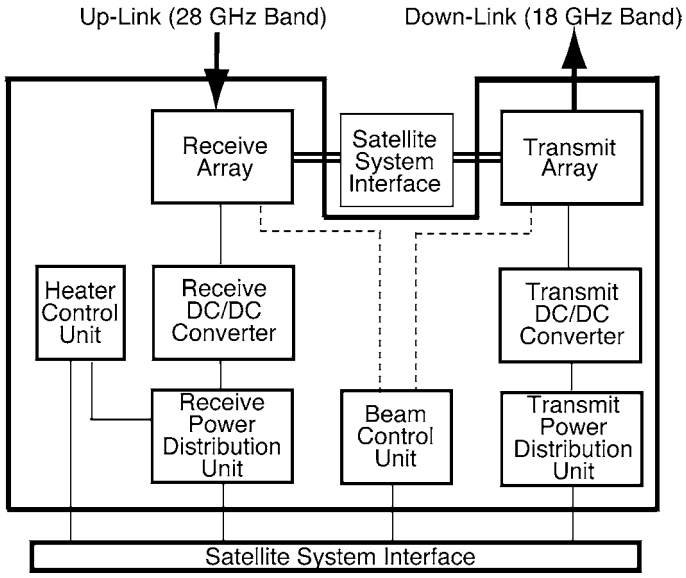


Fig. 17.14 Block diagram for the WINDS satellite transmit and receive system (Graphic courtesy of NICT of Japan)

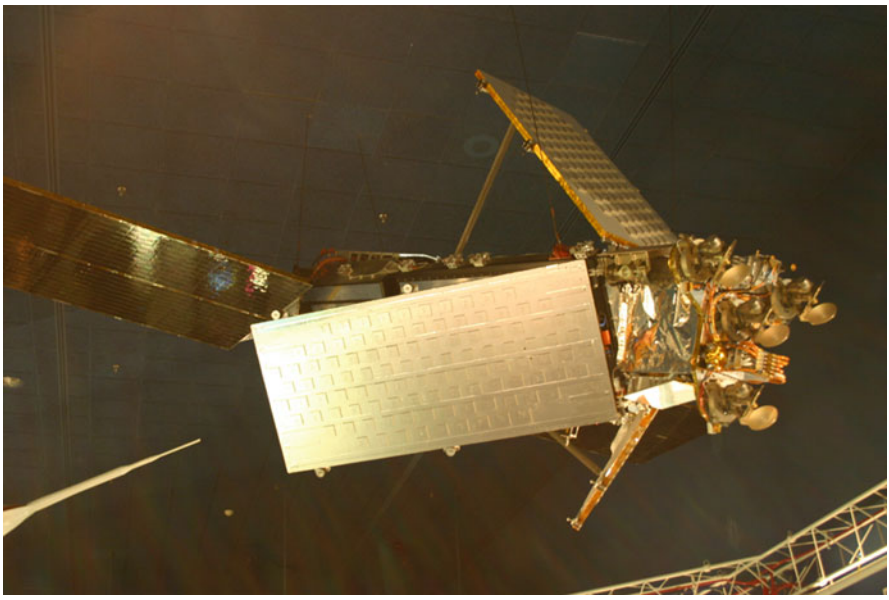


Fig. 17.15 An iridium satellite with three phase array panels and ISL antennas shown (Graphic courtesy of iridium)

These three phased array elements are arranged to be activated by a Butler-matrix-type beamformer that can simultaneously generate the 16 required shaped beams with minimal losses in gain. These systems that were manufactured by the Raytheon Corporation have proved not only able to generate clearly differentiated beams but with high reliability with many of these satellites operating more than 12 years in orbit and a survival record of over 600 satellite years of operation for the combined constellation. This is of particular note since the satellite antennas as well as the entire spacecraft were subject to very accelerated testing. These satellites have another feature that will be discussed later namely, the use of intersatellite links (ISLs) to allow the entire global constellation of 66 satellites plus spares to be managed by only two control stations. These RF based ISLs operate at 23 GHz and each satellite has four such ISLs. Two of the ISLs are used to connect to the satellites in "front" and "back" of the North-South latitudinal orbit, and two are used to connect to the satellites to the "sides" in parallel orbits in the longitudinal direction (Fig. 17.15).

Phased Array Antenna for the Quazi-Zenith Satellite System (QZSS)

Yet another example of a phased array antenna is one onboard QZSS (Quasi-Zenith Satellite System) as shown in Fig. 17.16 along with the major specifications (Jono et al. 2006). This satellite is in an unusual orbit that is like a geosynchronous orbit but inclined at a 45° angle to the orbital plane. This satellite carries experiments for space navigation services. The unusual orbit when populated by three satellites can provide



Fig. 17.16 Graphic of the quasi-zenith satellite (Graphic courtesy of JAXA)

extremely favorable look angles for coverage to the Japanese islands that allows service even in areas with extensive high rise buildings such as in downtown Tokyo.

The phased array antenna on the Quasi-Zenith Satellite (QZSS) operates in the L band. In the case of the active phased array antenna is an L-band helical antenna with 7 inner elements and 12 outer elements. The inner elements provide antenna gain with right circular polarization and the 12 outer elements form the antenna beam (Furubayashi et al. 2008).

In future years, there will likely be more phased array antennas and multi-beam reflector antennas with phased array feed systems as improved engineering and economies of scale allow such antennas to be designed and built at lower cost.

Large Deployable Antennas for Mobile Services

As noted earlier, one the most important elements for achieving truly high-gain satellite antennas in space, is the need to develop deployable satellite antennas. This is because much larger aperture antennas can be achieved with greatly reduced weight using a deployable and unfurlable design. Such a design allows very large aperture devices to be stowed compactly within the nose fairing of a launch system that might be 4–5 m in diameter. Deployable antennas in the range of 12–20 m in diameter could not otherwise be launched into orbit unless they were folded up and then deployed in space. The challenge has been not only to develop systems that can unfurl or deploy in space but to use extremely fine mesh and precision frameworks so

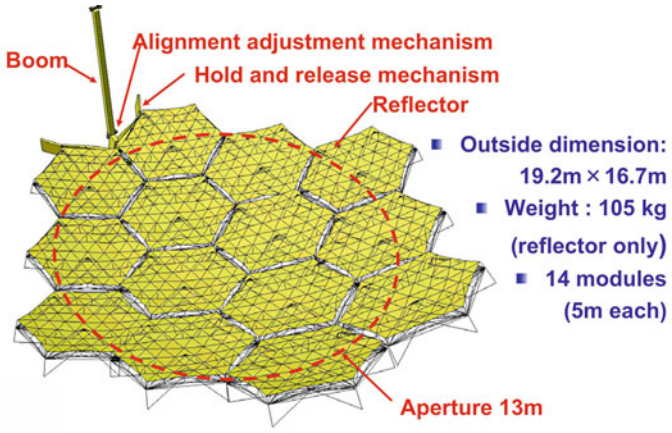


Fig. 17.17 Highly conformal large-scale deployable mesh antenna

that the deployed antenna can still conform to the parabolic shape with great precision to accommodate Ku- and even Ka-band frequencies (i.e., 12–30 GHz) as well as in lower S- and L-band systems (Fig. 17.17).

Although 18 m (or 60 ft) antennas represent the current industrial state of the art for the commercial communications satellite industry, there are apparently classified missions that have created unfurlable antennas that are up to 100 m in diameter that have been developed and deployed such as the reported National Security Agency NROL-32 mission[3].

In addition to the challenge of designing antennas that unfurl or otherwise unfold in space, there is also the difficulty of creating a complex feed system that might create hundreds of beams off the surface of such a large reflector. In designing for the future a phased array feed system may be the only way to create such a complex feed system within a condensed enough space. In the case of the Inmarsat 4 satellite that operates with a 9 m deployable antenna there is a 120 element helix cup array which was designed and built in Europe by EMS. This feed system is designed to create over 220 spot beams and some 20 zonal and broader coverage beams in the L band in the 2,500–2,600 MHz range. The design of the array feed system in this case emphasized the virtual elimination of passive intermodulation (PIM) products that would create interference among the large number of beams created.

The Antennas of Terrestar and Light Squared

The latest commercial technology in the field of extremely large satellite antennas that deploy in space are represented by the two U.S. based mobile satellite operators known as Terrestar and LightSquared (formerly SkyTerra and prior to that Mobile Satellite Ventures). These satellites have gigantic satellite antennas that are 18 m in diameter with fine gold-plated mesh parabolic reflectors. These satellite antennas were designed

and built by the Harris Corporation (Terrestar) and Boeing (Light Squared) respectively. These satellite both have extremely fine contoured deployable mesh antennas designed to provide the maximum gain in the assigned S-band frequencies. These deployable mesh reflectors are by far the largest commercial satellite antennas ever deployed. These satellites are also different in that they are part of an integrated space and terrestrial telecommunications network for mobile communications. Each of these operators will invest on the order of $7 billion in space and ground assets to provide blanket mobile cellular services for the entire continental United States, Canada, Puerto Rico, Hawaii, and Alaska. These satellite providers are offering what is known in the United States as Mobile Satellite Service with an Ancillary Terrestrial Component (MSS-ATC). The concept is to deploy and operate terrestrial cellular mobile services including video services via 4 G standards in urban areas, but have a dual mode satellite and terrestrial phone that will switch to satellite service as one makes the transition to rural or more sparsely populated suburb or exo-urban areas. These companies intend to provide phone, text, data, and video services. The gigantic structural framework for the 18 m Terrestar satellite antenna was deployed in space in July 2009. This unfurlable gold mesh antenna is shown in its "stowed configuration" prior to launch in the figure below (Fig. 17.18).

The Solaris Corporation of Dublin, Ireland has been licensed in a number of European countries to provide a similar hybrid MSS and terrestrial service in Europe that will, like the U.S. hybrid systems provide e-mail, text, voice and video services using a combination of terrestrial and satellite networks. This will, like in the case of the US-based MSS-ATC services require the launch and deployment of truly large 18 m (60 ft) satellite antennas to achieve the amount of gain to provide video to small dual mode 4 G satellite/cellular phones.

Optical Communications Systems

The rapid gains in terrestrial communications in the past two to three decades has come with the transmission of messages utilizing the light spectrum and sent through fiber optic cable. The very high quality light fiber carries modulated and encoded messages with very little attenuation and the enclosed cable prevents interference. Since the light frequencies are much higher than radio waves and thus their wavelengths very much smaller this allows extremely high throughput capability. The development of high quality blue and green lasers that are at even higher frequencies than red wavelength lasers opens up even more capacity for the future. One pathway to the future, in terms of higher capacity satellites, would thus seem to be the use of modulated light waves. The problem, in terms of transmission of modulated light signals via satellite, is that atmospheric conditions can easily interrupt these signals. In clear sky conditions, signals can be transmitted from Earth to an optical telescope and *vice versa*. In the case of cloud cover, rain, snow, or other conditions where the light signal can be blocked the transmission will not go through. Optical transmission above the atmosphere can be of very high quality. Since the speed of light is fastest in a vacuum and encounters no interference, the application of light signals for either



Fig. 17.18 Framework for the 18 m terrestar mobile satellite antenna (Graphics courtesy of Harris corporation)

intersatellite links or for interplanetary or cislunar links is a very logical application and both applications are under development in experimental R&D programs. The following discusses experiments that are being carried out using optical waves for satellite-based communications.

Optical Antenna of OICETS Satellite

One example of an onboard optical antenna is that of the Japanese Optical Intersatellite Communication Engineering Test Satellite (OICETS) (Jono et al. 2006). The experimental OICETS was launched in 2004. This small satellite had a weight of 570 kg and it was launched into a circular and nearly polar circular orbit with an inclination of 98°. It conducted successfully the optical intersatellite communication experiment with ESA's ARTEMIS satellite. The data transmission rate for the OICITS satellite was 50 Mbit/s (Fig. 17.19).

The mission equipment for OICITS was called as LUCE (for Laser Utilizing Communication Equipment) as shown in (Fig. 17.20).

The LUCE consists of an optical telescope on Earth and corresponding optical and electronic parts in the satellite. The optical part on Earth was a two axis gimballed, high-gain optical antenna with an inner part. This inner optical part consisted of

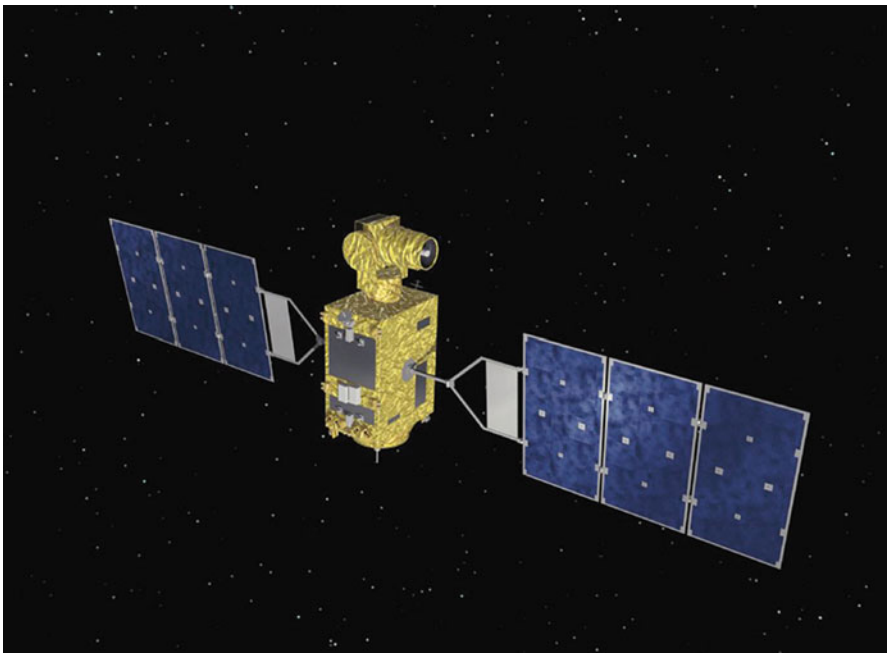


Fig. 17.19 The experimental optical communications satellite (OICETS) (Graphics courtesy of JAXA)

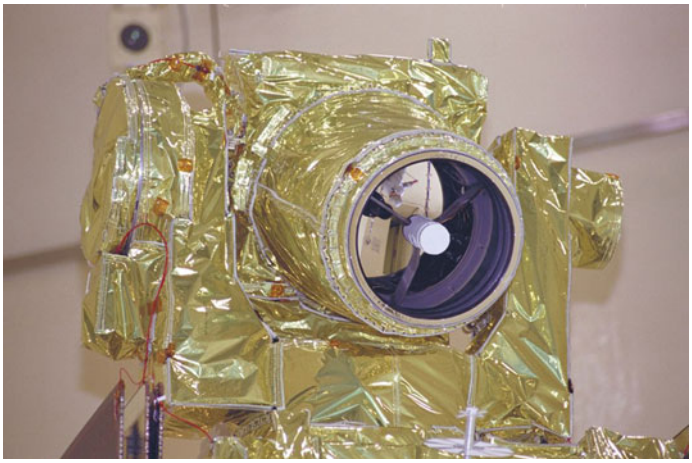


Fig. 17.20 The laser utilizing communications equipment (LUCE) of the OICITS satellite (Graphics courtesy of JAXA)

Table 17.3 Performance of OICETS optical antenna

Structure	Cassegrain-type telescope
Wave front error	Less than λ/20 rms within 1 mrad field of view (λ = 847 nm)
Magnification	20
Effective diameter	26 cm

optical elements including a high power output semiconductor laser as well as a high sensitivity signal detector. The major specification is shown in Table 17.3.

European Optical Antennas

The Artemis satellite data relay satellite, developed by the European Space Agency, has several payloads. This experimental satellite was eventually deployed in geosynchronous orbit in late 2001 using its ion engine to reach the desired altitude. This was accomplished over a period of several weeks after its initial deployment resulted in a lesser altitude elliptical orbit. The Artemis was designed and manufactured by Astrium to support data relay between low earth orbit satellites and geosynchronous orbit. The key payload on the Artemis satellite was the SILEX Semiconductor-laser Intersatellite Link Experiment (SILEX) that can send and receive signals between Artemis and lower earth orbit satellites. Specifically SILEX has been used for experiments to link to an airplane in flight and to connect to the Japanese OICETS satellite. Most significantly it was used once a day to connect with the French SPOT-4 remote-sensing satellite to obtain real time images from the ground or to empty the memory of this satellite for relay to the ground via conventional Ka-band radio frequency transmissions.[4]

The SILEX payload is based on Gallium Aluminum Arsenide (GaAlAs) laser diode as the transmitter and a photodiode detector, with a 25 cm aperture, fork-mounted telescope that can be used for establishing links with satellites in lower orbits. This system supports a data rate of 50 Mbps. The SILEX payload weighs about 160 kg and uses 150 W of power. Before the Artemis satellite ended operations over 1,000 successful laser transmissions were achieved between Artemis and other spacecraft and/or aircraft.[5]

Improved Large-Scale Satellite Systems with Large Reflectors with Phased Array Feeds and Nano Satellite Arrays

There is some thought that the continued development of high capacity and low cost fiber optic systems could in time make satellite technology obsolete. This, in fact, does not seem to be the case. This is because of two important reasons. One essential reason is that fiber optic technology does not support mobile

communications nor does it efficiently support broadcast and multicasting operations or services to remote and sparsely populated areas. Secondly, satellite technology has a large number of new concepts still be developed. There are ideas about how very large-scale passive reflectors combined with phased array feed systems to increase the capabilities of satellites even beyond the nearly 20 m antenna systems deployed with the Terrestar and LightSquared MSS-ATC systems. There have further ideas that thousands of nanosatellite elements might be deployed to form a very large "virtual satellite" in the sky that might be square kilometers in size. Some of these concepts are described in ▶ Chap. 20, "Satellite Communications Regulatory, Legal and Trade Issues" on the future of satellite systems.

Notes

1. http://www.**lascom**.or.jp/member/iomver/iomver4_303.pdf
2. The ETS-8 Project http://www.jaxa.jp/projects/sat/ets8/index_e.html
3. http://www.liquida.com/page/13583910/
4. Artemis Satellite Reaches Geostationary Orbit, ESA News www.**esa**.int/export/**esa**CP/SEMCVY1A6BD_index_0.html
5. Silex: The First European Optical Communication Terminal in Orbit, www.esa.int/esapub/bulletin/bullet96/NIELSEN.pdf also see Silex update in Space Reference, "SILEX: More than one thousand successful optical links," http://www.spaceref/com/news/viewpr.html?pid=17298

References

Y. Demers, A. Liang, E. Amyotte, E. Keay, G. Marks, Very large reflectors for multibeam antenna missions, in 15th ka and broadband communications, navigation and earth observation conference, Cagliari, Italy, 23–25 Sep 2009

T. Furubayashi, O. Amano, T. Takahashi, H. Noda, E. Myojin, M. Kishimoto, Test result of onboard L-band antenna assembly for high accuracy positioning experiment system, in 52nd association conference on space science and technology, 3F04, 2008. Also see http://www.jaxa.jp/projects/sat/qzss/index_e.html

T. Iida (ed.) *Satellite Communications –System and Its Design Technology*, (Ohmsha Ltd., Tokyo/IOS Press, Amsterdam, 2000)

T. Jono, Y. Takayama, K. Ohinata, N. Kura, Y. Koyama, K. Arai, Demonstrations of ARTEMIS-OICETS inter-satellite laser communications, in 24th AIAA international communications satellite systems conference (ICSSC), No. AIAA 2006–5461, 11–14 Jun 2006

M. Yajima, T. Kuroda, T. Maeda, M. Shimada, T. Hasegawa, S. Kitao, K. Hariu, Ka-band active phased array antenna. Special Issue on WINDS, NICT Review, 53(4), 49–55 (2007) (in Japanese)

Satellite Earth Station Antenna Systems and System Design

18

Jeremy E. Allnutt

Contents

Introduction 428
Basic Antenna Concepts 428
Dipole and Monopole Antennas 428
The Concept of Antenna Gain 431
Parabolic Reflector Antennas 434
Antenna Sidelobes 437
Antenna System Aspects 437
Blockage 437
System Noise Temperature 439
Rain Attenuation 444
Depolarization 446
Modulation 451
G/*T* 452
Tracking 452
Shielding 456
Weather Protection 456
Feed Systems 458
References 460

Abstract

This chapter reviews the design and operation of user antennas for satellite communications – for fixed, mobile, and broadcast services. This review includes simple dipole antennas, progresses to Yagi-Uda antennas, and then on to high-gain parabolic reflector antennas that are the most commonly used in satellite communications systems. The trade-off between antenna gain and beamwidth is explored in detail. The key differences in the design process for

J.E. Allnutt
Department of Electrical & Computer Engineering, George Mason University, 4400 University Drive, Fairfax, VA, 22030, USA
e-mail: jallnutt@gmu.edu

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_20, © Springer Science+Business Media New York 2013

very small aperture terminals (VSATs) and large earth stations are explained. The influence of blockage on whether to choose offset-fed over on-axis fed antennas is seen to be key. This is particularly true for small aperture antennas with diameters of less than 100 wavelengths. Frequency reuse through dual-polarization operation is presented, with the different system advantages of dual-linear and dual-circular operation set out. The impact of the choice of modulation on the power margins required for a given bit error rate (BER) is seen to be significant. Noise temperature contributions from the atmosphere, the ground, and particularly from lossy feed runs that reduce antenna performance are explored. Reducing the feed losses is key to the design of very large earth station antennas. The difference in the impact of noise temperature on the uplink and the downlink is explained, and the differences between antenna design and performance with regard to fixed satellite service, mobile satellite service, and broadcast satellite services are noted. Finally, some additional aspects of earth station designs that are affected by the environment, both meteorological and interference, are discussed.

Keywords

Antenna Systems • Bit error rate (BER) • C-Band • Depolarization • Directional antennas • Ka-band • Ku-band • Micro-terminals • Modulation • Optical systems • Rain attenuation • Satellite earth stations • Site shielding • UHF • USAT • V-band • VSAT

Introduction

All telecommunications systems need to have a transmitter at one end of the path and a receiver at the other in order to complete the link. This is true whether audio frequencies are used (as in human speech, with a mouth transmitting at one end and an ear receiving at the other) or radio frequencies are used (e.g., a cell phone at one end and a base station at the other). The material between the transmitter and the receiver is referred to as the *transmission medium* and in some cases as the *propagation channel*. In order to successfully send the required message over the transmission medium, an efficient mechanism needs to be used to *launch* the message from the transmitter into the medium. Such a mechanism is an antenna.

Basic Antenna Concepts

Dipole and Monopole Antennas

An antenna is simply a device for taking energy, usually electrical energy in the form of either a current or a field, and transferring that energy as efficiently as possible from the physical structure in which it originated (e.g., a wire or a waveguide) out into the transmission medium, which is often air or space.

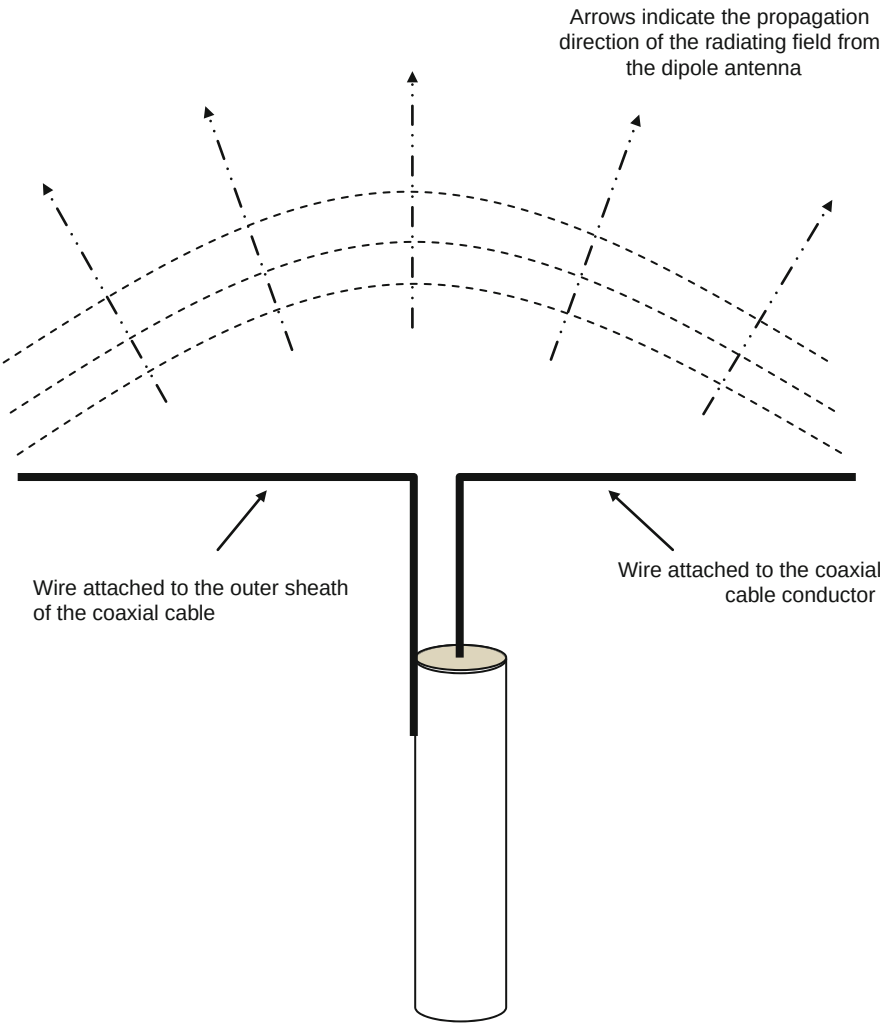


Fig. 18.1 Schematic of a simple dipole antenna. The example above shows a simple dipole antenna that is attached to a coaxial cable. The inner conductor of the coaxial cable is attached to one of the wire antennas and the coaxial cable's sheath to the other. Together, the two wire extensions – the dipole antenna – create a radiation field around the antenna that propagates away from the antenna, with a maximum approximately normal to the orientation of the dipoles. Most antennas that operate at frequencies below about 2 GHz tend to use just one wire extension, rather than two, and so are more commonly referred to as monopole antennas (see Fig. 18.2)

Antenna engineers call it "launching" the field. The simplest antenna for launching a field is a dipole antenna (see Fig. 18.1).

Dipole literally means two poles. Usually the "poles" consist of electrical conductors, one of which is attached to the source that is to be radiated and the other to a ground, or earth. The outer sheath of a coaxial cable is usually connected to the

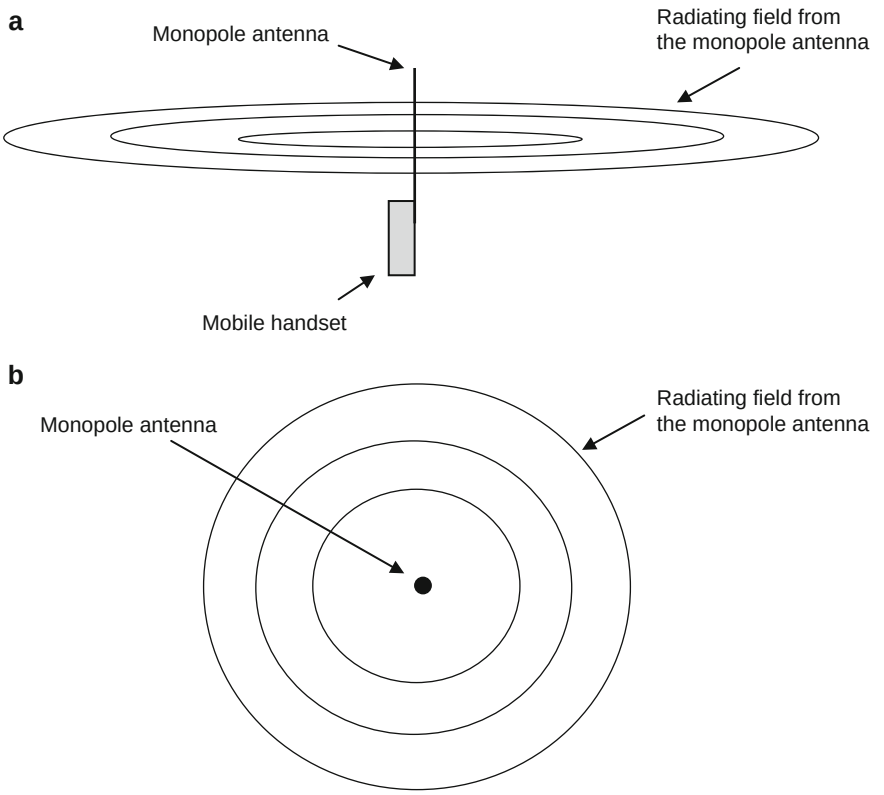


Fig. 18.2 (**a**) Schematic isometric depiction of the field radiating outward from a monopole antenna. The monopole antenna of a mobile handset is radiating equal power in every direction about its long axis. In modern mobile radio handsets, the monopole antenna is formed within the case of the radio unit or is part of the structure carrying the operator (e.g., an automobile). (**b**) Plan view schematic of the field radiating from a monopole antenna. In this plan view of the radiating field from a monopole antenna, it is clear that the energy radiating out from the monopole antenna has equal power in each direction that is normal (i.e., at right angles) to the long axis of the monopole antenna

ground: Thus a simple dipole antenna (see Fig. 18.1) can consist of one wire attached to the center conductor of the coaxial cable and the other to the sheath. The wires are bent to form a straight line normal (i.e., at 90° angle) to the coaxial cable, and the greatest energy is radiated normal to the long axis of the dipoles, as shown in Fig. 18.1. To make dipole antennas radiate as efficiently as possible, the length of the wires should be a submultiple of a wavelength of the signal to be radiated, usually half-a-wavelength or a quarter-wavelength. In many cases, the wire attached to the earth of the coaxial cable is omitted, and only the central core of the coaxial cable is used to radiate. Since just one radiating element is used, this is usually referred to as a monopole antenna (see Fig. 18.2), and the antenna is left straight in what is called a *whip* antenna. (First used at very high frequencies (VHF) – 30–300 MHz – the monopole antennas were long and flexible, looking like a whip).

In Fig. 18.2a, the monopole antenna can be seen radiating a field that is like ripples on a pond when a stone is dropped in: electromagnetic energy moves outward from the antenna with ever-increasing diameter. If the field can be measured at any point on any one of these circles radiating outward from the monopole antenna, we would find that the energy, or more properly the *power flux density*, is the same in any direction from the monopole point of origin. This is depicted schematically in Fig. 18.2b. An antenna that creates such a uniform field in any direction is called an *isotropic antenna*. Isotropic antennas are very popular for mobile handsets and electronic personal assistants, whether for terrestrial or satellite services, since they do not require the user to know in which direction the antenna needs to be pointed to pick up a signal. This versatility, however, comes at a very high price: much of the radiated energy is wasted since most of it does not reach the intended antenna at the other end of the link. To improve on this, the radiating element – the antenna system – must increase the radio energy radiated in the desired direction, or received from the desired direction when compared with all the other possible directions. This enhanced performance antenna will now have a preferred direction in which to transmit and receive radio energy. The increase in energy radiated on the preferred direction is referred to as antenna gain.

The Concept of Antenna Gain

To understand the concept of antenna gain, it is important to know that a directive antenna does not create energy: it just focuses it in the desired direction. This is illustrated in Fig. 18.3.

The simplest antenna that provides energy in a preferred direction is the Yagi-Uda antenna, named after its two Japanese inventors. The Yagi-Uda antenna is essentially a half-wavelength dipole that has a slightly larger dipole behind it to act as a reflector and smaller elements in front to provide additional gain in the forward direction. A three-element Yagi-Uda antenna is shown in Fig. 18.4.

A Yagi-Uda antenna is most often employed at frequencies in the UHF (Ultrahigh Frequency) band, which is from 300 MHz to 3 GHz. At these frequencies, the physical size of this type of antenna is convenient: In the lower bands the antenna dipoles would become large and somewhat clumsy, while in higher bands, the antenna dipoles would become small and have high losses. Table 18.1 illustrates the different wavelengths for the popular communications bands.

The *Gain* of a Yagi-Uda antenna, and that of any other antenna, is the ratio of the energy transmitted in the particular direction specified to that which an isotropic antenna would provide. By definition, the gain of an isotropic antenna is unity, since it has no preferred radiation direction (see Fig. 18.3), and the uniform radiation pattern can be thought of as creating a sphere of radiated energy around the antenna. Antenna gain can be expressed as an analog ratio or as a decibel value. The analog ratio is simply the arithmetic value the gain exceeds that of an isotropic antenna, the gain of which we have seen is unity. If the antenna now has some value of gain that

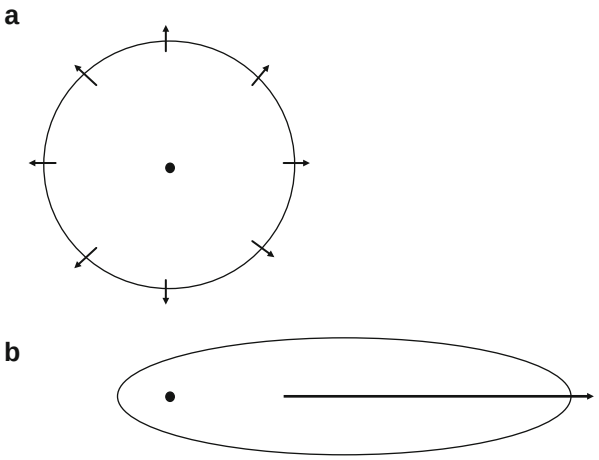


Fig. 18.3 Concept of preferred radiation direction. In (**a**) and (**b**), an antenna (shown as a *black dot*) is radiating energy. (**a**) The antenna is isotropic and it radiates energy equally in every direction. The *arrows* show the radiation direction and their length indicates the amount of power radiated in that direction. It can be seen that energy is radiated equally in each direction. (**b**) There is a preferred radiation direction, shown by the *long arrow*. The radiated energy can be thought of as being like air in a balloon. The balloon can be squeezed to give a nonuniform shape, and this is essentially what has happened in (**b**). Different antenna types "squeeze" the energy in different ways and with differing degrees of focusing, or *gain*

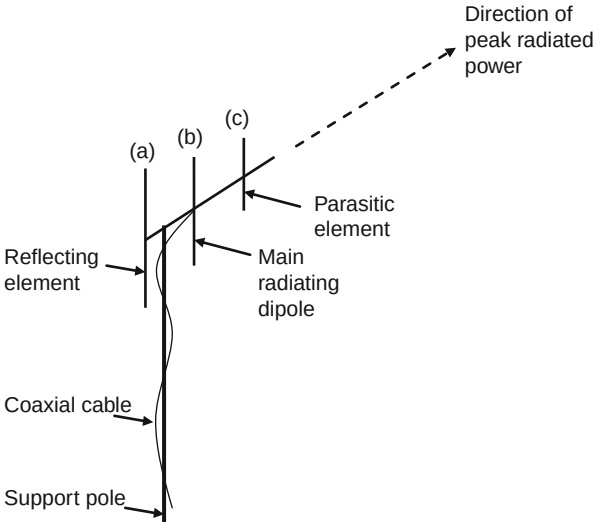


Fig. 18.4 Simple Yagi-Uda Antenna. The simplest Yagi-Uda antenna, sometimes just called a Yagi Antenna or abbreviated to Yagi, consists of a half-wave, radiating dipole with a reflecting dipole behind it and a parasitic element in front of it. The lengths of the dipoles, and the separation between them, are carefully chosen to maximize the peak radiated power, which is highest in the direction shown. Increasing the number of parasitic elements in front of the main radiating dipole will increase the forward gain and slightly narrow the beamwidth

Table 18.1 Illustration of frequency vs. wavelength for typical communications bands

Frequency	Wavelength	Half-wavelength	Quarter-wavelength
890 MHz	33.71 cm	16.85 cm	8.43 cm
1.5 GHz	20 cm	10 cm	5 cm
6 GHz	5 cm	2.5 cm	1.25 cm
14 GHz	2.14 cm	1.07 cm	0.54 cm
30 GHz	1 cm	0.5 cm	0.25 cm
200 THz	0.00015 cm	0.000075 cm	0.0000375 cm

A frequency of 890 MHz is close to that used by terrestrial mobile wireless systems and over-the-air TV broadcasting services. Satellite Mobile Services operate around 1.5 GHz (L-Band), while the uplinks for Fixed Satellite Services are at 6 GHz, 14 GHz, and 30 GHz (C-band, Ku-Band, and Ka-Band). Free-space optical communications is usually in the infrared region, close to 200 THz.

exceeds unity in one direction, the radiated energy can no longer be described as a sphere. It is more like a cone of energy symmetrical about the preferred radiation direction, in some respects like the beam of a flashlight. Usually when one speaks of the gain of an antenna, it is normally the maximum gain that is being referenced. This is the gain along the electrical axis, or *boresight*, of the antenna. Since the gain of an isotropic antenna is unity (i.e., there is no preferred radiating direction), the analog gain of any antenna is simply the ratio of the energy radiated in a given direction to that of an isotropic antenna. For example, if the amount of energy radiated in the preferred direction is ten times that which an isotropic radiator would transmit in the same direction, then the analog gain is $10/1 = 10$. For a Yagi-Uda antenna, the maximum analog forward gain is between about 15 and 100, with the higher gain being for an antenna with many more than just three dipoles. The more dipoles there are, the longer the antenna is and the more bulky it becomes. The beamwidth of a Yagi-Uda antenna – the angular dimension between points in the beam either side of boresight where the radiated energy is half the maximum – is fairly broad. Typical values are 18–25°.

Analog values of gain between 15 and 100 that we saw above for the Yagi-Uda antenna are common numbers met in everyday life, and so are easy to work with. However, when these analog gain numbers start to get large – values exceeding 10,000 being common for large earth station antennas – satellite engineers resort to the decibel notation to describe antenna gain. Converting an analog value into a decibel value is straightforward. There are two steps. First a logarithm to the base 10 is taken of the analog value, and then this logarithmic number is multiplied by 10 to arrive at the decibel value. An example is shown below for an antenna with an analog gain of 2,000.

- Step 1 – Gain of 2,000 is first converted into a logarithm: $\log_{10}(2{,}000) = 3.3$
- Step 2 – Logarithm of 3.3 is converted to a decibel value: $3.3 \times 10 = 33$ dB

Hence an analog gain of 2,000 is equivalent to a decibel gain of 33 dB. Decibel units are normally written as "dB" and so the decibel gain of this antenna = 33 dB. Antenna systems have become ubiquitous in all aspects of modern life. Small Inmarsat "M" terminals, essentially briefcases with their lids used as an antenna, are employed for satellite news gathering (SNG) and almost everyone has seen video reports from the field using these terminals. UHF and VHF antennas uplink

geophysical information (water flow rates, temperatures, humidity, etc.) to low earth-orbiting satellites for onward distribution to meteorological organizations. Anyone who has swiped their credit card at a gasoline station to pump gasoline has (probably without even knowing) used a Ku-Band VSAT link to the credit card center for ID verification and billing. USAT antennas are everywhere on tens of millions of homes receiving TV from Direct Broadcasting Satellites. More recently, two-way Internet links at Ka-Band using micro-terminals operating in what has become known as SOHO – Small Office/Home Office – are becoming widespread. And of course there is the ultimate USAT, a handheld mobile communications handset that can operate to either Low Earth Orbit (LEO) satellite constellations like Iridium and Globalstar or to Geostationary Earth Orbit (GEO) satellites like Inmarsat. As we saw earlier, a mobile wireless handset usually has unity gain, which has to be compensated for either by having the satellite in LEO, to reduce the path loss between the satellite and handheld device, or have the satellite use a huge, deployable antenna (probably 12 m in diameter) if the satellite is in GEO. Whenever there is a need to have high gain in a satellite communications link, parabolic reflector antennas are used to focus the energy.

Parabolic Reflector Antennas

A parabola is a geometric shape that has two focal points: one is at infinity and the other is close to the parabola. The concept is illustrated in Fig. 18.5.

Two principal parameters describe the performance of a parabolic reflector: the gain of the antenna and the *beamwidth* of the antenna. The beamwidth of an antenna is normally the angular distance between two points on the beam either side of boresight where the power has dropped by half (or 3 dB in decibel units) from the maximum. The terms *half-power beamwidth* and *3 dB beamwidth* are used interchangeably. Both the boresight gain and the 3 dB beamwidth of an antenna are directly related to the aperture diameter of the antenna "D," the wavelength of the signal "λ," and the efficiency of the antenna "η" as (Allnutt 1989, 2011):

$$\text{Gain} = \left(\frac{\pi D}{\lambda}\right)^2 \times \eta \tag{18.1}$$

$$\text{Beamwidth} = 1.2 \times \left(\frac{\lambda}{D}\right) \text{radians} \tag{18.2}$$

Antenna peak forward gain and the 3 dB beamwidth are also closely related, as is shown in Eq. 18.3:

$$\text{Antenna gain} = \frac{30,000}{(3 \text{ dB beamwidth})^2} \tag{18.3}$$

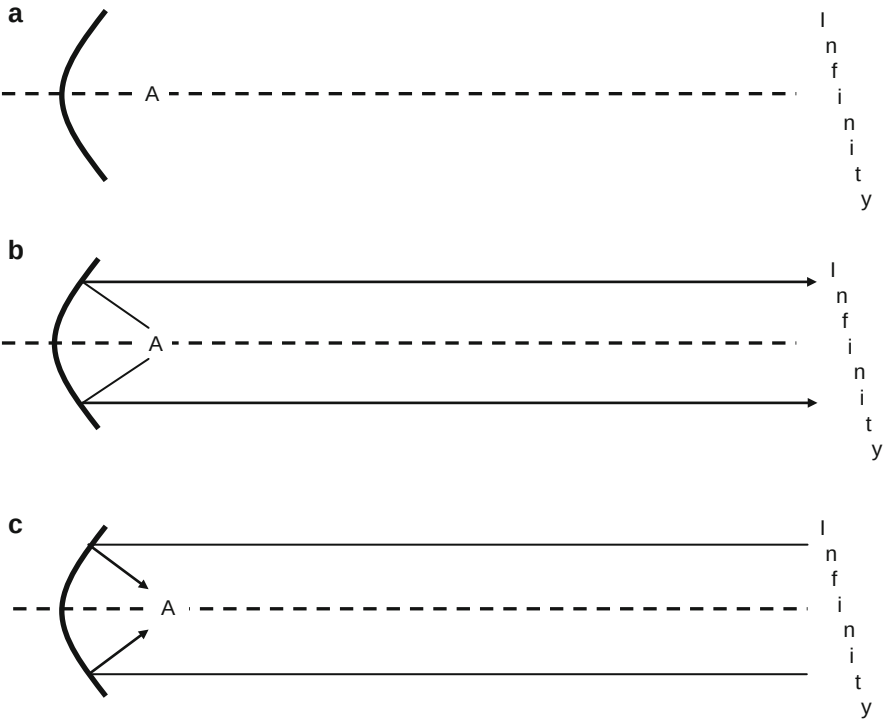


Fig. 18.5 Illustration of a parabolic antenna. A parabola has two focal points, one at infinity and the other close to the parabola (**a**). If a signal is radiated from focal point "A," then the energy will "bounce" off the surface of the parabolic reflector and be transmitted outward parallel to the antenna axis to infinity (**b**). Likewise, if energy arrives from infinity along the antenna axis, it will arrive at the focal point "A" after being reflected off the parabolic antenna (**c**). To capture the energy at point A, or to radiate it, a device known as a *feed* or *feed horn* is used. Coaxial cables or waveguides are attached to the feed to connect the feed to the earth station equipment

Note that the 3 dB beamwidth in Eq. 18.3 is expressed in degrees. Further, a final formula connecting the 3 dB beamwidth with the wavelength and aperture diameter, this time with the 3 dB beamwidth in degrees, is shown in Eq. 18.4:

$$3\text{ dB beamwidth} = \left(\frac{75\lambda}{D}\right)\text{ degrees} \tag{18.4}$$

Note that all dimensions are in meters and Eq. 18.2, 18.3, and 18.4 are dependent on the shape of the amplitude illumination over the aperture. The equations above assume an aperture distribution between $\cos\theta$ and $\cos^2\theta$ (see Fig. 18.7 and Table 18.2, discussed in the next subsection). The parameter η, the efficiency of the antenna, relates the performance of a perfect antenna to that of a real antenna. The parameter η takes the value of 1 (perfect antenna) through 0 (antenna radiates nothing), but is usually expressed in percentages. Thus a perfect antenna would have

Table 18.2 Relative peak sidelobe level to maximum forward gain for some typical aperture distributions (Tim Pratt, 2006, private communication)

Aperture distribution	Beamwidth in degrees	Peak gain relative to uniform distribution	Peak sidelobe level (in dB)
Uniform	51 λ/D	1.0	−13.2
Cosθ	69 λ/D	0.81	−23
$Cos^2\theta$	83 λ/D	0.67	−32
Triangular	73 λ/D	0.75	−26.4
$Cos^2\theta$ distribution on a pedestal	63 λ/D	0.88	−26

an efficiency of 100%. For large, standard A, antennas in the Fixed Satellite Service (FSS) with $D = 18$ m or greater, the efficiency can be as high as 70%, since forward gain is maximized. With such large aperture diameters, the *sidelobes* generated by the antennas are very close to the main beam direction – the *boresight* – and so interference into adjacent satellites is virtually nonexistent for correctly pointed antennas. For smaller antennas, particularly Very Small Aperture Terminals (VSATs) and Ultrasmall Aperture Terminals (USATs), and micro-terminals, cost is the main determinant and the efficiency is consequently lower, on the order of 50% or even less. The boundary between VSAT and USAT is not clear. Most designers generally consider any antenna aperture that is 100 wavelengths across, or less, to be a VSAT. A USAT and a micro-terminal are typically much less than this and 20 wavelengths is a common size. A mobile handset is even smaller.

Mobile Satellite Service (MSS) has a significant challenge when it has to compete with terrestrial mobile service. The average distance between a terrestrial base station and a mobile user is less than 20 km. For MSS, the distance can range from about 700 km for LEO MSS satellites to 37,000 km for GEO MSS. The only way the MSS can compete is by offering service to regions where no terrestrial infrastructure exists. Even then, significant EIRP – Equivalent Isotropic Radiated Power – has to be generated at the satellite to compensate for the low gain of the user's handset and the huge distances involved. Nevertheless, a number of GEO MSS satellites have been launched and are providing service to large numbers of users. What may make MSS satellites more successful would be the development of so-called smart antennas for user handsets. Smart antennas would employ small phased array elements that would confer two valuable attributes: an ability to steer the beam toward a satellite and a forward gain of about 3–4 dB. A gain of 3–4 dB does not sound much, but 4 dB is a ratio of 2.5. Imagine if the mileage your car obtained increased from 30 miles per gallon to 75 miles per gallon: this is a 4 dB improvement in miles per gallon. Nevertheless, the greatest gain that can be achieved will be by using a parabolic antenna, and for small aperture diameters this means that the beamwidth will be fairly broad. If the VSAT or USAT terminal is used in a receive-only mode – that is, it does not transmit – then interference into adjacent satellite systems is not a concern. However, if the earth terminal is used in both receive and transmit modes, the sidelobes can illuminate adjacent satellites in geostationary orbit. Suppression of antenna sidelobes is therefore a primary requirement for VSAT antennas that transmit to satellites.

Antenna Sidelobes

In Fig. 18.5b, it can be seen that energy radiated from the near-in focal point of a parabola by a feed element toward the reflector will be redirected along the main beam axis in a series of parallel lines to infinity. However, this is only true for an infinitely small feed aperture. A feed aperture is generally at least on the order of a wavelength across, and usually much more. Clearly an antenna feed is not infinitely small. Because of this, a *diffraction* pattern will be set up in the energy radiated from the parabolic reflector (Stutzman and Thiele 1998). A diffraction pattern causes a maximum of energy to be directed along the main axis of the reflector, the boresight direction. Away from boresight, the energy density falls off in a series of ripples, and the peaks of these ripples are called the sidelobes. This is illustrated in Fig. 18.6. In most antenna systems, there are far more than just the three sidelobes seen in Fig. 18.6. Paradoxical as it may seem, there is also what is referred to as a *back lobe* where radiated energy exists almost diametrically opposite the peak forward energy – the *main lobe*. The gain (actually it should really be called a loss since it is negative) of a parabolic antenna is generally about −10 dB in the back lobe.

All high-gain antennas that use a parabolic reflector strive to keep the received (or transmitted) signal in phase, so that the entire signal is received (or transmitted) coherently across the aperture. The same is not necessarily true of the amplitude distribution: by manipulating this distribution across the antenna aperture, the sidelobe amplitudes can be significantly reduced, thus lowering the potential for off-axis interference. This can be seen by reference to Fig. 18.7 and Table 18.2.

Figure 18.7 shows four common amplitude distributions and Table 18.2 compares the beamwidth obtained with these four amplitude distributions with that obtained using an antenna with a uniform amplitude distribution across the aperture. Amplitude distributions across the aperture are not the only consideration when building up an understanding of antenna systems and how to optimize their performance.

Antenna System Aspects

When the parameters of the satellite transponder are known and the link budget has been developed for a range of antennas operating to that satellite, there are some additional design aspects that need to be considered for the overall antenna system. These are considered below.

Blockage

There are two basic ways a parabolic antenna can be configured: on-axis and off-axis. In an on-axis configuration, the feed is located on the main axis of the antenna. In this way, the mechanical axis of the main reflector is the same as the electrical axis. The feed for an on-axis antenna will prevent some of the incoming energy from reaching the main reflector. This is referred to a *blockage*. Blockage will occur

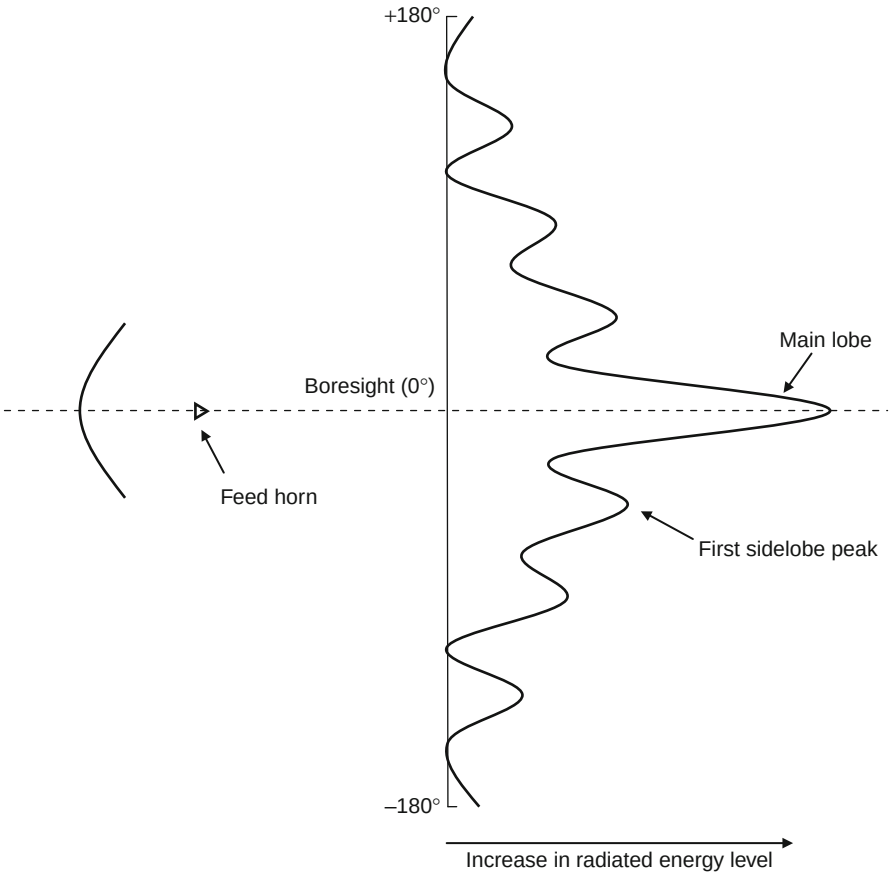


Fig. 18.6 Schematic of sidelobes from a parabolic reflector antenna in the far field. A diffraction pattern is set up when a source is not infinitely small. There are really three, somewhat ill-defined regions where different diffraction patterns exist. Very close to the reflector antenna there is a *near field* region where a constant beam from the antenna is not yet set up and energy levels can fluctuate significantly over small distances (laterally and in the direction of propagation). The next region from the antenna is called the *Frésnel region*. The energy levels still fluctuate significantly, but in a less irregular fashion than in the near field. And finally, there is the *far field* region, also called the Fraunhofer region, where the diffraction pattern from the antenna is well defined. The far field radiation levels are shown. It can be seen that there are many sidelobes either side of the main lobe peak. These sidelobes can cause interference into other satellite systems if not suppressed

on both receive and transmit. To avoid blockage, the feed can be offset from the electrical axis. On-axis and offset-fed (or off-axis) parabolic, front-fed reflectors are illustrated in Fig. 18.8.

The designs shown in Fig. 18.8 can also be characterized as single-reflector antennas, that is, there is a single feed and a single parabolic reflector. The focal length, F, of such antennas is simply the distance between the feed horn and the parabolic reflector. The longer the focal length of a reflector antenna is, the

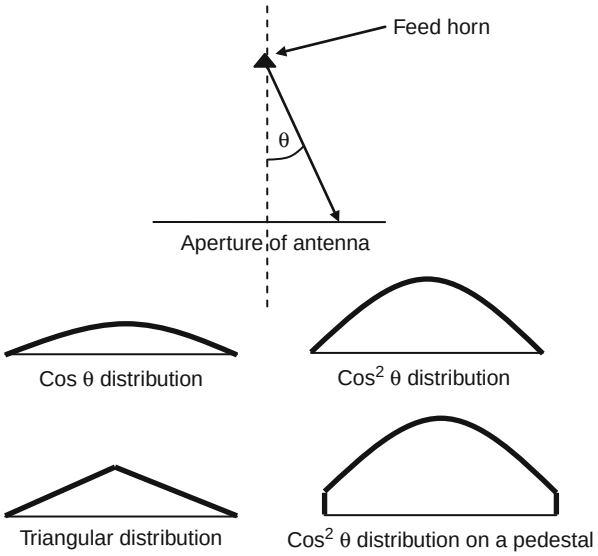


Fig. 18.7 Illustration of different aperture amplitude distributions. The thin horizontal lines are a cross section of the antenna aperture. The thick lines represent the amplitude distribution across each of the antenna apertures. An increase in energy is shown as an increase in the vertical direction (toward the feed element, not shown). As can be seen, in each of the examples, the maximum power is developed at the center of the antenna aperture. The more energy there is at the center of the antenna aperture compared with the edge of the aperture, the broader the beamwidth, the lower the sidelobe amplitudes, and the lower the peak forward gain. The angle θ refers to the angle between the feed horn and the antenna rim

better the off-axis performance. To create a longer focal length, dual-reflector configurations can be employed. The two principal dual-reflector configurations are Cassegrain and Gregorian, and these designs can be both on-axis (as shown in Fig. 18.9) and off-axis (as shown in Fig. 18.10).

Dual-reflector configurations provide easier siting of the feed and receiver, which can both be behind the antenna, thus easing maintenance requirements in terms of accessibility.

System Noise Temperature

A key parameter in the design of a receiver in a communications link, and possibly *the* key parameter, is the carrier-to-noise ratio at the input to the demodulator (Pratt and Bostian, 1986; Pratt et al. 2002). This is often written as C/N, where C is the carrier power and N is the system noise power, both expressed in watts. System noise power $N_{\text{syst}} = kT_{\text{syst}}B$, where k = Boltzmann's constant = 1.38×10^{-23}J/K $\Rightarrow$ -228.6 dBW/K/Hz, T_{syst} is the system noise temperature in degrees Kelvin, and B is the bandwidth in Hz. The calculation of system noise temperature for the receiving system is not the topic of this chapter, but system noise temperature is

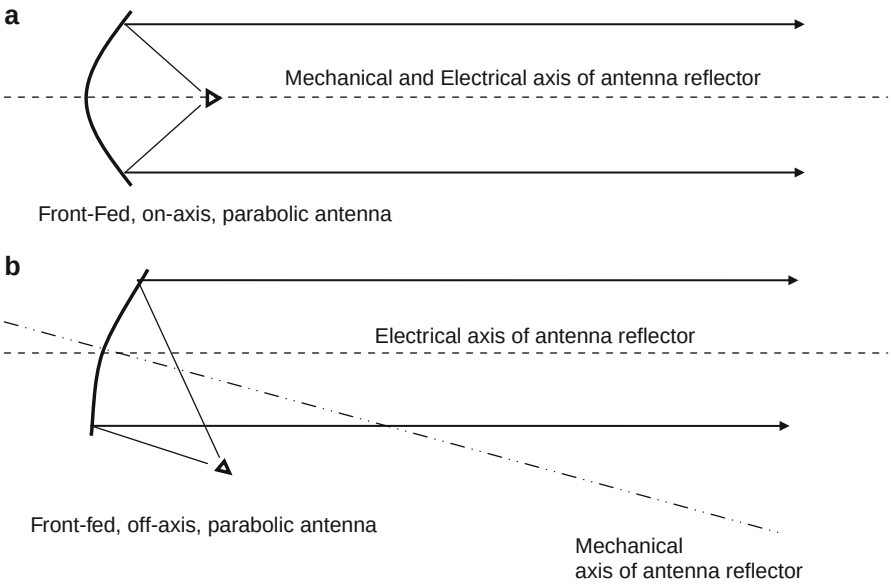


Fig. 18.8 Illustration of the two main types of parabolic antennas: front-fed and offset-fed. (**a**) The antenna is symmetrical, with the mechanical axis pointing in the same direction as the electrical axis, the boresight. The feed is placed in the center, in front of the antenna reflector, and so will block both incoming energy and transmitted energy that is on the electrical axis. (**b**) The parabolic shape is not symmetrical, being a section of the reflector to one side of the center of the parabola. The feed is displaced to one side and generates a main beam that is away from the mechanical axis. The feed in (**b**) is offset away from the electrical axis and so causes no blocking of the main beam direction. The reflector surface is still parabolic, but it is essentially a portion of the parabola away from the center of revolution of the parabola shown in (**a**)

made up of two major components: the *internal* noise temperature of the receiving system and the *external* noise temperature. The internal system noise temperature of the receiver is generally constant, with the major components being the noise temperature of the front-end amplifier and the noise temperature induced by the feed run into the low noise amplifier. The external noise temperature is the additive sum of a number of noise temperature components entering the antenna feed, and the sum is called the *antenna temperature*.

Noise temperature provides an incoherent source of unwanted energy into the receiving system. Since noise energy is not coherent, all components will sum arithmetically rather than as a vector addition. The noise temperature of an antenna will therefore consist of the sum of all noise components external to the antenna that enter via the feed from any direction. These components include the noise remnants from the *Big Bang* (around 2.8 K when the antenna is not pointed toward a radio star that emits significantly more noise than this, or when the antenna is pointed toward the center of the home galaxy, which we know as the *Milky Way*), gaseous atmospheric constituents (mainly emission from oxygen and water vapor), and from the

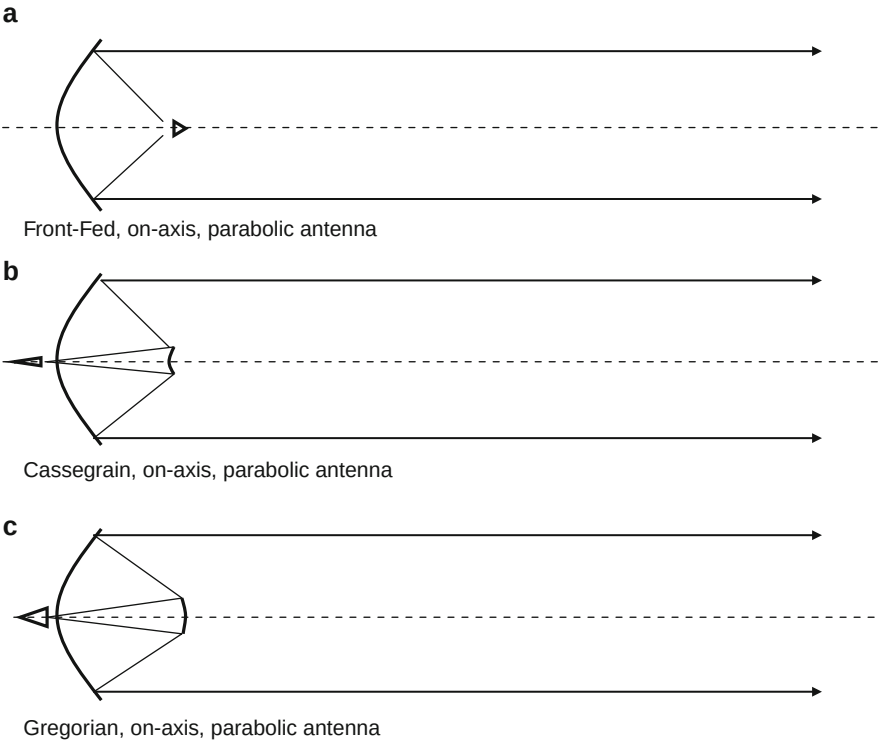


Fig. 18.9 Illustration of the three types of parabolic antennas that have their main feed system on-axis: front-fed, Cassegrain, and Gregorian. All three of the parabolic antenna types shown are symmetrical, on-axis, antennas, with both the feed and the antenna reflectors being rotationally similar about the mechanical axis of the main reflector. In (**a**), the feed is in front of the antenna reflector, while in (**b**) and (**c**) the feed is behind the antenna reflector. Not shown is a hole in the main reflector that permits the energy radiated by the feed in (**b**) and (**c**) to pass through to the sub-reflector antenna, and from thence onto the main reflector. Note that the sub-reflector for the Cassegrain antenna is hyperbolic in shape, while that of the Gregorian is elliptical in shape. Also note that the sub-reflector of the Cassegrain antenna has a curvature that is oriented in a similar direction to the parabolic main reflector, while the elliptical sub-reflector of the Gregorian antenna is oriented in an opposite sense to the parabolic main reflector

ground (the latter mainly through sidelobes that intercept the ground) (Recommendation ITU-R P.676-4, 1999). This is illustrated schematically in Fig. 18.11.

The attenuation produced by oxygen and water vapor on a satellite-to-ground link has been carefully calculated and can be found in International Telecommunications Union (ITU) Recommendation 676 (Seema Sud 2008, private communication). The gaseous attenuation at microwave frequencies is largely absorptive and the absorbed energy will lead to an enhanced noise temperature. The enhanced noise temperature is related to two parameters of the gaseous cloud: the physical temperature of the gaseous cloud and the specific transmissivity, σ, of the gaseous cloud. The parameter σ varies between zero (i.e., nothing is transmitted through

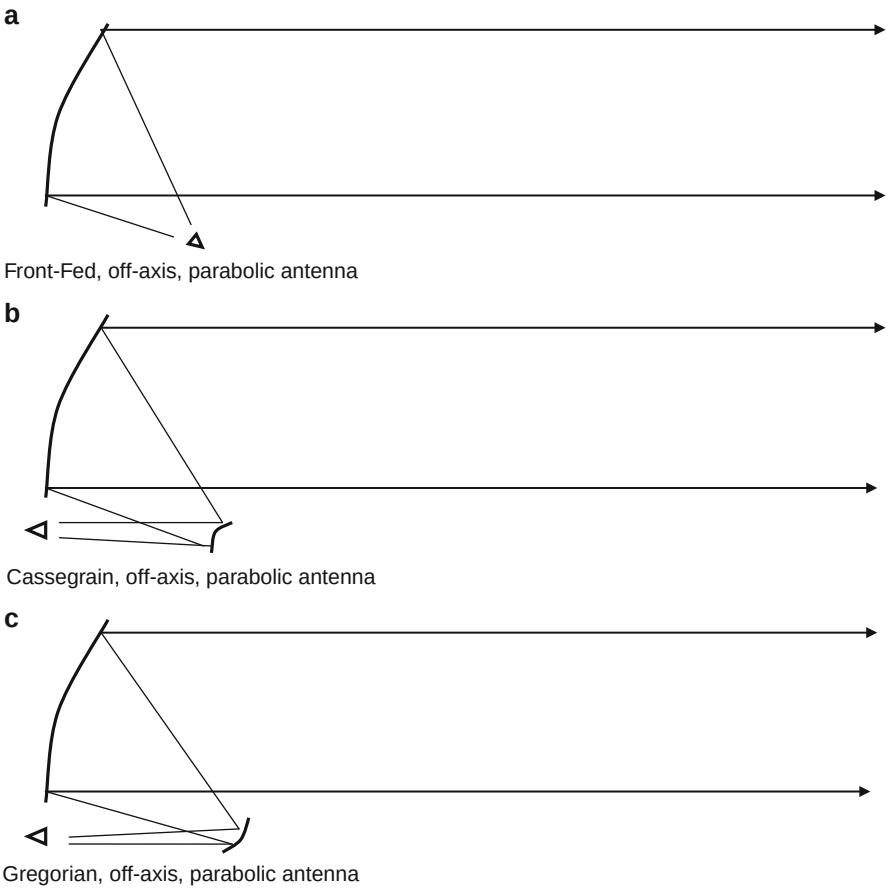


Fig. 18.10 Illustration of the three types of parabolic antennas that have their main feed system off-axis: front-fed, Cassegrain, and Gregorian. All three of the parabolic antenna types shown in are off-axis antennas. In (**a**), the feed is in front of the antenna reflector, while in (**b**) and (**c**) the feed is behind the antenna reflector. As in Fig. 18.9, the sub-reflector of the Cassegrain antenna has a curvature that is oriented in a similar direction to the parabolic main reflector, while the elliptical sub-reflector of the Gregorian antenna is oriented in an opposite sense to the parabolic main reflector

the gaseous cloud) and one (i.e., everything is transmitted through the gaseous cloud). This is illustrated in Fig. 18.12.

The apparent sky-noise temperature of the gaseous cloud, T_{sky}, can be directly calculated from a knowledge of the physical temperature of the gaseous cloud, T_m, by:

$$T_{\text{sky}} = (1 - \sigma) \times T_m \tag{18.5}$$

For example, if the gaseous cloud causes 0.2 dB of attenuation, we can convert 0.2 dB to an analog value $\Rightarrow$ 1.047. Inverting this analog value provides the value

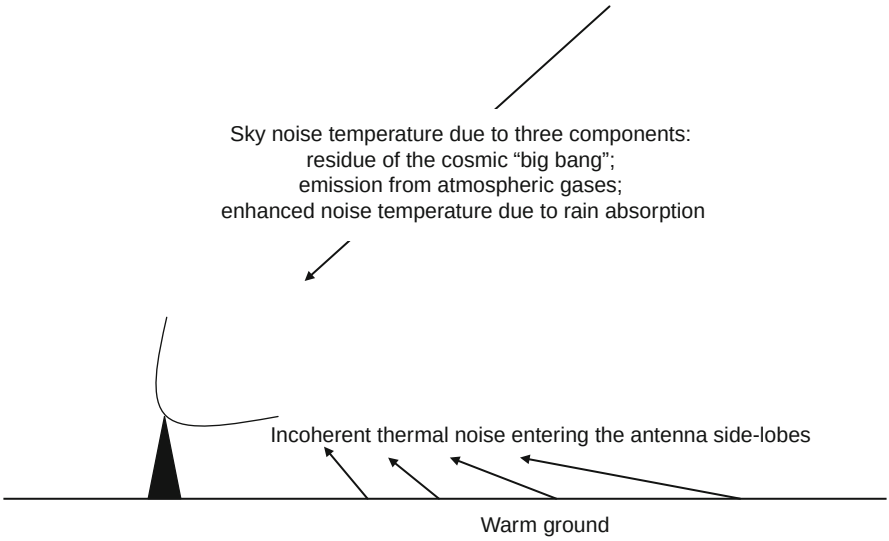


Fig. 18.11 Schematic of the external noise temperature components that can add to the system noise temperature of the receiver. The noise temperature components from directly in front of the antenna are due to three components: residue from the big bang (2.76 K), emission from oxygen and water vapor in the atmosphere, and enhanced emission from rain clouds. An additional component of noise temperature is that from the ground that is hot. The hot ground noise temperature enters through the antenna's sidelobes

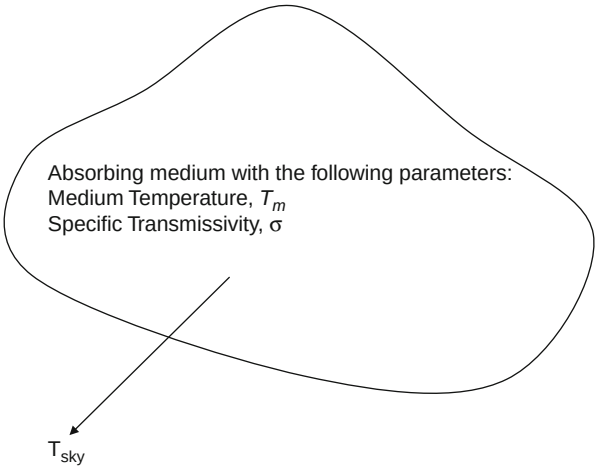


Fig. 18.12 Illustration of thermal emission due to absorption. An absorbing medium, which in this case is a rain cloud, causes a signal passing through it to lose some energy. If the signal entering the rain cloud has a power, S, then the signal leaving the rain cloud is $\sigma \times S$, where σ is the specific transmissivity of the rain cloud. Since energy is being absorbed by the rain cloud, it will emit at an enhanced temperature compared with clear sky, which is given by $T_{\text{sky}} = (1 - \sigma) \times T_m$

of $\sigma = 0.95$ and $(1 - \sigma)$ yields 0.05. If $T_m = 280$ K, then the value of $T_{sky} = 0.05 \times 280 = 14$ K. In this instance, a loss of 0.2 dB along the path due to atmospheric gaseous absorption gives rise to an increase in the antenna temperature contribution of 14 K.

Exactly the same increase in sky-noise temperature will occur in the presence of rain attenuation. If a rain cloud causes 3 dB attenuation ⇒ an analog loss of 2, which inverted gives a value of $\sigma = 0.5$, then $(1 - \sigma) = 0.5$. If the rain cloud is at a physical temperature of 280 K, then an additional 140 K will enter the feed of the antenna as an enhanced noise temperature contribution due to the presence of rain in the path. *This is a critical point to remember in antenna system design*.

If the system noise temperature of the receiving system is, say, 140 K in clear sky, then under the rain attenuation in the example above, the total noise temperature will be 280 K, made up of 140 K from the receiver and 140 K from the noise temperature emitted from the rain cloud. The system noise temperature has doubled under rainy conditions (a 3 dB increase) at the same time that the received signal power went down by 3 dB. The *C*/*N* has therefore dropped by 6 dB under a rain attenuation of 3 dB. This should always be remembered when calculating the system fade margin appropriate for the percentage time the receiving system should provide adequate *performance* and *availability*. Performance and availability are not the same when considering communications systems. Performance is the level of service, often measured as the Bit Error Rate (BER), delivered for a very high percentage of the time, generally 99%, or higher. Availability is the time the service is available above a usable threshold, again often measured in BER. For example, the performance level of a typical Ku-band digital satellite communications link will be set at a BER of 10^{-8} to 10^{-10} for at least 99% of the time, while the availability would be a BER of 10^{-6} for 99.7% of the time. For the early, and even some of the current, C-Band (6/4 GHz) satellite communications systems, achieving a BER of better than 10^{-6} is very difficult, but the propagation impairments are not significant, and so the availability margin is easily met. Conversely, because Ku-band (14/11 GHz) and Ka-Band (30/20 GHz) satellite communications systems have to provide fairly large fade margins to overcome rain attenuation, achieving a BER of 10^{-10} in clear-sky conditions is relatively straightforward. Achieving the availability margins is a different matter altogether due to significant rain attenuation at Ku-Band and above.

Rain Attenuation

Rain attenuation becomes significant at frequencies above about 10 GHz (Recommendation ITU-R P.618-9 2007). Rain attenuation is not closely correlated with the *total* amount of rainfall that accumulates on the ground over a relatively long period of time (e.g., 2 h), but it is directly correlated with the *rate* at which the rain falls. Rain fall rate "*R*" is measured in millimeters per hour (mm/h), and the rainfall rate is related to the *specific attenuation* "γ" by Eq. 18.6:

$$\gamma = kR^{\alpha} \text{dB/km} \tag{18.6}$$

Table 18.3 Regression coefficients for estimating the attenuation coefficients for specific attenuation, γ, where $\gamma = kR^{\alpha}$ (From Table 1 of reference [Recommendation ITU-R P.838-1 1999])

Frequency (GHz)	k_H	k_V	α_H	α_V
1	0.0000387	0.0000352	0.912	0.880
2	0.000154	0.000138	0.963	0.923
4	0.000650	0.000591	1.121	1.075
6	0.00175	0.00155	1.308	1.265
7	0.00301	0.00265	1.308	1.312
8	0.00454	0.00395	1.327	1.310
10	0.0101	0.00887	1.276	1.264
12	0.0188	0.0168	1.217	1.200
15	0.0367	0.0335	1.154	1.128
20	0.0751	0.0691	1.099	1.065
25	0.124	0.113	1.061	1.030
30	0.187	0.167	1.021	1.000

The subscript "H" refers to horizontal polarization and "V" refers to vertical polarization. The parameters γ is the specific attenuation in dB/km

Specific attenuation is a term used to describe the amount of path attenuation, in decibels, experienced over a kilometer, hence the units dB/km. The parameters k and α are given in Table 18.3 for both vertical and horizontal polarization (Recommendation ITU-R P.838-1 1999).

The rainfall rate, R, is usually measured on the ground by a rain gauge. It is therefore the rainfall rate at a point. Attenuation on a satellite-to-ground link, however, takes place along the path through the rain. The rainfall rate along this path is not uniform, but Eq. 18.6 assumes a uniform rainfall rate. The key step, therefore, in moving from specific attenuation, given by Eq. 18.6, to path attenuation (i.e., the total attenuation experienced along the path due to rain) is to find the equivalent distance "L" over which the rainfall rate can be assumed to be constant. The ITU-R prediction method (Seema Sud 2008, private communication) describes a procedure to calculate the effective pathlength L. The total path attenuation will therefore be given by:

$$\gamma = (kR^{\alpha}) \times L \text{ dB} \tag{18.7}$$

ITU-R Recommendation 618 (ITU-R Recommendation P.618-9, 2007) uses the point rainfall rate measured for 0.01% of an average year to calculate the path attenuation experienced for an average year at the percentage time 0.01%. This value is then extrapolated to lower, or higher, time percentages. It can be seen from Table 18.3 that the specific attenuation is very low for UHF frequencies, but increases rapidly above a frequency of about 10 GHz. Between about 10 GHz and 30 GHz, the path attenuation can be very approximately scaled as the square of the frequency, using dB to describe the attenuation. For example, if the path attenuation at 10 GHz is 3 dB, the scaled path attenuation along the same link for

Table 18.4 Examples of path attenuation for three frequencies. In the examples given below, the rainfall rate, R, is 50 mm/h, the effective pathlength, L, is taken as 4 km, and vertical polarization is assumed

Frequency (GHz)	$\gamma = (kR^{\alpha}) \times L$	Path attenuation (dB)
1	$0.0000352\ (50)^{0.880} \times 4$	0.004
10	$0.00887\ (50)^{1.264} \times 4$	5.0
20	$0.0691\ (50)^{1.065} \times 4$	17.8

Note that the more accurate formulation of path attenuation gives the attenuation at 20 GHz to be 3.56 times the attenuation at 10 GHz rather than a factor of 4 that would be given by the simple frequency squaring formula

the same time percentage at 20 GHz is simply given by the ratio of the square of the frequencies. That is, the attenuation at 20 GHz $= (20/10)^2 \times$ (the attenuation at 10 GHz) $= 4 \times 3 = 12$ dB. In this example, the attenuation at 20 GHz was 12 dB when the attenuation at 10 GHz was 3 dB. Table 18.4 gives some other examples using the parameters k and α from Table 18.3.

The effect of polarization is not critical for path attenuation, although attenuation for vertical polarization is generally less than or equal to that for horizontal polarization. However, the effect of rain and ice crystals on the polarization of a signal can be significant.

Depolarization

The polarization of an electromagnetic signal is given by the orientation of the electric vector, given the symbol E. There is always a magnetic field associated with any electric field, and this is commonly given the symbol M. Since both E and M propagate together, this is why the research area is called Electro-Magnetics. The E and M fields are oriented at right angles to each other (i.e., *normal* or *orthogonal* to each other). They are also mutually orthogonal to the direction of propagation of the electromagnetic wave. This is illustrated in Fig. 18.13.

The example shown in Fig. 18.13 is for a *linearly* polarized signal. That is, the direction of the electric (and thus magnetic) field is always oriented in the same direction, unless acted on by an external medium (e.g., rain or ice crystals) or by an ionized medium, such as the ionosphere. An electromagnetic signal can also be launched in such a way that the electric (and magnetic) fields are not oriented in a constant direction, but are rotating about the direction of propagation. This is shown in Fig. 18.14.

The E and M fields can rotate about the propagation direction in one of two directions. If the propagation direction is away from you, and the fields are rotating in a clockwise direction as viewed from your perspective, the polarization sense is called Right Hand Circular Polarization (*RHCP*). If the fields are rotating in the opposite direction to this, then the polarization sense is called Left Hand Circular polarization (*LHCP*).

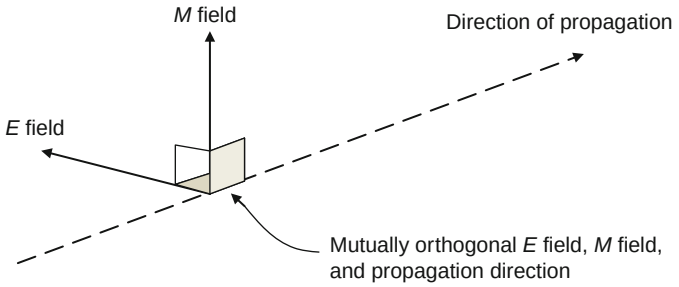


Fig. 18.13 Schematic of a linearly polarized electromagnetic signal, shown as *E* and *M* fields, propagating in the given direction. In the example above, the *E* and *M* fields are always oriented in the same direction, unless acted upon by an external medium (such as rain) or ionized media (such as the ionosphere). This form of polarized signal is referred to as a *linearly* polarized signal. If the electric (*E*) field is vertical to the local horizontal direction, then the signal is said to be *vertically* polarized. Similarly, if the *E* field is parallel to the local horizontal direction, the signal is said to be *horizontally* polarized.

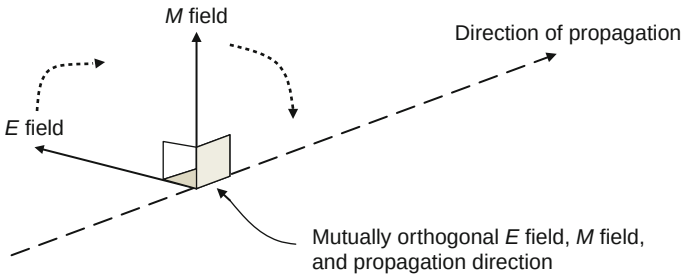


Fig. 18.14 Schematic of a circularly polarized electromagnetic signal, shown as *E* and *M* fields. In the example above, the *E* and *M* fields are rotating about the direction of propagation, while maintaining their mutually orthogonal orientation with respect to each other and the propagation direction. This form of polarized signal is referred to as a *circularly* polarized signal. In the example shown, the *E* and *M* fields are rotating in a clockwise direction with respect to the direction of propagation. This form of circular polarization is known as *Right Hand Circular Polarization* (*RHCP*). If the *E* and *M* fields are rotating in the other direction, they form what is known as *Left Hand Circular Polarization* (*LHCP*).

It is very rare that a polarization is "pure," that is, there is no residual energy in the opposite polarization sense. The general polarization state is elliptical polarization. There are two special cases for elliptical polarization: circular and linear. This is illustrated in Fig. 18.15.

Linearly polarized signals can be resolved into two orthogonal orientations (usually linear vertical and linear horizontal). For a purely polarized signal, there is no component in the orthogonal sense. If the signal is not a purely polarized signal, or if the signal encounters a propagation medium that causes the signal to

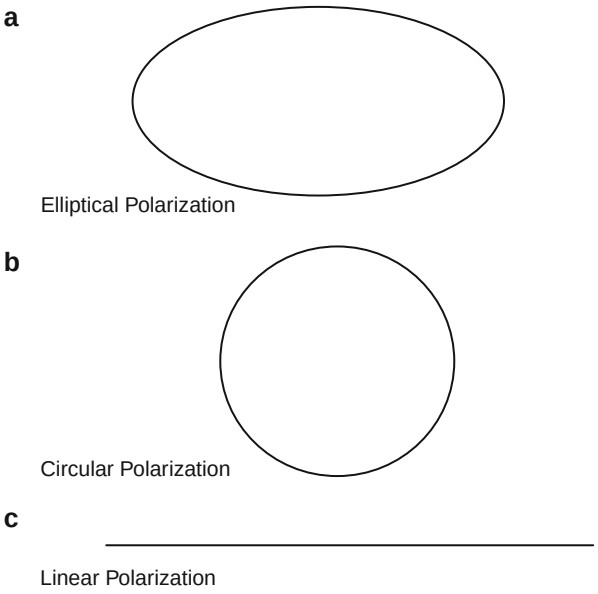


Fig. 18.15 Three examples of signal polarization. The general case for the polarization of a signal is elliptical polarization (**a**). The elliptical shape forms the locus of the electric (or magnetic) field vectors. There are two special cases for elliptical polarization: circular and linear. (**b**) The two axes of the ellipse (the semimajor and the semiminor) become the same and are equal to the radius of the circle. This is an example of pure circular polarization. By "pure" we mean that there is no component in the opposite sense of the polarization. The circle could represent a pure RHCP or pure LHCP signal, depending on the rotation direction of the signal with respect to the propagation direction. (**c**) Indicates the case when the semiminor axis of the ellipse becomes zero and the entire signal is represented by the semimajor axis. Since there is no component of the signal in the orthogonal sense, the signal is said to be a pure, linearly polarized signal

lose its purity of polarization (i.e., exhibit a depolarized component), then a signal component will be apparent in the orthogonal sense. This is illustrated in Fig. 18.16.

In Fig. 18.16a, the linearly polarized signal, shown as a vertically oriented linear signal, L_V, has no component in the orthogonal direction, L_H. The pure, linearly polarized signal then encounters something in the propagation medium (e.g., rain) and some of the energy in the vertically oriented signal is *depolarized* into the horizontal direction, L_H. This is shown in Fig. 18.16b. The resultant signal that emerges from the propagation medium, L_R, is the vector addition of the two orthogonal components, L_V and L_H. A receiver that is set up to receive both polarization senses (linear vertical and linear horizontal) will therefore receive two components of the original signal: one in the originally polarized sense (L_V) and one in the oppositely polarized, or depolarized, sense (L_H). If another signal is supposed to be entering the receiver in the L_H sense, then it will encounter interference from the depolarized component of the original L_V signal. Figure 18.16c shows a case of severe depolarization where half of the received energy from the original L_V signal appears as an L_H signal in the receiver.

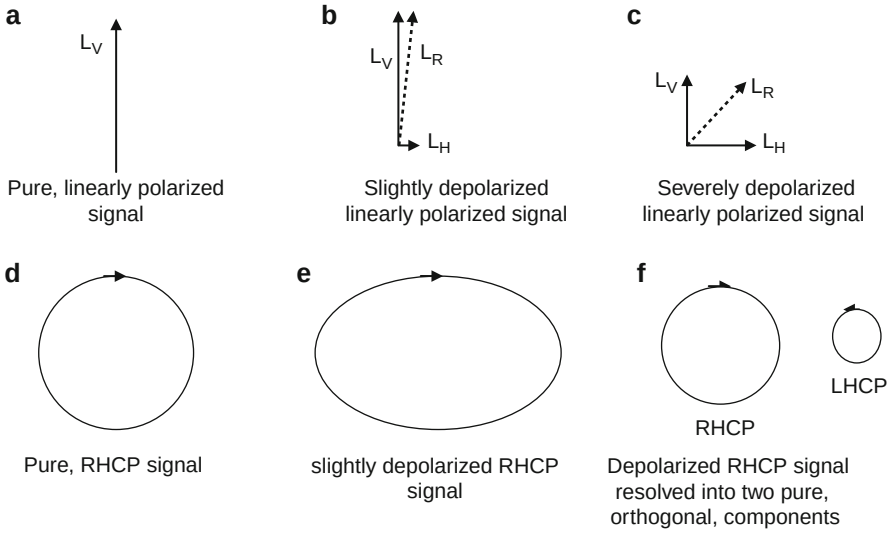


Fig. 18.16 Schematic representation of depolarization for linearly polarized and circularly polarized signals

In Fig. 18.16d, the circularly polarized signal, shown as a right-hand circularly polarized (RHCP) signal, has no component in the opposite sense, left-hand circularly polarized (LHCP). The pure RHCP signal then encounters something in the propagation medium (e.g., rain or ice crystals) and some of the energy in the RHCP sense is depolarized into the LHCP sense, leading to an elliptically polarized signal, shown in Fig. 18.16e. If the receiver is set to receive signals in both RHCP and LHCP, it will receive an LHCP component that has been depolarized from the original RHCP signal. This will cause interference in the receiver. In Fig. 18.16f, the elliptical polarization shown in Fig. 18.16e has been resolved into two orthogonal circularly polarized senses, RHCP (the wanted signal) and LHCP (the unwanted signal).

Telecommunications systems need to be able to discriminate between the wanted polarization, which carries the information signal, versus the unwanted polarization, which carries the residual, depolarized signal. For a dual-polarized receiver that is set up to receive two different signals at the same frequency, but in oppositely polarized senses, it is essential that the energy in the wanted signal is well above that of the interfering signal, which has been depolarized into the wanted signal's channel from the other polarization sense. The term *cross-polarization discrimination* (XPD) is used to describe the power difference between the wanted polarization and the interfering signal. If the rotation of the wanted signal (we will assume it is L_V) in Fig. 18.16b to the resultant signal (L_R) is $\Delta\theta$, the cross-polarization discrimination, XPD, is given by the equation $\text{XPD} = -20 \log_{10} \tan(\Delta\theta)$. The minus sign is used by convention to produce a positive value of decibels. The multiplier 20 is used instead of 10 because the XPD is a power ratio

and it is necessary to square the electric fields. For example, if the apparent rotation $\Delta\theta = 1°$, the XPD = 35 dB. An XPD of 35 dB means that more than 1,000 times more power is in the wanted orientation than in the unwanted orientation. This is an excellent value of XPD for a dual-polarized system. Typically, rain and ice crystals can cause significant depolarization along a path. A minimum operating XPD for digital communications systems is about 12 dB.

The apparent rotation of the linear vector in Fig. 18.16 is caused by two effects in the propagation medium, differential phase, and differential attenuation. *Differential* in this case means the difference in the level of the phenomenon between the two polarizations. In Fig. 18.16b, if the propagation medium has a different attenuating effect in the linear vertical polarization to that in the linear horizontal polarization, this is referred to as *differential attenuation*. Similarly, if there is a phase difference between the two linearly polarized vectors in Fig. 18.16b, this is referred to as a *differential phase*. Differential phase effects are dominant at C-band (6/4 GHz), since the attenuation in rain is very small to begin with. As the frequency increases to Ku-band (14/11 GHz) and to Ka-Band (30/20 GHz), rain attenuation starts to dominate and differential attenuation is the primary depolarization mechanism. The change from a differential phase-dominated depolarization mechanism to a differential attenuation-dominated depolarization mechanism as the carrier frequency used increases from 4 GHz to 30 GHz has an interesting systems effect. For each decibel of attenuation at C-band there is a much higher resultant depolarization effect than at Ku-band, and especially at Ka-band. The result is that depolarization is the dominant performance and availability limiting parameter at C-band while depolarization can be largely ignored as a limiting phenomenon at Ka-band: attenuation effects dominate the performance and availability margins at Ka-band. The margin provided to account for signal loss and depolarization in adverse propagation conditions is also significantly affected by the choice of modulation.

The choice of whether to use linear or circular polarization for an operational system involves a number of parameters to be considered. Linearly polarized antenna feeds are much simpler to design and build, and are thus cheaper. They also have generally much better on-axis XPD properties. A good dual-polarized linearly polarized antenna system can generally achieve a clear sky XPD of 30 dB without much difficulty, and more than 40 dB can be achieved. A comparable circularly polarized antenna system is normally limited to 27 dB, unless extraordinary care is taken in the design and construction. Circularly polarized antennas, particularly small earth station antennas, do not need to have their feed system aligned to the orientation of the satellite signal. The receivers are also largely unaffected by rotation of the electric vector, particularly due to Faraday rotation at C-Band. Geostationary satellite antennas appear to have better off-axis properties when circular polarization is employed.

Another factor to consider is the *rms* surface tolerance of the antenna being used for dual-polarization operation. To achieve adequate to good performance in single-polarization operation, an rms tolerance of about a quarter-of-a-wavelength is required for the antenna reflector surface. If dual-polarization operation is

contemplated, the rms surface tolerance must be on the order of a tenth-of-a-wavelength. The better the rms surface tolerance, the higher the manufacturing costs will be.

Modulation

Modulation is the technique used to modify one or more parameters of the transmitted signal so that information can be placed on the carrier. More importantly, it will permit the information to be retrieved at the receiver using a process called demodulation. There are a number of techniques used to modulate a digital carrier, the principal three being Amplitude Shift Keying (ASK), Frequency Shift Keying (FSK), and Phase Shift Keying (PSK). There are also combinations of the modulation techniques where, for example, both amplitude and phase are used to define a symbol, and these forms of modulation have the generic name *Quadrature Amplitude Modulation* (QAM).

A key point to realize in digital communications is that the information is *not* being sent as bits (ones and zeros) between the transmitter and the receiver: it is being sent as *symbols*. A symbol is a particular state that is impressed onto the carrier signal: it can be a change in level for ASK, and change in frequency for FSK, or a change in phase for PSK. A simple formula that connects the number of symbol states M in a digital modulation scheme with the number of bits that are used in each symbol, n, is:

$$M = 2^n \tag{18.8}$$

For example, if you are using Binary Phase Shift Keying (BPSK), there is 1 bit for every symbol, (i.e., $n = 1$) and so the total number of symbol states is $2^1 = 2$. For Quadrature Phase Shift Keying (QPSK), $n = 2$ and so $M = 4$, that is, there are four symbol states. These states can be expressed in bit form as 00, 01, 10, and 11. The more bits there are per symbol, the smaller the occupied bandwidth becomes. However, there is a downside to having a smaller occupied bandwidth: the more bits there are per symbol, the more carrier-to-noise power the systems need to develop to provide the same BER. Table 18.5 illustrates the increase in raw power needed to provide the same BER for a given modulation.

Clearly, there is a trade-off between occupied bandwidth and the C/N required to develop the required BER. In addition, as the frequency increases, so does the level of path attenuation in a given rain event, and so achieving the same BER for the same percentage of the time at Ka-band as was achieved at C-band will require a much larger fade margin. And, as we have seen, as the frequency increases, there will be a concomitant increase in the perceived antenna temperature under rain conditions. In many system designs, a useful combination parameter is used to characterize an earth station designs: G/T. The parameter G/T is the ratio of the gain of the antenna divided by the system noise temperature of the receiver.

Table 18.5 Carrier-to-noise required for m-QAM (Tim Pratt, 2006, private communication)

Modulation	Bits/symbol	C/N for BER $= 10^{-6}$ (dB)	Relative bandwidth[a]
BPSK	1	10.6	1.0
QPSK	2	13.6	0.5
16-QAM	4	20.5	0.25
32-QAM	5	24.4	0.20
64-QAM	6	26.6	0.17
256-QAM	8	32.5	0.125
1024-QAM	10	38.5	0.10

[a]Relative bandwidth means the bandwidth occupied relative to a modulation of BPSK

G/T

The link budget for a satellite-to-ground link can be expressed as shown in Eq. 18.9:

$$\frac{C}{N} = \frac{P_t G_t G_r}{kT_s B}\left[\frac{\lambda}{4\pi R}\right]^2 \tag{18.9}$$

where (C/N) is the carrier-to-noise ratio, P_t is the power transmitted in watts, G_t is the gain of the transmitting antenna, G_r is the gain of the receiving antenna, λ is the wavelength of the signal in meters, k is Boltzmann's constant, T_s is the system noise temperature in Kelvin, B is the bandwidth of the receiver in Hz, and R is the distance between the transmitting and receiving antennas in meters. It is useful to note in Eq. 18.9 that (C/N) is proportional to (G_r/T_s). Increasing the receiving antenna gain, G_r, will increase C/N and reducing the system noise temperature, T_s, will also increase C/N, and vice versa, of course.

The external noise temperature – the antenna noise temperature – of the earth station system will vary with the perceived noise temperature emitted by constituents along the path to the earth station. The antenna noise temperature will therefore vary with the total pathlength through the atmosphere. If the elevation angle of the earth station is reduced so that it can operate with a different satellite, the total pathlength through the atmosphere will increase, as shown in Fig. 18.17. If the pathlength increases through the atmosphere, the total absorption will increase due to additional constituents in the path. As a result, the perceived sky noise will also increase. This is shown schematically for a typical standard A earth station in Fig. 18.18.

As the elevation angle of a large earth station antenna is reduced, additional problems occur that cause tracking of a satellite to become more difficult.

Tracking

A communications satellite is considered to be in geostationary orbit if it is at geostationary altitude with an eccentricity of ≤ 0.001 and an inclination of $\leq 0.05°$.

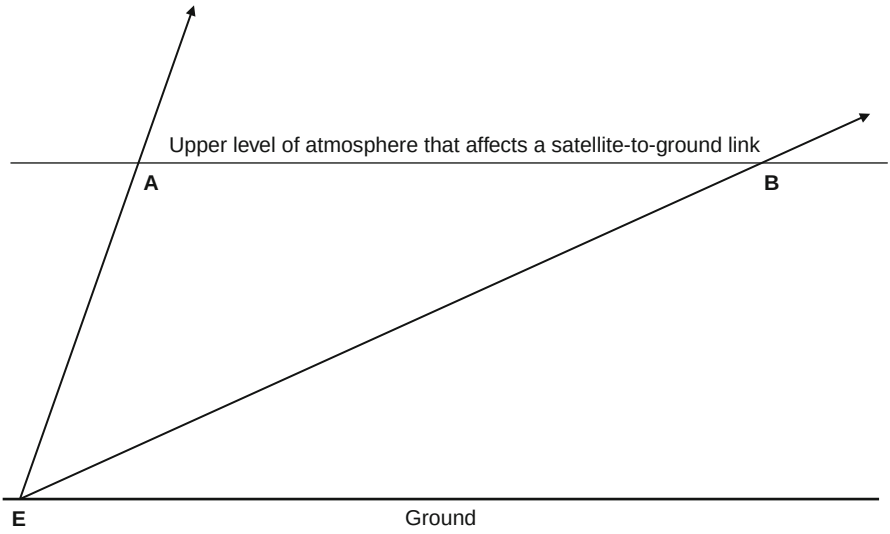


Fig. 18.17 Schematic showing the change in pathlength through the atmosphere. A high elevation angle link from the earth station, E, exits the atmosphere at A, and a relatively low elevation angle link from an earth station, E, exits the atmosphere at B. Path EA is shorter than path EB. Since the specific transmissivity, σ, of the atmosphere will decrease as the elevation angle is reduced, the resultant noise temperature emitted by the sky will increase as the elevation angle reduces. This can be seen in Fig. 18.14

For a geostationary satellite, orbital height + Earth radius = 35,786.03 km + 6,378.137 km (average) = 42,164.17 km. The *station-keeping box* for a geostationary satellite can therefore be seen to be $\pm 0.05^\circ$ east-west and north-south. Using Pythagoras' equation, the largest movement of the satellite in this box is 0.14°, from one corner of the box to the other, diagonally opposite, corner. If the earth station has a 1 dB beamwidth that is smaller than this, then the earth station will have to use tracking. (The 1 dB beamwidth is approximately half that of the 3 dB beamwidth.)

Tracking can be active (i.e., use is made of the incoming signal from the satellite to update the pointing of the antenna) or passive (i.e., use is made of the satellite ephemeris data to predict the position of the satellite and software code used to passively point the antenna toward the predicted position of the satellite). The cheapest form of tracking used initially for earth stations operating to geostationary satellites was *Step-Tracking*, sometimes called *Sequential Lobing* or *Hill-Climbing*. In this form of tracking, the satellite is initially acquired under manual control. The earth station tracking is then put under automatic control. The automatic controller then waits a given interval (15 min, sometimes longer), and the antenna is steered a given amount east and west about the nominal position of the satellite, and then north and south of that same position. These movements do not lose the satellite signal, since the angular movement is small. The antenna is then steered back to the point where the signal appeared to be a maximum.

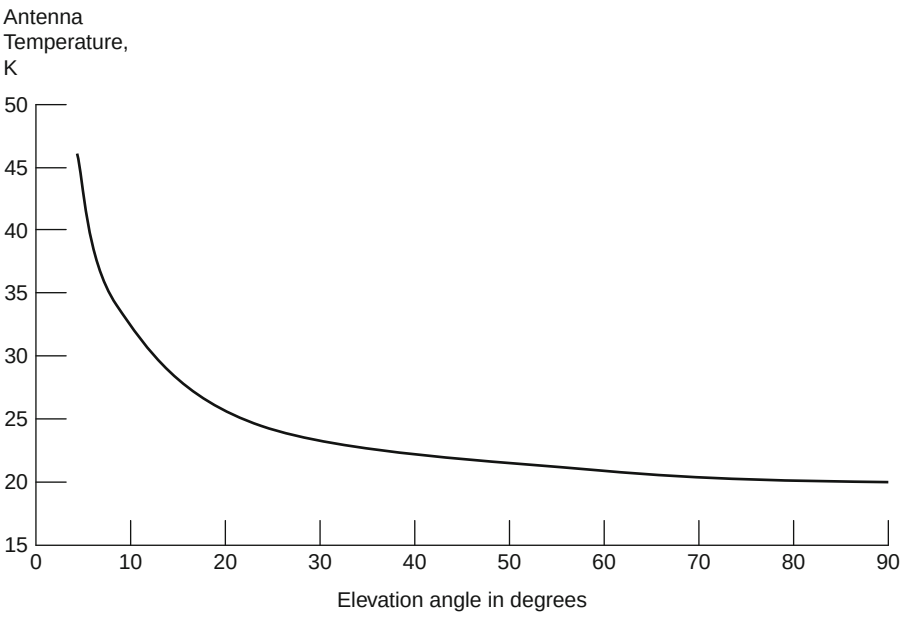


Fig. 18.18 Plot of a typical Standard A antenna noise temperature vs. elevation angle. The increase in the antenna noise temperature as the elevation angle reduces is due to two principal effects: (**a**) an increasing number of the antenna sidelobes intercept the ground, which is often at a temperature well above freezing; and (**b**) the path through the atmosphere becomes longer and so the absorption of the gaseous constituents leads to a concomitant increase in the noise temperature of the sky that is picked up by the antenna (see Fig. 18.11)

While this form of tracking works for targets that are moving very slowly (like a geostationary satellite) problems start to occur when the elevation angle becomes relatively low, especially below 15°. Below 15°, and especially as the elevation angle gets close to 5°, clear air propagation effects become increasingly significant. These effects can be summarized as ray bending, defocusing, angle of arrival, atmospheric multipath, antenna gain reduction, tropospheric scintillation, and low angle fading. The cumulative effect of these propagation problems is to prevent step-tracking antennas from operating effectively. Low-cost, program tracking can overcome much of these propagation problems that affect step-tracking, although they will not reduce the effect of the propagation impairments.

Active tracking that is used for some earlier radar systems is *conical scan*, but for large earth stations operating to geostationary satellites the best form of tracking is *monopulse* tracking. Conical scanning requires the main beam of the antenna to be spun about its mechanical axis, forming a "cone" around the target, and if the target moves, the energy difference between the sides of the cone allows corrective action to be taken. This type of tracking always receives a lower signal than that which

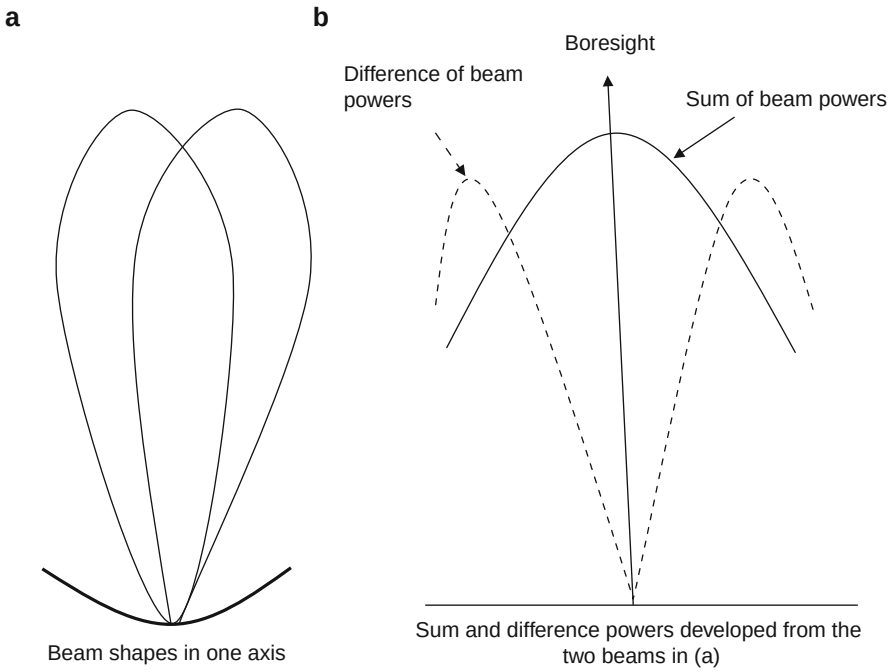


Fig. 18.19 Schematic representation of one axis of monopulse tracking. A monopulse antenna, in its simplest form, consists of four feed horns close to the focus of the parabolic main reflector. Two of the feed horns are orthogonal to the axis of the other feed horns. One of the axes is depicted. (**a**) The shape of the two main lobes created by the two feeds is shown, with the angular separation of the two beams exaggerated for clarity. (**b**) The power of the sum beam (adding the two beam powers together) and the difference beam (subtracting the two beam powers from each other) are shown. It can be seen that the antenna has only to move off-track by a very small amount for the power in the difference beam to increase significantly. The feedback tracking loop seeks to minimize the difference beam at all times

would be received on-axis, and so is not employed in satellite systems, where received power has to be maximized. Monopulse tracking, so called because it derives the pointing commands from one pulse (if it is radar) or one set of input signals received at the same time (such as in satellite antenna systems), uses four sets of input signals to develop sum and difference channels. Figure 18.19 illustrates the principle.

Monopulse tracking is the most accurate form of tracking available to earth station antenna systems. Whether the tracking is absolutely precise, or somewhat relaxed, it is very likely that the earth station will be located within a region where other systems operate on similar frequency bands, and so it may be necessary to protect the earth station antenna with what is known as *site shielding*.

Shielding

Earth stations often have to be sited in areas where there already exists a significant interference potential, not just from other satellite systems, but from terrestrial systems. An example of interference into a satellite earth station from a terrestrial source is shown in Fig. 18.20a.

Siting of earth stations is closely controlled by the national organization of the country the earth station is to be sited in. For the USA, this is the Federal Communications Commission (FCC). The control process is referred to as *coordination*. New antenna or satellite systems are required to *coordinate* with all preexisting systems to ensure that interference potential is at a minimum. In cases where potentially interfering signals can exist into, or from, a system operating in another country, *international coordination* is required between the affected countries, and this is generally administered through the offices of the International Telecommunications Union (the ITU). The ITU is based in Geneva, Switzerland.

The process of coordination, whether national or international, requires the calculation of the likely interference levels. In some cases, the distance between the two interfering systems is not sufficient to provide the required level of protection, and operators need to resort to ways in which they can artificially protect their antennas from interference. One of the most popular methods is *site shielding*.

Site shielding can be natural or artificial. One of the best natural shields is rolling terrain, or even better, a mountain range. In the absence of natural shielding, artificial shielding is resorted to. The simplest artificial shield is a metal fence, as depicted in Fig. 18.20b. Diffraction of energy is highest over an obstacle when the obstacle has a sharply defined edge, so site shields should be rounded, if possible. If sufficient space is available at the proposed earth station site, a shallow hole should be dug with the soil removed from the hole placed around the hole to form a raised rim, called a *burm*. This is depicted in Fig. 18.20c. Earth station operators often make the mistake of thinking the best location for an earth station is on the top of a hill, but this exposes the earth station to the maximum potential for interference. Locating the earth station in a shallow valley would be better than on a hill top.

Whether in a valley, a hill top, or the center of a city, the earth station antenna will be exposed to the elements and so consideration must be given to providing adequate protection to the antenna and feed from the weather.

Weather Protection

There are essentially three meteorological elements that the antenna systems will possibly need protection from, depending on the climate it is operating in: water, snow/ice, and the sun.

Water: Liquid water can cause significant attenuation at frequencies of 10 GHz and above. It can also cause oxidation on metallic surfaces. Feed covers are often used over feed horns to prevent water entering the feed. Care must be taken to ensure the feed covers are cleaned regularly.

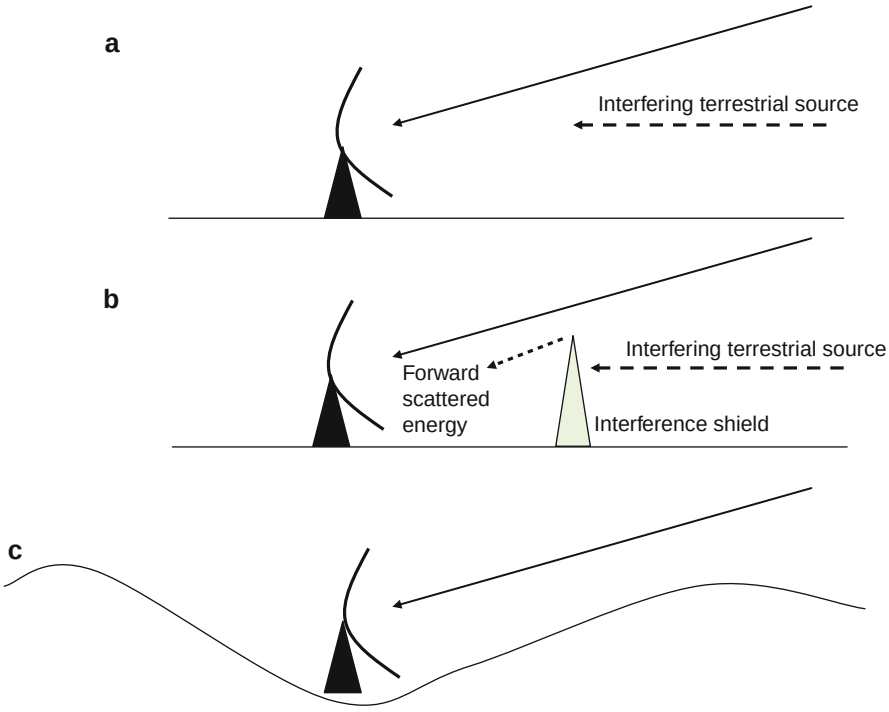


Fig. 18.20 Illustration of the use of site shielding. (**a**) A terrestrial interference source is shown entering the antenna main beam. In some situations, erecting a metal barrier in the form of an interference shield will provide adequate protection (**b**), but if the shielding fence is incorrectly designed, forward scattered energy can still disrupt communications in the earth station antenna. Possibly the best solution to use if there is sufficient space available is to dig a shallow hole and, with the soil removed from the hole, build a berm around the hole, as depicted in (**c**). The antenna would be located inside the shallow hole. Forward scattered interference is significantly reduced if the shield has a rounded top: the bigger the radius of curvature, the better

Snow/Ice: Snow and ice buildup on feed covers and antenna reflector surfaces can cause two effects. The first really is not in evidence until the temperature rises above 0°C, at which point the frozen particulates melt, and a layer of water will cover the feed cover or antenna reflector surface. The second is if there is a significant fall of snow onto the surface of a large parabolic antenna. A heavy layer of snow in one part of a large reflector can cause the reflector to distort out of a parabolic shape, thus lowering the gain of the antenna. If the earth station is sited in a climate where the temperature regularly falls below freezing point for many weeks, consideration should be given to heating the feed cover and the reflector surface.

Sun: Many medium-sized earth stations and VSATs are located in hot regions of the world. To maintain stable operation of the receiving equipment, it is not unusual to have the equipment box heated to a temperature above the maximum expected outside temperature. However, if the equipment box is located where it is possible to receive direct heating from the sun for several hours in the day, the temperature

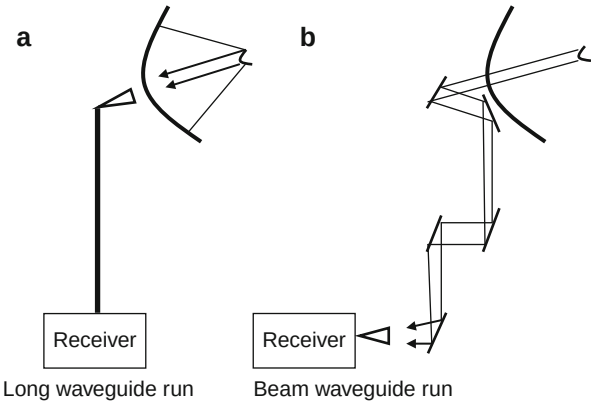


Fig. 18.21 Illustration of the difference between a standard waveguide run and a beam waveguide. Traditionally, a long waveguide run will connect the feed horn of the Cassegrain antenna to the receiver (**a**). The loss of the waveguide feed can be as high as 0.3 dB, leading to an increase in the noise temperature at the input of the receiver of about 18 Kelvin. A beam waveguide reduces the loss considerably, and a beam waveguide configuration is shown in (**b**). (Note: the reflector surfaces in the beam waveguide are not flat, but slightly concave to focus the beam within the narrow confines of the earth station physical structure. The earth station structure is not shown for clarity.)

inside the equipment box can go well above the anticipated temperature. In a VSAT located in Hong Kong, the equipment box was heated to 45°C, but direct heating from the sun caused the equipment box to reach an internal temperature of 70°C. To solve the problem, a sun shade was erected over the equipment box.

Many operators who have earth stations located in regions of the world where snow and freezing temperatures persist for several months house the complete antenna system inside a radome shelter that protects the entire antenna system from the elements. In such cases, care should be taken to ensure that the radome, and any particulates that adhere to the outside surface, do not degrade the performance of the link. In particular, if dual-polarized operation is contemplated with an antenna inside a radome, the depolarizing effects of the radome should be characterized.

Feed Systems

A simple rule of thumb to decide whether to employ a single-reflector or a dual-reflector configuration is as follows: if the aperture diameter is $\geq 100\lambda$, then a dual reflector is preferred; if the aperture diameter is $< 100\lambda$, then a single-reflector configuration should be used. In the section on system noise temperature, we saw that one of the major contributions to system noise temperature is the feed run that connects the feed horn to the receiver. In VSAT systems, where the antenna reflector is fairly small, the feed horn is usually offset from the mechanical axis



5.4 m antenna
© ViaSat Inc., Satellite Ground System Division
Reproduced with permission



11 m antenna
© ViaSat Inc., Satellite Ground System Division
Reproduced with permission

Both of the above earth station antenna systems were developed for operation to low earth orbiting satellites, although they could equally well be used for MEO and GEO satellites. In the example on the left, a Cassegrain configuration is used, with the sub-reflector attached to a cone that extends from the primary feed. In the example on the right, the much larger antenna employs four struts to hold the sub-reflector in a Cassegrain configuration. Both antennas are "Az-over-El", that is the elevation axis is supported above the azimuth axis. The smaller antenna on the left does not require counter-balancing loads to offset the weight of the main reflector, while the 11 m antenna on the right has two counter-balances to reduce the wind, and other, loading forces on the steering mechanism.

Fig. 18.22 Examples of X-band (approximately 12 GHz) earth station antenna systems developed by ViaSat

to minimize blockage. For a dual-reflector antenna, offset configurations are employed in some situations, but for really large antenna systems, with the aperture diameter 18 m, or larger, the dual-reflector configuration is almost always on-axis. The aperture blockage from the sub-reflector is not significant enough to warrant an offset-fed design. However, for large antennas, the feed run from behind the main reflector to the receiving system below can be long – perhaps more than 100 ft – and so the waveguide loss can create a significant noise temperature contribution (see Fig. 18.21).

To reduce the feed loss, and hence the noise temperature contribution of the feed, a *beam waveguide* configuration can be used. This is shown in Fig. 18.21b. A beam waveguide consists of a series of concave reflectors that conduct the received signal through the antenna support structure, down to the receiver. The system of reflectors looks like a type of periscope, and so antennas that use beam waveguides are sometimes called *periscope antennas*.

Conclusion

Antenna systems form a key link in the transmission and reception of signals from satellites. The antenna is designed to maximize the efficiency of receiving, or sending, electromagnetic signals. The design changes with the operating frequency, and the need to provide high gain and good sidelobe characteristics. While omnidirectional antennas permit mobile systems to operate without the user needing to know where the satellite is located, they considerably reduce the communications throughput due to their low gain. A parabolic main reflector is seen to provide the highest gain, with a number of designs available for specific systems (two examples are shown in Fig. 18.22). While offset-fed designs are preferred for small aperture sizes, as the size of the main reflector increases, so symmetrical, on-axis designs become optimum in terms of gain, design, and ease of construction. The location of the antenna is also important when considering interference issues, and whether to adopt a natural or artificial site shield. The choice of modulation is a critical system design parameter as it will impact the occupied bandwidth, the fade margin required, and the ability to resist interfering signals. Frequency reuse systems that employ dual polarization confer a significant increase in operational bandwidth, but at some cost. Weather characteristics in the operating region should also be part of the overall design of the antenna and its feed system.

References

J.E. Allnutt, *Satellite-to-Ground Radiowave Propagation* (Peter Peregrinus, London, 1989). ISBN 0 86341 157 6

J.E. Allnutt, *Satellite-to-Ground Radiowave Propagation*, 2nd edn. (IET, London, 2011). The 10 digit ISBN is 1849191506 and the 13 digit ISBN is 9781849191500

T. Pratt, C. Bostian, *Satellite Communications* (Wiley, New York, 1986)

T. Pratt, C. Bostian, J. Allnutt, *Satellite Communications*, 2nd edn. (Wiley, New York, 2002)

Recommendation ITU-R P.618-9, Propagation data and prediction methods required for the design of earth-space telecommunications systems (2007)

Recommendation ITU-R P.676-4, Attenuation by atmospheric gases (1999)

Recommendation ITU-R P.838-1, Specific attenuation model for rain for use in prediction methods (1999)

W.L. Stutzman, G.A. Thiele, *Antenna Theory and Design*, 2nd edn. (Wiley, New York, 1998). ISBN 0-471-02590-9

Technical Challenges of Integration of Space and Terrestrial Systems

19

John L. Walker and Chris Hoeber

Contents

Introduction 462
Mobile Satellite Systems and Ancillary Terrestrial Component 463
GBBF Development and Implementation 470
GBBF Calibration Scheme 474
Feeder Link Doppler Correction 475
Return Feed Element Path Gain and Phase Imbalance Correction 477
GBBF Ground Equipment 479
Inclined Operation of MSS Geo Satellites 481
MSS User Terminal Links and MIMO 482
Broadband Satellite Systems and Internet Access 484
Protocols and Network Performance 487
Reference Models 487
Role of Reference Models 489
ISO 7-Layer Reference Model for Open System Interconnect 490
TCP/IP Protocol Suite 491
Broadband Satellite Multimedia Protocol Architecture 492
Basic TCP 494
Network Environment and TCP Optimization 495
Classical TCP 496
Satellite Link TCP 497
TCP Performance-Enhancing Proxies (T-PEP) 498

J.L. Walker (✉)
Space Systems/Loral, 4706 Harbor Cover Ct., 94587 Union City, CA, USA
e-mail: walkerjo@ssd.loral.com

C. Hoeber
Space Systems/Loral, 3825 Fabian Way, 94303 Palo Alto, CA, USA
e-mail: hoeber.chris@ssd.loral.com

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_22, © Springer Science+Business Media New York 2013

Cross-Layer Signaling 501
TCP and Web Acceleration: Turbopage® 501
IP Routers in Space 501
References 504

Abstract

This chapter discusses the challenges of integrating space and terrestrial systems as well as some of the unique solutions and approaches to solving those challenges. While the first satellite systems were stand-alone and akin to a private network in today's terminology, virtually all current satellite systems are interconnected through some component of the terrestrial infrastructure: e.g., the Internet, PSTN, or private fiber. This chapter presents examples of the current challenges in integrating space and terrestrial systems by considering two satellite system classes which have unique requirements for interconnection and interoperability: the Mobile Satellite Systems (MSS) and the Broadband Satellite Systems for Internet Access.

Keywords

Ancillary terrestrial component (ATC) • DBSD • doppler • ground-based beam forming (GBBF) • inclined orbit • internet over satellite (IpoS) • internet protocol (IP) • internet router in space (IRIS) • ISO 7-layer • lightsquared • mobile satellite systems (MSS) • multiple input/multiple output (MIMO) • O3B • performance-enhancing proxies (PEP) • protocols • satellite • satellite phone • TCP/IP • terrestar • web acceleration

Introduction

This chapter discusses the challenges of integrating space and terrestrial systems as well as some of the unique solutions and approaches to solving those challenges. The History of Satellite Communications was presented in an earlier chapter, and began in the 1950s. The first satellites celebrated the ability for man to place a satellite in orbit and were used to gather information to evolve the art of space communications. The first active communication satellite was not launched until 1960, and the next few years' satellites were launched to demonstrate and provide a relay capability for teletype, voice, and even television. Since then, satellites and satellite systems have evolved to provide a variety of services.

In today's world, more and more things are interconnected. A person can surf the Web on their cell phone, track the location of their pet on their computer, and monitor and adjust the temperature of their home while they are away. They can remotely program their satellite television receiver to record a show their friend at the office recommends and watch it when they get home, or even watch the show remotely on their laptop or cell phone.

While the first satellite systems were essentially stand-alone and akin to a private network in today's terminology, virtually all current satellite systems

are interconnected through some component of the terrestrial infrastructure: e.g., the Internet, PSTN, or private fiber. The complexities of satellites have evolved since their beginning, and so has the need and depth for satellite systems to connect and operate with terrestrial systems and infrastructure.

This chapter presents examples of the current challenges in increasing this interconnection and interoperability by considering two satellite system classes which have unique and intricate requirements for terrestrial integration: the Mobile Satellite Systems (MSS) and the Broadband Satellite Systems for Internet Access.

In the first section, the challenge of market economics for mobile satellite operators is discussed along with the advent of the Ancillary Terrestrial Component (ATC). Key challenges for MSS operators are summarized, followed with a discussion into current and future solutions to solving those challenges. A synopsis of the recent MSS deployments is provided along with the observation of the importance of ground-based beam forming (GBBF) as an enabler for this industry. The GBBF approach is described, including an overview of the architecture and the challenges in developing the concepts, implementing and deploying the system. All of the recent geosynchronous MSS satellites are in an inclined orbit, and the reasons are discussed.

The MSS integration challenges are not limited to connecting the Gateway with the PSTN and Internet, but include the additional challenge of integrating the satellite user terminal (satellite phone) with the ATC. Finally, Multiple Input/Multiple Output (MIMO) systems are described in the context of addressing challenges in satellite link performance.

The chapter moves next to address the challenges in broadband satellite systems for Internet access. The challenges are categorized into three areas, followed by discussions of current and future approaches to their resolution. A brief discussion on the market challenges is provided as this leads to the innovations required by the technical solutions.

Protocols and reference models are discussed with the goals: to indicate the development frameworks in which protocols have been and are being developed; to indicate many of the protocols currently in use in both terrestrial and satellite terrestrial integrated systems; and to discuss some of the specific challenges in current protocol development.

One of the ongoing challenges facing the integration of satellites and terrestrial networks concerns the protocol implementation of the Transport Layer Protocol, TCP. Since a large amount of Internet traffic is directly related to TCP, the issues and solutions envisioned are discussed including some examples of current implementations.

Finally, a section is provided to present the challenges, potential advantages, and disadvantages of placing an Internet router in space.

Mobile Satellite Systems and Ancillary Terrestrial Component

Mobile Satellite Systems (MSS) present unique challenges to both the space and terrestrial cellular industries. The MSS category refers to satellite systems where

the user communication terminals are mobile and portable. The user frequency band is smaller than the broadband category to be discussed later, although the term "broadband" is sometimes used in the context of particular MSS data services such as Web browsing. The user frequency bands typically span 20–40 MHz one-way. For example, the FCC licenses MSS services in the 2 GHz (1990–2025 MHz and 2165–2200 MHz), the L-band (1525–1544 MHz/1545–1559 MHz and 1626.5–1645.5 MHz/1646.5–1660.5 MHz), and the "Big LEO" (1610–1626.5 MHz and 2483.5–2500 MHz) band.

Market economics are extremely challenging for mobile satellite operators, as the build-out of terrestrial cell phone services has been rapid. In addition, the terrestrial cell phone technologies and services have evolved rapidly through the progression of 2 G, 3 G, and 4 G digital mobile communications.

In the 1990s it was recognized that MSS operators were challenged with the market economics to provide mobile services with commercially successful systems. The build-out of terrestrial cell phone coverage was expanding, as well as the March to improve cell phone technologies, increase the services offered, and the enhance the quality of service provided. Nevertheless, there were areas unserved by terrestrial towers that a satellite-based solution could address.

The FCC recognized the value of MSS to provide advanced communications to areas not readily or economically served by terrestrial systems and released a notice of proposed rulemaking in 2001 (FCC 2001). The consideration was to allow the MSS operator to use terrestrial base stations to augment the satellite coverage. In this rulemaking, the satellite provides the primary communications. However, when the link performance is poor due to foliage and terrain, or when the users are in buildings, the terrestrial fill-in base stations provide the communications link. In 2003, the first rulemaking was released authorizing MSS operators to add these base stations, referred to as the Ancillary Terrestrial Component (ATC), to their networks.

The approach has continued to evolve with the latest rulemaking in 2005 (FCC 2005). The FCC set a gating criterion that all MSS/ATC equipment must be able to communicate via both the satellite and the ATC, and that the services offered must be available through both the MSS and the ATC. While this provided economic opportunity for MSS operators, key technical challenges needed to be solved, some of which are still ongoing as of the time of this writing.

Interestingly, the challenges presented in developing and integrating Mobile Satellite Systems are not limited to integrating the satellite gateway with the terrestrial Internet and telephony networks. These challenges extend to providing adequate quality of service to handheld users, coordination and non-interference between satellite systems, and efficient sharing of the user spectrum between those users linked through the ATC and those linked through the satellite. In addition, challenges exist in the development of user terminals which must consider dual mode (satellite and terrestrial) and the rapid advance of terrestrial cellular phone applications and services.

To summarize, the key challenges for MSS are:

1. User terminal handset compatibility with terrestrial and satellite communications
2. Uplink and downlink interference management within and between MSS/ATC operators



Fig. 19.1 TerreStar GENUS phone

3. Commercially successful viability in offering services via MSS/ATC with the continuing rapid advance of terrestrial networks and service evolution: 2 G, 3 G, 4 G
4. Flexibility of system architecture to adapt to market evolution

For challenge 1, the FCC rulemaking is limited to addressing the need for the user terminal to have compatibility and common services within the MSS/ATC operator network. However, when challenge 3 is folded in, consideration must be given to multi-mode user terminals which can communicate not only with the MSS satellite and corresponding ATC, but also have roaming and compatibility with a standard terrestrial network provider.

Multi-mode phones eliminate the need for the consumer to have both an MSS handset for rural and remote areas, and a terrestrial cell phone for more economic connection as well as the likely advanced services which have evolved more rapidly in the terrestrial networks. TerreStar networks pursued this approach and developed the GENUS phone which is a multi-mode smart phone as depicted in Fig. 19.1. In addition, they developed business relationships with AT&T to offer both satellite and cellular communications on one device with a single phone number on a single bill. Smartphone features such as text, email, contacts, and calendar were included in both satellite and cellular mode (TerreStar).

In considering challenge 3 and challenge 4, along with meeting aggressive milestones to maintain FCC licensing, the need was prevalent for a very flexible satellite system architecture. MSS operators needed spot beams and frequency reuse plans that could be modified as the ATC roll-out was implemented and as the traffic and service demands evolved. Combining this requirement for flexibility

with the need for shorter satellite schedules and the necessity of keeping the total satellite costs as low as practical led to the development of the Ground-Based Beam Forming (GBBF) solution. Although all the MSS-ATC systems employ similar GBBF technology, these U.S.-Based networks have all suffered suffered financial reversals for a variety of different reasons as discussed in ▶ Chap. 4 "Space Telecommunications Services and Applications".

At the time of this writing, all of the recent MSS programs (DBSD, TerreStar, LightSquared) employ Ground-Based Beam Forming (GBBF). With GBBF, beams can be added, removed, or reconfigured to enable a satellite to operate from different orbital locations or to adapt to changes in traffic patterns and evolve to service new applications. Further, with beam forming performed on the ground, the cost and time to deliver a highly flexible satellite is significantly reduced.

The worlds' first satellite to utilize a two-way Ground-Based Beam Forming (GBBF) system, DBSD G1 (previously named ICO G1), was launched from Cape Canaveral, Florida on April 14, 2008. Designed and built by Space Systems/Loral (SS/L), G1 has a 12 m unfurlable reflector and provides 250 fully configurable transmit and receive beams covering the continental US (CONUS) including Alaska, and Hawaii and Puerto Rico. SS/L was first in developing and deploying a two-way GBBF and the on-orbit performance met or exceeded all the original requirements. SS/L selected and subcontracted to Hughes Network Systems LLC (Hughes) to develop and implement the high-speed signal conditioning and processing ground equipment that enables GBBF.

The Terrestar-1 satellite, also built by SS/L, was launched a little over 1 year later in July, 2009 on an Ariane-5 in French Guiana. This was the largest, most advanced commercial communications satellite launched at the time. It hosts an 18 m unfurlable reflector and extends the coverage over the DBSD G1 to include Canada.

Light squared, formerly MSV Skyterra, launched SkyTerra 1 from the Baikonur Cosmodrome, Kazakhstan on November 14, 2010. This satellite incorporates a 22 m unfurlable antenna and an on-board digital channelizer to ease with interoperability of other L-Band systems within the fragmented L-Band spectrum. The GBBF ground equipment is provided by Viasat. The coverage area is extended even further to include Mexico and Central America.

Central to these MSS/ATC systems is the need to share the spectrum between the satellite and the ATC base stations. Commercial viability requires a large numbers of users. The frequency reuse schemes, along with the flexibility of the satellite network to place beams, adjust beam power and to perform interference suppression is key to maximizing the overall ATC and satellite capacity.

In an optimized architecture the ATC cell size is much smaller than the satellite beam cells. Figure 19.2 illustrates the hybrid architecture of a MSS with ATC. The GBBF provides flexibility of assigning user frequencies to beams, adjusting the size and shape of the formed beam and varying the amount of power applied to each beam. The only limitation on the flexibility is the geometric constraints of the antenna optics on the satellite. Large unfurlable reflectors are used to achieve high G/T values and provide the link performance required to communicate reliably with cell phone style satellite user terminals.

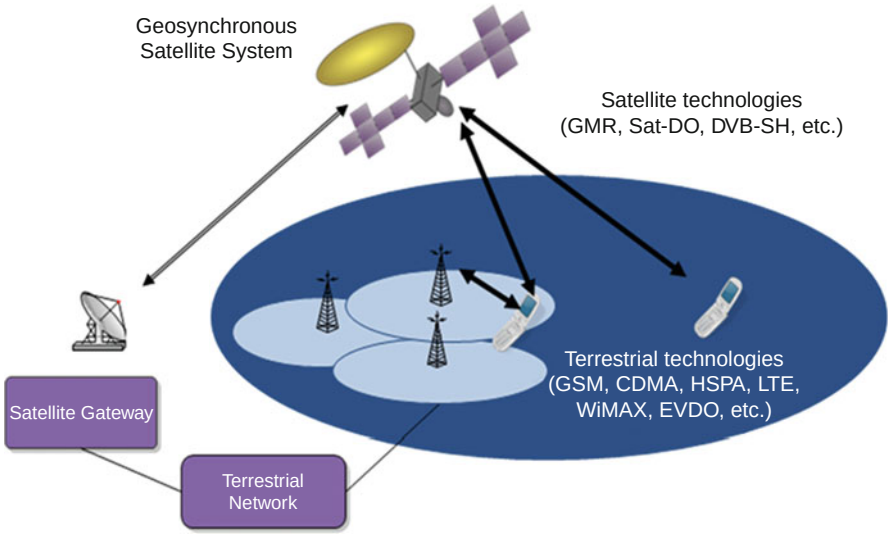


Fig. 19.2 Hybrid Mobile Satellite System (MSS) with ancillary terrestrial component (ATC)

Interestingly, the three MSS systems mentioned deployed increasingly larger reflector size and coverage area. Going from 12 m to 18 m to 22 m and increasing the coverage area drives additional payload equipment to handle the increased feed element count in the feed array that illuminates the reflector. This trend is likely to continue with further miniaturization of payload hardware and the advantages in link performance to mobile users.

Perhaps an even more important factor to drive an increased reflector size is related to challenge 2: uplink and downlink interference management within and between MSS/ATC operators. The size of the smallest spot beam that can be formed is not limited by the GBBF, but rather is limited by the aperture size of the satellite antenna. For example, a 12 m reflector would have a .47 deg 1-dB beamwidth, or 1.2° null-to-null and span 775 km on the Earth. Whereas, A 22 m reflector can form a spot beam at S-Band with a 0.26° 1-dB beamwidth, or 0.66° beamwidth null-to-null. The null-to-null angle would span approximately 430 km in diameter on the Earth from the geosynchronous orbit.

The larger reflector size and correspondingly smaller spot beam results in increased capacity for the satellite system. The smaller spot beam directivity rolls off much quicker over the geography than a larger beam, so the frequency can be reused more times over the coverage area. In addition, the smaller beam provides increased ability for interference cancelation to enhance the capacity of the satellite and ATC total network and increased directivity for the RF links.

In summary, the GBBF offers several advantages to conventional satellite systems. First, it simplifies the satellite design when compared to on-board processing solutions. Second, the GBBF architecture is very robust to changing markets and business plans. The GBBF equipment on the ground uses flexible

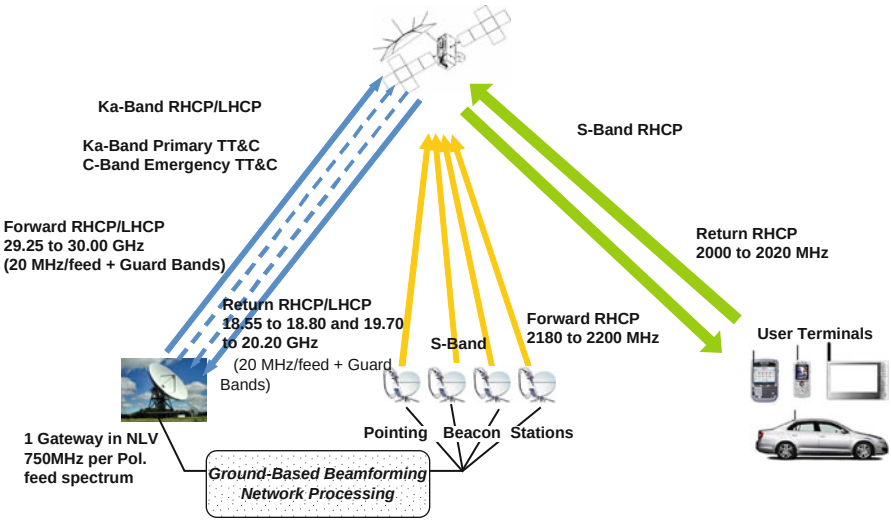


Fig. 19.3 DBSD space segment architecture with GBBF

digital signal processing hardware and software which allows beam patterns to be easily changed, allowing the system to provide EIRP and G/T to match traffic demands. Also, the ground algorithms and equipment can be readily upgraded to enhance system capability as needed and is readily accessible for repair and replacement.

As an example to understand the architecture and terminology of these GEO MSS GBBF systems, consider the DBSD G1 Satellite and GBBF architecture shown in Fig. 19.3.

The DBSD Space Segment and GBBF consists of the G1 Satellite, the Gateway, and four Pointing Beacon Stations (PBS). The Gateway includes the Radio Frequency Subsystem (RFS) and the Ground-Based Beam former (GBBF) subsystem. The basestation complement to the user terminals is also included at the Gateway. The G1 Satellite services S-Band user terminals and employs a Ka-Band feeder link. The GBBF is integral to the workings of the G1 Satellite and is therefore included with the satellite in what is defined as the Space Segment. Consequently, specifications were developed to include the GBBF and RFS in what is normally specified for the satellite alone: e.g., G/T, EIRP, pointing error. The GBBF assemblies are physically located within the Gateway in North Las Vegas and the PBS Stations are located at four locations within the contiguous United States (CONUS). The communication signal flow places the GBBF assemblies in between the user terminal head-end, named the Mobile Satellite Systems (MSS) Communications Processing equipment (MCP) and the Gateway RF Subsystem.

The return link is defined as the signal path transmitted from the user terminal on the ground up to the Satellite and down to the Gateway, through the GBBF and to the MCP. Therefore, the return link consists of both an uplink from the user

terminal to the Satellite and a downlink from the Satellite to the Gateway. The forward link is defined as the signal path transmitted from the MCP source through the Gateway up to the Satellite and down from the Satellite to the user terminals on the ground. Likewise, the forward link consists of both an uplink from the Gateway to the Satellite and a downlink from the Satellite to the user terminal receiver.

The Space Segment provides service to the user terminals via the forward (transmit) and return (receive) beams in the coverage area. The number of beams and their shapes and sizes are controlled by the GBBF.

The communication paths between the Satellite and Gateway are referred to as the feeder links. The forward feeder link corresponds to the Ka-Band transmission from the Gateway to the Satellite (uplink). Similarly, the return feeder link corresponds to the Ka-Band transmission from the Satellite to the Gateway. The Ka-Band feeder link requires both polarizations simultaneously in order to accommodate the forward and return signals to and from each feed element of the satellite S-Band antenna used to form the beams to communicate with the S-Band user terminals.

The Gateway consists of a Radio Frequency Subsystem (RFS) which includes a 13.2 m Ka-Band dish, High Power Amplifiers (HPAs), and up and down conversion equipment which connect to the GBBF at an intermediate frequency of 140 MHz.

The GBBF subsystem consists of two GBBF assemblies with 4 racks of equipment, and four Pointing Beacon Station (PBS) transmitters. The GBBF assemblies are located in the Gateway and the PBS transmitters are positioned at four locations over CONUS to aid with spot beam pointing control.

The Satellite S-Band antenna consists of a 12 m reflector and a 46 element feed array for CONUS coverage, with two additional feed clusters which employ a traditional on-board beam forming network and provide coverage for Hawaii and Puerto Rico.

As will be described, the Satellite in the GBBF Space Segment architecture is essentially a bent-pipe transponder converting each feed element in the Satellite Feed Array between S-Band and Ka-Band. The GBBF ground equipment processes the received signals from each feed element to form the beam outputs in the return direction. Likewise, in the forward direction, the user traffic signal is applied to the desired beam input to the GBBF, where the GBBF decomposes the signal into the individual signals to be applied to each of the Satellite S-Band feed array elements. These signals are then sent up to the Satellite on the Ka-Band feeder link where they are transponded, amplified, and sent to the S-Band feed array. The individual signals from each feed element radiate and combine in the far field to form the desired transmitted user traffic signal level over the coverage area (beam).

In summary, the DBSD G1 Satellite and GBBF architecture and implementation offers the following features: (Walker et al. 2010)

- GBBF can form up to 250 beams independently in each direction.
- GBBF can form beams of various sizes and shapes.
- GBBF can use between 0 and 20 MHz in each beam.
- GBBF is flexible to assign spectrum to any beam.

- GBBF is flexible to assign power to beams.
- GBBF is flexible to use any modulation scheme.
- GBBF is flexible to accommodate Ancillary Terrestrial Component (ATC).

GBBF Development and Implementation

To understand GBBF operations, we will begin with a discussion of beam forming independent of where it is accomplished. Beam forming is commonly used in radar and communication systems by combining the signals from a multiple element array (Johnson and Jasik 1984). Figure 19.4 illustrates beam forming operations in the return direction.

In the return direction a signal transmitted from the ground is received at the feed element array depicted on the left of the figure. Each feed element of the array receives the signal at a different amplitude and phase due to the spatial geometry between each feed element of the satellite antenna and the point on the ground where the transmitter is located. The beam forming operation is shown on the right of the figure and consists of taking the signal from each feed element array and applying a beam forming coefficient weight to shift the amplitude and phase of the received signal by A_c and Θ_c. A spot beam is formed by combining the beam coefficient weighted feed element signals as shown by the Sum block. By choosing different coefficient weightings (A_c, Θ_c), different beams may be formed.

Each A_c and Θ_c value is required to be a specific number to form any particular beam contour. By changing the values of A_c, and Θ_c the formed beam is changed. For example, by changing the values of the weighting, a beam may be steered in pointing, narrowed to a spot, broadened to a regional beam, or designed to have additional sidelobe suppression or even nulls at specific geographical points. Further, virtually any number of beams may be formed by replicating the weighting and sum operations with various combinations of the feed element signals.

Beam forming by weighted combination of feed elements is graphically illustrated in Figs. 19.5 and 19.6. Figure 19.5 shows the raw feed element antenna pattern on the ground. By amplitude and phase weighting these feed elements with the proper coefficient weights, the raw feed patterns are converted into spot beams as shown in Fig. 19.6.

The beam forming operation, as described so far, is well understood and has been used extensively in communication systems of all types. In Ground-Based Beam Forming, the beam forming operations described are moved to the ground as illustrated in Fig. 19.7.

The signals from each feed element of the array are transponded to the Gateway through the feeder link and sent to the GBBF assembly. The GBBF assembly has an input corresponding to each feed element in the array. Within the GBBF, the beam forming operations occur as shown in the diagram and the GBBF provides output ports for each of the formed beams.

By moving the beam forming operations to the ground, any amplitude and phase mismatch between the transponded pathways must be determined and compensated

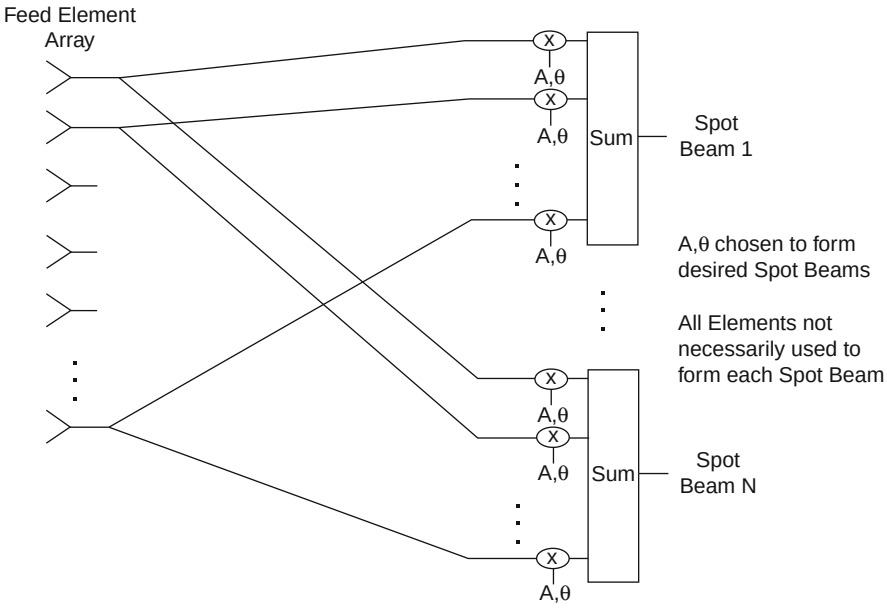


Fig. 19.4 Return basic beam forming operations

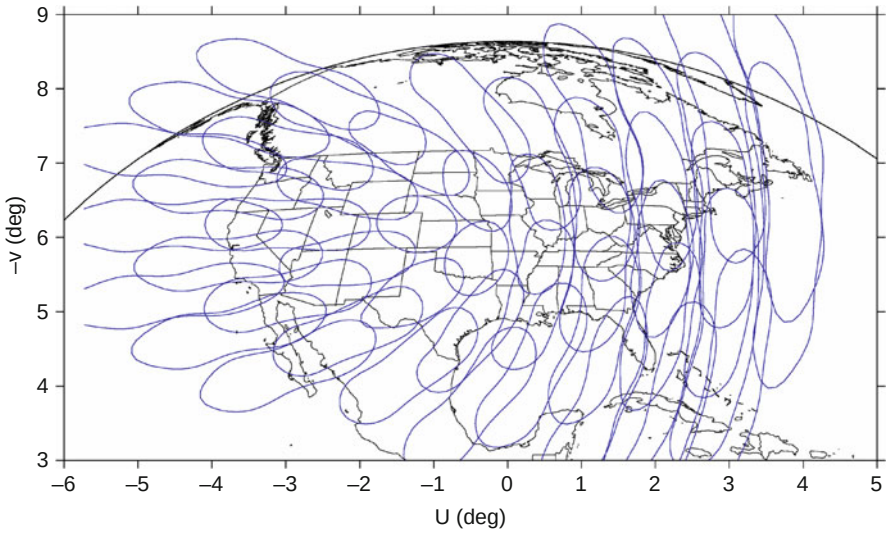


Fig. 19.5 Forty-six return feed element 3 dB contours before beam forming

for to ensure the desired beams are formed correctly. These pathways include the Satellite payload, the Ka-band feeder link propagation, and the Gateway RF subsystem. Any amplitude and phase fluctuations in any of these three components will alter the formed beam pattern if left uncompensated.

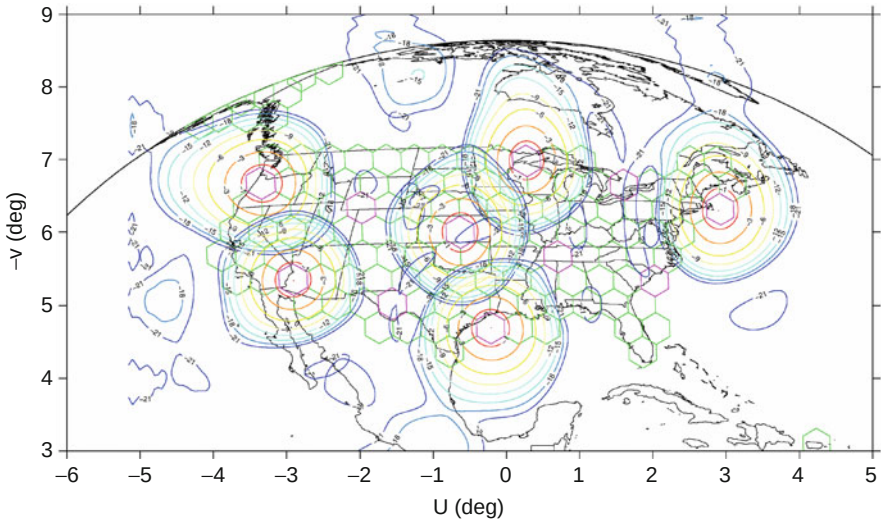


Fig. 19.6 Six return spot beam contours after beam forming

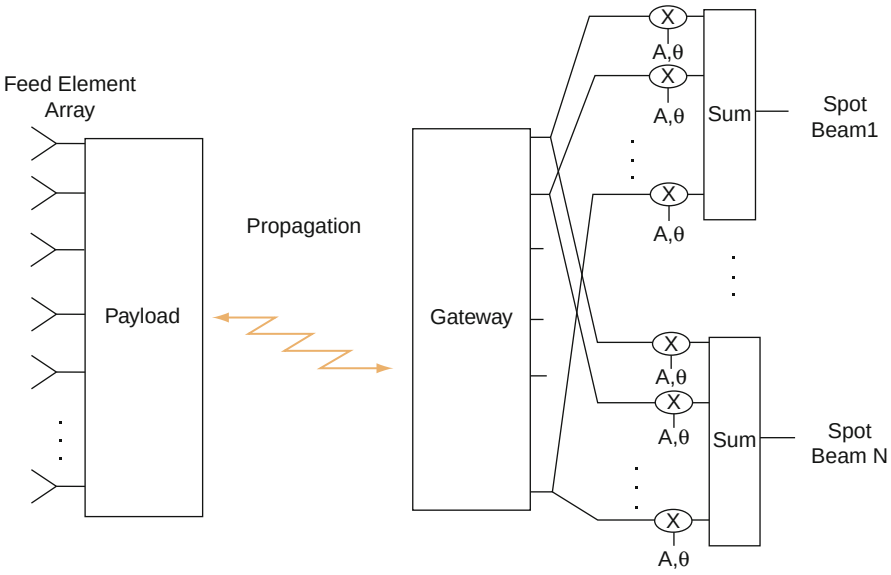


Fig. 19.7 Ground-Based Beam Forming

The Ka feeder link contains each feed element frequency division multiplexed along with special calibration channels. Thus, the satellite payload may be thought of as a bent-pipe transponder for each feed element as depicted in Fig. 19.8. This approach not only simplifies the payload design, but allows the

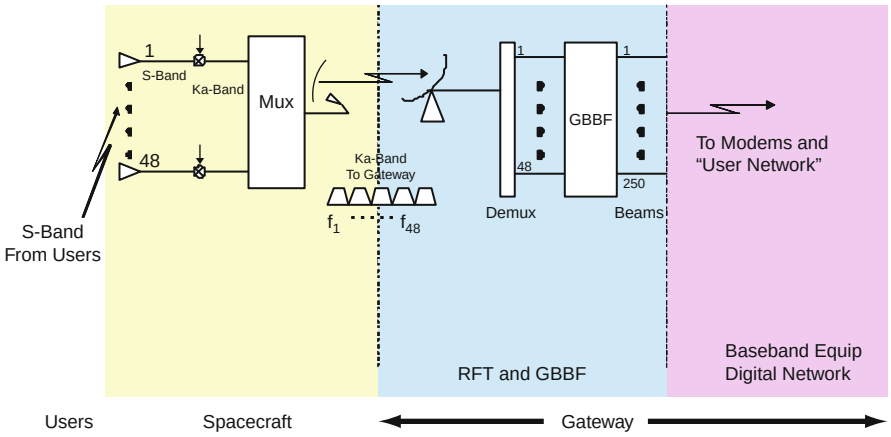


Fig. 19.8 Return payload is a bent-pipe transponder for each feed element

payload manufacturing and test operations and procedures to be performed in the usual manner for transponder satellites.

With such flexibility as provided by the architecture to form any number, size, or shape of beams, it is necessary to define a reference set of beams at the beginning of the program to drive the design and to quantify the performance of the Satellite and GBBF. DBSD and SS/L chose the reference case depicted in Fig. 19.9 and developed the primary specifications for performance for this scenario.

This reference beam set consists of 135 beams total, with 124 beams over CONUS, 9 beams over Alaska, plus 1 beam each over Hawaii and Puerto Rico. There are 133 hexagonal cells over CONUS and Alaska which correspond to geographic areas. The Space Segment forms spot beams, using the GBBF and the Satellite S-Band antenna, which are approximately centered over these geographic cells. The communications performance of the Space Segment is defined in terms of the beam performance achieved in each of these geographic cells.

It is important to understand that the cells represent the geographic definition and the actual formed beam contours extend beyond the cell as shown by the beam contour lines in Fig. 19.6. The reference case cell edges correspond to approximately 1 dB down from the peaks of the formed beam contours. The system provides full flexibility to redefine the cell layout and form different sized beams to provide the best quality of service to the users. Nevertheless, a reference case must be defined such that the standard metrics of communication performance may be specified, e.g., G/T and EIRP.

The reference set is used for design budgets and verification tests to quantify the performance of the system and to provide the allocations to the subsystems and elements. However, the architecture easily allows different beams to be formed by simply changing the beam coefficient weights. The only limits being the amount of hardware processing deployed and the satellite antenna optics. Additional hardware processing may be included in the initial roll-out to accommodate expected growth, or added at a later date.

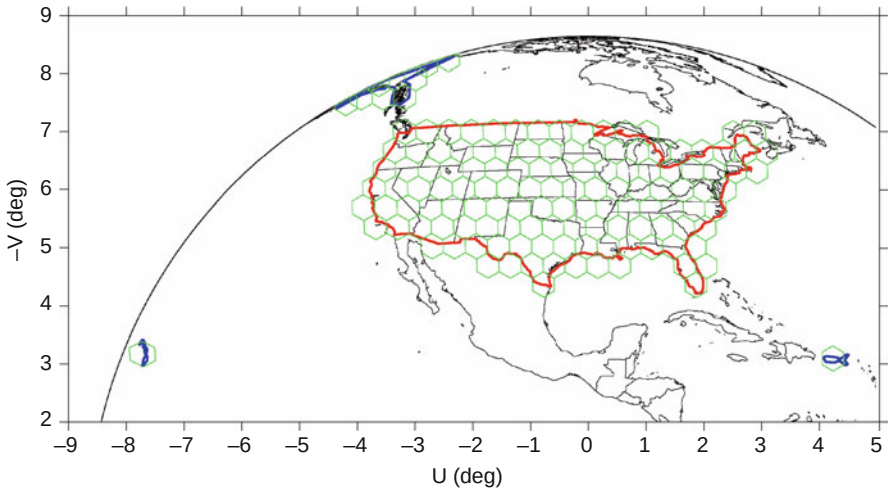


Fig. 19.9 Example of a reference cell layout over the coverage area

GBBF Calibration Scheme

To form a specific beam the amplitude and phase weighting must be set to the appropriate numbers for each element. Further, these weightings must effectively be applied at the feed element aperture. For example, take two elements and assume the desired beam is formed with an amplitude of 1 and a phase of 50° for element 1 and an amplitude of ½ and a phase of 100° for element 2. In typical on-board beam forming, the beam forming operation is very close to the feed aperture, so it is much simpler to set these values correctly. However, with GBBF, the signals traverse down independent payload conversion paths, through the propagation media at different frequencies which may have differing amplitude and phase channels, and then through independent conversion paths in the RFS. With this example, let us say that the element 1 path experiences 3 dB additional attenuation and 50° of additional phase shift as it traverses through the payload, RFS, and propagation path. Without knowledge and compensation of this difference, the beam forming weights at the aperture will not be the desired values but instead element 1 would be equivalent to ½° and 100°, the same as element 2. Depending on the error experienced, the desired beam may be mispointed, misshaped, or even dispersed so grossly as to not be recognizable as a spot beam. However, if the value of the amplitude and phase difference between the element paths between the feed element aperture and the GBBF beam forming operation is known, it may simply be compensated for by adjusting the feed coefficient weights or compensating for the shift before applying the feed weights. Consequently, in order for the GBBF to function as required, a calibration scheme must be implemented to determine and compensate for the amplitude and phase variations between the feed element paths.

This distributed calibration architecture provides many benefits. First, the satellite design consists of traditional bent-pipe transponders. Consequently, the

payload may be designed, built, and verified to traditional specifications without any specific knowledge of GBBF. This approach minimizes the cost, schedule, and complexity of the satellite payload. Second, much flexibility is built into the architecture to allow changes to the calibration signaling and algorithms as needed to solve any unforeseen challenges and to evolve the system as markets and traffic profiles change. Third, the testing of the satellite and GBBF after launch is greatly simplified as the satellite and GBBF operate together in a transparent fashion allowing them to be tested together using traditional satellite IOT methods.

Next, let us discuss the details of the calibration architecture. The calibration scheme contains five processes: feeder link Doppler correction, return feed element path gain and phase imbalance correction, forward feed element path gain and phase imbalance correction, pointing error estimation and correction, and uplink power control.

Feeder Link Doppler Correction

The diagram in Fig. 19.10 is used to describe the Doppler correction scheme. The upper right of the diagram illustrates the satellite-based components of the distributed calibration architecture, whereas, the lower left of the diagram illustrates the ground-based components. The upper left and lower right images show the Ka-band feeder link frequency plan: the uplink and downlink.

The Doppler effect is well known and refers to a frequency shift that occurs between a transmitter and receiver when they are in relative motion. The relationship between the received and transmitted frequency is given by:

$$f_r = (1 - \frac{v_r}{c}) f_{tx}$$

The satellite is launched in an inclined orbit of 6° which results in the Doppler shift varying between plus and minus 3.2 KHz at the Ka-Band feeder link frequencies over the 24-h period. As discussed earlier in order to form the desired beam, the amplitude and phase weights must be set at the S-band feed aperture to the correct values. With a differential Doppler shift of even 1 Hz between the signals in the feed element channels, this would mean that the phase is changing 360°/s. In order to form the beam the differential phase between elements must be kept small, e.g., 5°. Furthermore, in the forward direction, the signals leave the GBBF with their set amplitude and phase coefficients in advance of their arrival at the feed aperture by the propagation delay from the GBBF output to the feed aperture of approximately 250 ms. Thus the Doppler correction must be precise.

The method of Doppler correction is as follows. First, the GBBF includes a very precise low phase noise Master Reference Oscillator (MRO) which is the master frequency reference for the entire system. A continuous wave (CW) signal derived from this reference is sent up to the satellite in the special calibration channels highlighted in red on the uplink frequency plan of Fig. 19.10.

The satellite receives this CW signal at Ka-band and routes it to a Tracking MRO (TMRO). The TMRO follows the Doppler frequency shift of the CW reference signal

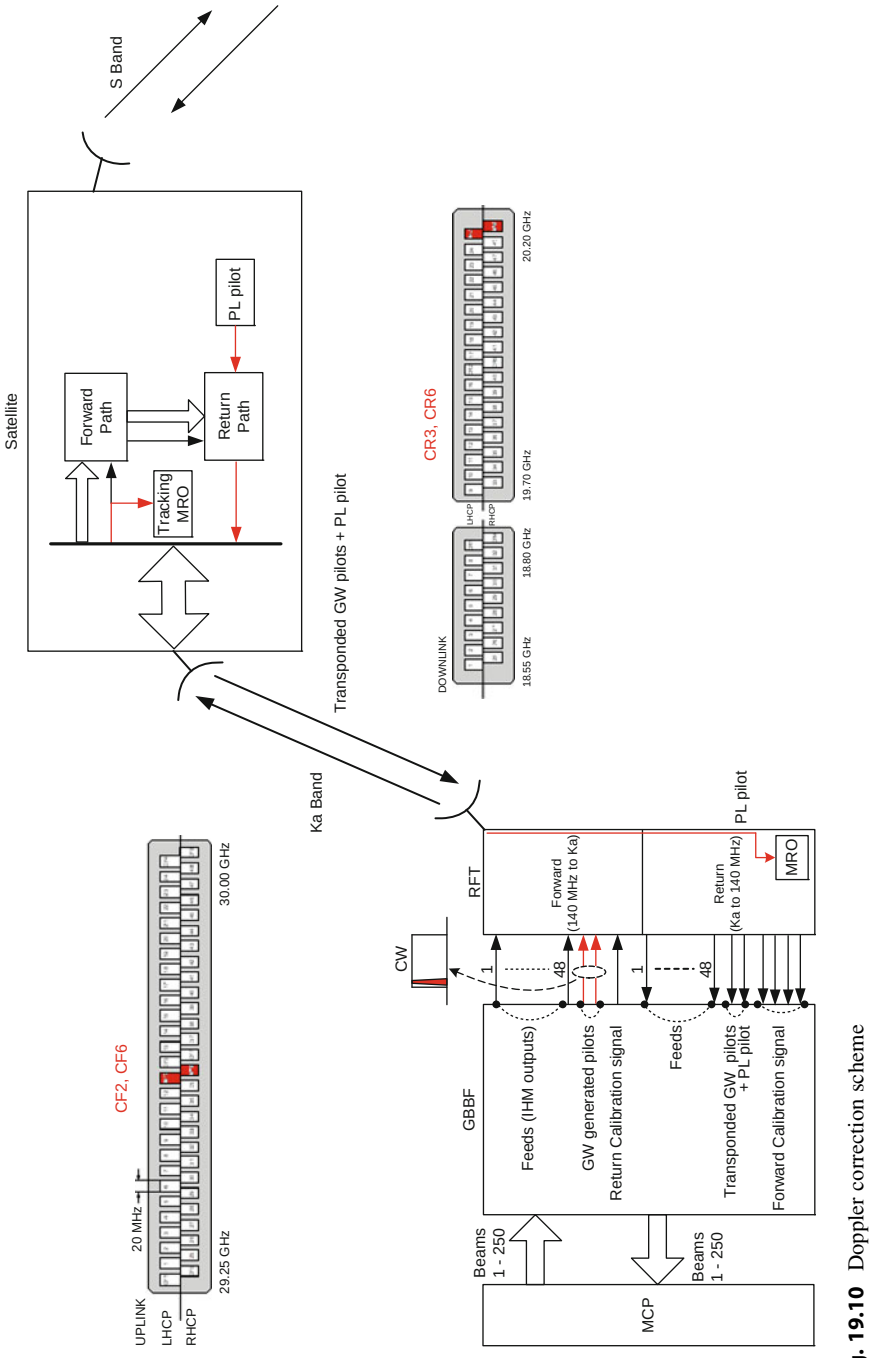


Fig. 19.10 Doppler correction scheme

as the satellite moves through the orbit. The satellite payload uses the TMRO as the master reference for all frequency conversions on the satellite. Thus, as the frequency of the calibration signal moves higher, the local oscillators used to translate the Ka-Band signal to S-Band also move higher. The Doppler effect causes the 20 MHz channels to move higher in frequency and also to be spaced further apart. However, the local oscillators on the satellite move in a likewise fashion such that the signals in each 20 MHz channel arrive at each S-Band feed at the same frequency.

In the return link the process is similar. A payload pilot CW signal is generated on the satellite and sent down to the Gateway RF Terminal subsystem in the special calibration channels highlighted in red in the downlink frequency plan. The RFT also has a TMRO which receives the payload pilot and tracks the feeder downlink Doppler shift. The return frequency conversions in the RFT are locked to the Gateway TMRO, so that all the return channels arrive at the GBBF lined up in frequency.

Return Feed Element Path Gain and Phase Imbalance Correction

The return calibration scheme is illustrated in Fig. 19.11. Notice the two red X's on the diagram. One is at the S-band feed aperture on the satellite in the upper right. The second is at the GBBF feed element input where the RFT translates the transponders to provide the 48 feed element channels.

In order to form the desired beam the amplitude and phase weight for each beam must be set at the feed aperture: the first red X. The S-band signals from users on the ground arrive at each feed element and travel through the return conversion path where they are upconverted to Ka-band and frequency division multiplexed as highlighted in red on the Ka-band downlink frequency plan. The signals from each feed element are then received by the RFT and translated in frequency from Ka-band to 140 MHz and applied to the GBBF. The signal from the feed elements experiences the amplitude and phase shifts that occur along the path from the satellite to the GBBF: the payload, the Ka-band propagation, and the RFT conversions. The calibration scheme's goal is to determine the amplitude and phase for each element path from the aperture, the first red X, to the GBBF beam former, the second red X. Once this value is known for each path, it is easily compensated for in the GBBF.

The value is determined by measuring the amplitude and phase of a calibration signal which traverses the complete path. Specifically, a single calibration signal is coupled into the feed aperture and travels down each return feed element path along with the S-band return traffic. This signal is then received in each of the GBBF feed element inputs. The GBBF processes the calibration signal using a correlation receiver and determines the relative amplitude and phase offset between each of the elements.

An important goal during the calibration architecture design was to have flexibility to solve any unforeseen challenges that might occur during the deployment. This goal, along with the complementary goal of simplifying the satellite and putting complexity on the ground, drove the decision to generate the return

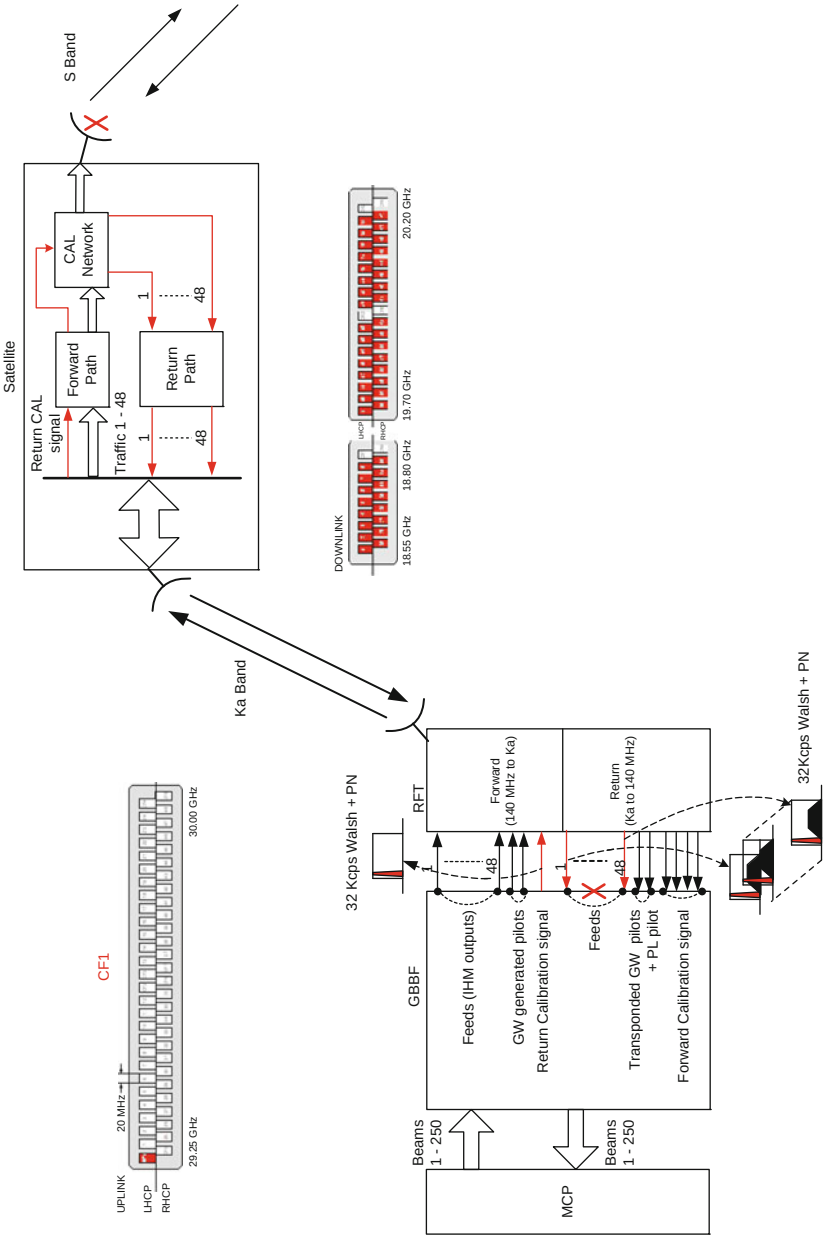


Fig. 19.11 Return feed element path gain and phase imbalance correction scheme

calibration signal on the ground where changes could easily be made. In Fig. 19.11 this is shown as the return calibration signal, 32 Kcps Walsh + PN waveform, which travels up to the satellite on a special calibration channel. The satellite receives the signal, converts it to the S-band receive band, and couples it into the feed aperture of each element.

A calibration network is used on the satellite to couple the calibration signal into each of the feed apertures. Any amplitude and phase variations that the calibration signal experiences on the uplink path before coupling into each feed element will be common between each of the return feed element paths. Therefore the uplink variations will not contribute to the differential amplitude and phase measurements of each feed element path in the GBBF and the desired beam will be formed correctly.

GBBF Ground Equipment

The GBBF Subsystem implementation is shown in Fig. 19.12. The GBBF ground equipment consists of four elements: the Beam forming Element (BFE), Management Element (ME), Diagnostics and Test Element (DTE), and Pointing Beacon Station (PBS) Element.

The BFE is the computation engine of the GBBF Subsystem and performs the high-speed real-time computations. On the return path, the BFE converts and shapes the analog element signals, received from the satellite through the Radio Frequency Subsystem (RFS). The BFE sends the signals for each beam to the MSS Communications Processing (MCP) Subsystem, where the user beam signals are processed and provided to the end user. In the forward path, the BFE processes user beam signals from the MCP and generates the signal data to apply to each feed element. The BFE develops the channel signals from the feed element signal and converts each to an analog Intermediate Frequency (IF). The channel signals are then transmitted to the satellite by the RFS. The BFE includes the generation and processing of the calibration signals as described in the architecture description.

The ME is responsible for the monitoring, control, and management of the GBBF Subsystem and the PBS element. The ME communicates with the Gateway System Controller (GSC) for the purpose of system monitoring and control and with Satellite Control Center (SCC) in order to obtain ephemeris data update.

The DTE performs the role of GBBF Subsystem diagnosis and testing. It has both online and offline diagnostic capabilities. The ME exercises the diagnostic tests and records the outcome.

The PBS Element is comprised of four Pointing Beacon Stations (PBSs) that are positioned in four designated geographic locations within the Continental US (CONUS). The PBSs transmit pointing beacon signals which are received by the satellite in each feed element and passed through to the GBBF along with the return user signals. The BFE forms special monopulse beams at the PBS locations to track and electronically steer the user formed beams to compensate for satellite motion and beam pointing errors.

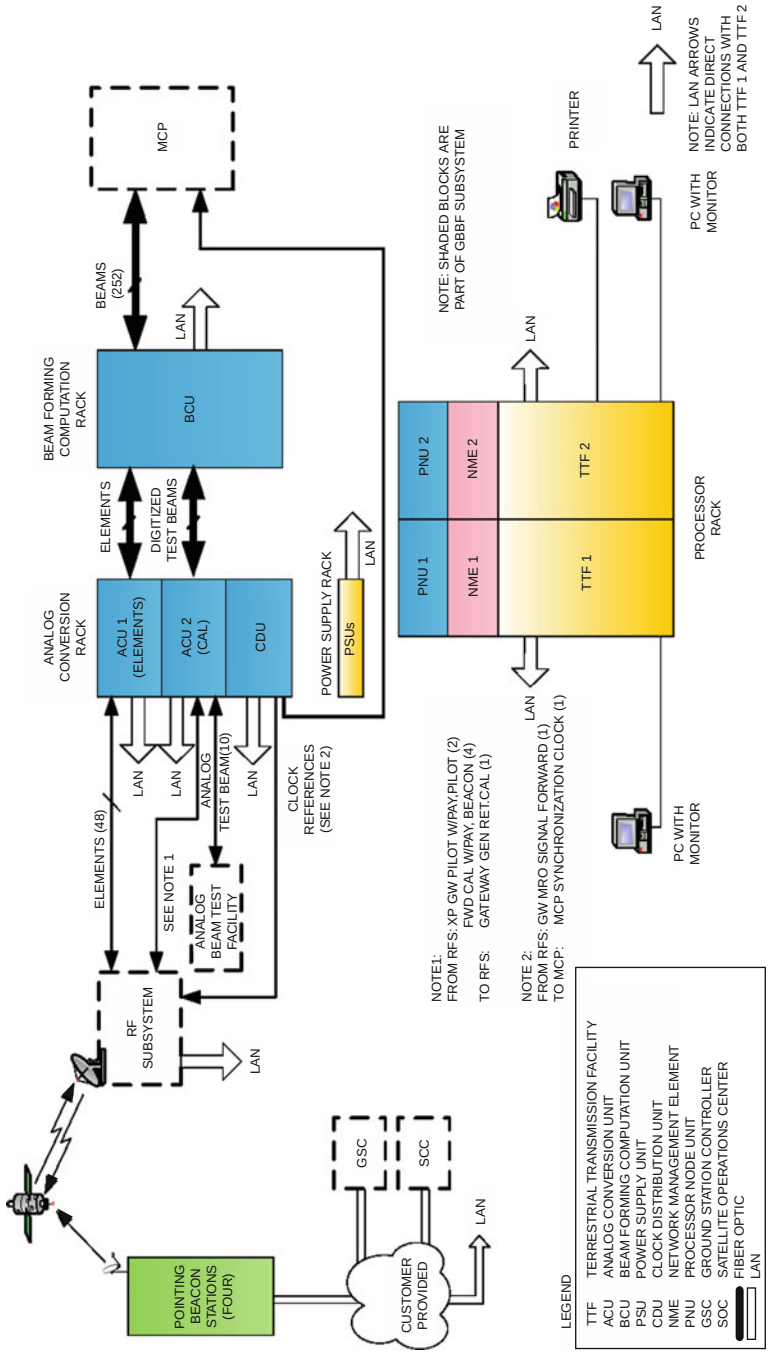


Fig. 19.12 GBBF subsystem implementation

Inclined Operation of MSS Geo Satellites

One of the challenges in system design is balancing the cost and complexity of the satellite including launch with the cost and complexity of the ground segment. An interesting solution in the MSS to this problem is to allow the Geo satellite to drift in north-south inclination. This reduces launch costs and provides more options of launch vehicles by eliminating the need to launch expensive station keeping fuel. Then, the GBBF is employed to compensate for the orbital inclination by reshaping and pointing the spot beams as needed.

This problem arises due to a combination of a need for a large satellite and the need to operate the satellite served user terminals with those using the same frequency band in the ATC. Although the GBBF simplifies the satellite when compared to on-board processing approaches, the satellites are still very large due to the multi-element beam array and the corresponding payload hardware. In order to operate with the ATC and ease the frequency reuse planning, it is important to keep the satellite spot beams aligned with the geographic cells.

When a Geo satellite is used to provide Fixed Satellite Services (FSS), such as the broadband satellites, the orbital position is controlled and held within tight limits referred to as the station keeping box. For example, the box dimensions might be +/− 0.05° horizontally and vertically from the orbital center. This is needed for FSS systems, so the user terminals can have inexpensive fixed pointed dish antennas which only require alignment to point toward the satellite during initial installation.

Since the MSS Systems are designed to communicate with user terminals with near omni-directional antennas, the satellite north-south station keeping can be removed. North–south station keeping is expensive in fuel and is often the limiter on the operational life of a satellite. Without station keeping, the inclination of the satellite orbit increases continually at about 0.8°/year. Thus, for a 15-year operational life, the satellite may be initially placed in a (−6°) inclination, and allowed to drift up to (+6) degrees.

While the satellite drift in inclination does not cause a problem for the mobile user terminal pointing to the satellite, it does give rise to an issue with the spot beams from the satellite to the user. As the satellite moves through the inclination, a fixed spot beam from the satellite will project on the earth with a different size and location. This is illustrated in Fig. 19.13 which shows that in the center of CONUS, the effect is small; however for beams near the edge, the difference is significant.

The GBBF has the ability to compensate for this effect by using different beam forming coefficient weights to adjust for the satellite's position in the inclined orbit. Thus, the spot beams can remain in alignment with the geographic cells and use the same frequency reuse tables.

The combination of eliminating the need to fly fuel for north–south station keeping and using the GBBF flexibility to keep the beam shape and pointing aligned with the cells addresses this challenge.

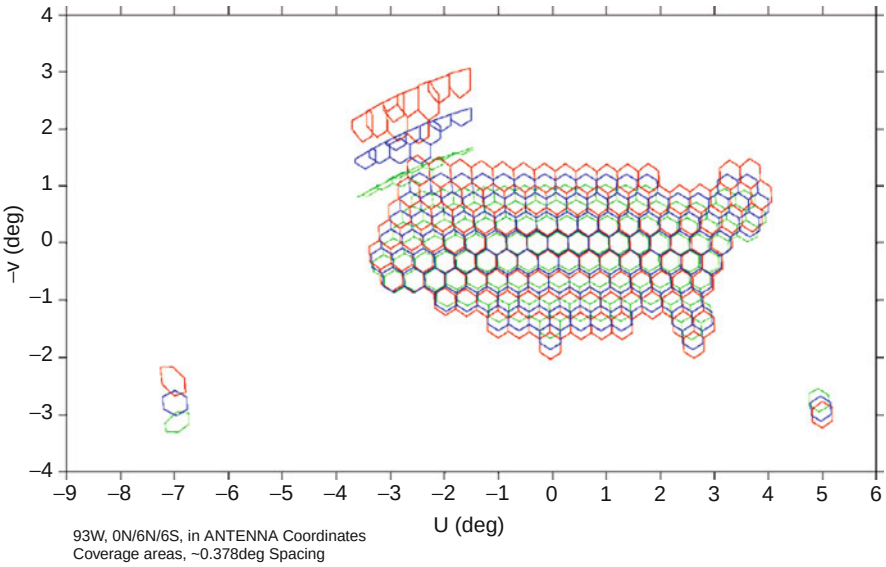


Fig. 19.13 Projection of beam cells at –6, 0, +6 inclinations

MSS User Terminal Links and MIMO

Even with ATC, there is a need for the satellite phone to provide a similar performance and user experience as the terrestrial network for commercial viability. Thus a key challenge is to improve the link performance between the satellite and the satellite phone. In addition to larger antenna apertures on the satellite and adaptive coding and modulation on the links, consideration of advanced satellite phone antenna technologies and diversity is underway.

When Multiple Input/Multiple Output (MIMO) systems were described in the mid-to-late 1990s by Gerard Foschini and others, the astonishing bandwidth efficiency of such techniques seemed to be in violation of the Shannon limit. But, there is actually no violation because the diversity and signal processing employed with MIMO can, in effect, transform a point-to-point single channel into multiple parallel channels as depicted in Fig. 19.14. The gains however come not from an increase in power but from an opportunistic mitigation of channel fading. The required SNR of a fading channel is often more than 20 dB higher than an Additive White Gaussian Noise (AWGN) channel. MIMO designs allow us to operate in a fading environment but reduce the required SNR significantly through introduction of space and time diversity.

Pieces of what we now call a MIMO system have existed for the last 30 years in the form of phased array antennas. Adding additional capabilities to phased array antennas, such as adaptive change of weights, leads to the evolution of smart antennas. The first application of these smart antennas was with digital TV set-top

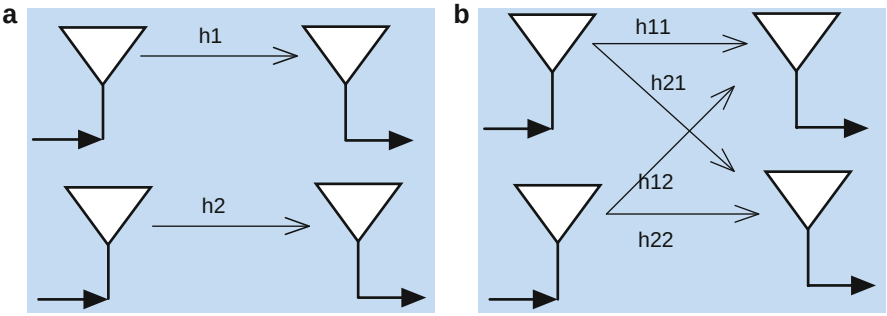


Fig. 19.14 Doubling of capacity by doubling the number of paths, (**a**) SISO system, (**b**) same design that allows reception of the signal from each transmitter by both receivers, making it a MIMO system

boxes where the antenna automatically adjusts array weights to maximize received signal gain. On the transmit side, phased arrays or Multiple Element Antennas (MEA) are used on transmit towers to transmit multiple channels over multiple antennas as well for beam forming. For satellite application, beam forming can be an enabler for forms of MIMO.

In fading channels, the multiple transmit and receive antennas create both Tx and Rx diversity. This is a way to increase data rates, by transmitting data through N different channels. And even if higher data rates are not the goal, the use of different paths provides greater robustness. MIMO systems use *space-time signal processing*, whereby time is complemented with the spatial dimension inherent when using several spatially distributed antennas (at the transmitter and receiver). Such systems can improve bit error rate (BER), or they can increase capacity, or both, without expending any additional power or bandwidth.

For wireless communications, the new paradigm introduced with MIMO is: Instead of combating multipath, we exploit it by transforming a point-to-point channel into multiple parallel channels (a matrix channel), thereby achieving greater capacity (and/or robustness). Special version of trellis coding that operates over the multiple antennas is an integral part of the capacity enhancement usually described as a sum of increased array gain, higher SNR, and multiplexing gain (higher data rates). The capacity calculation of MIMO indicates that adding a small amount of diversity (more antennas) is sufficient to provide large gains, with diminishing returns as more antennas are added. In Fig. 19.15, the performance is seen to approach that of an AWGN channel as the number of antennas keeps increasing.

MIMO systems are employed in channels that experience several types of fading; multipath and Doppler being two main areas. A typical satellite link is a line-of-sight link. However, as many new systems have direct-to-user components, the channels have to cope with Doppler and fading. Due to the large delays, the channel knowledge that is required for MIMO is a challenge to both the transmitter and the receiver in a satellite link. However, it is not an insurmountable problem, and some satellite systems such as DBSD have explored technology to overcome it.

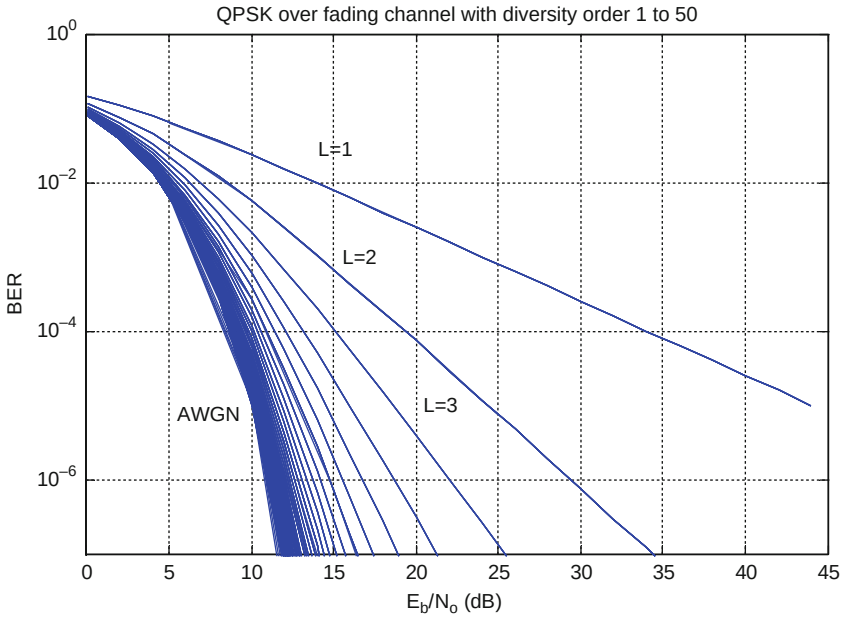


Fig. 19.15 Gains possible as more antennas are added to a link, at BER of 10-2, A SISO channel requires 14 dB, with two antennas, this goes down to 8 dB and with one more antenna, the total gain is (14 – 6.5)–7.5 dB

For a pure line-of-sight (LOS) link, MIMO does not offer a huge benefit, but when a link suffers from fading such as in an urban environment, mountainous terrain with shadowing, and during rain, MIMO can increase link performance by a significant amount. A normal SISO link can be turned into a SIMO link when receiver is equipped with multiple antennas located at just half a wavelength apart. The multipaths created by this environment can then be advantageously combined to reduce the losses usually allocated for these channels. As we see in Fig. 19.15, using three antennas instead of one can reduce the required Eb/N0, under a fading condition, by approximately 8 dB, enough to make up for rain loss at Ku-band. Future work in developing MIMO receivers for satellite applications is expected to help address the discussed challenges in MSS systems.

Broadband Satellite Systems and Internet Access

Broadband Satellite Systems are used to provide Internet access to users via a satellite terminal. The physical integration of the satellite system with the terrestrial network occurs at the gateways. The satellite user terminal side, commonly called Customer Premise Equipment (CPE), communicates via two-way links through the satellite to the Gateway which is connected to the Internet.

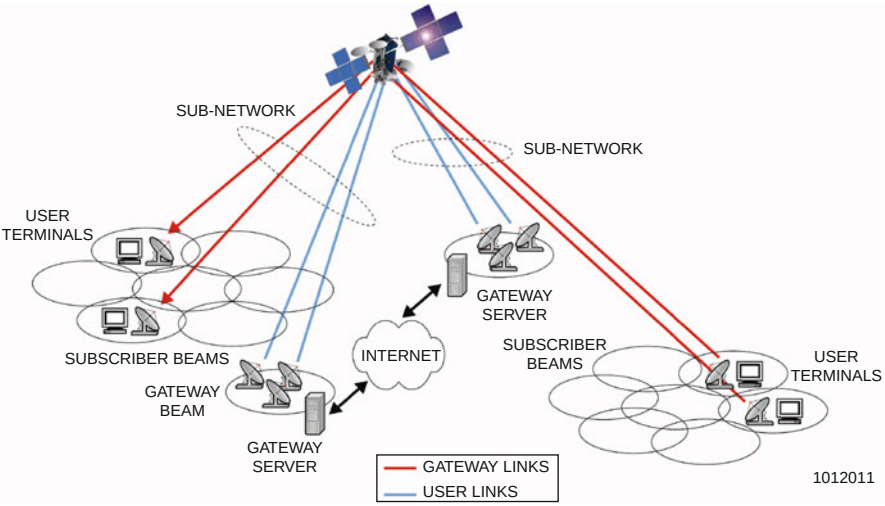


Fig. 19.16 Broadband satellite for internet access

Since the user terminals only communicate with the satellite, the MSS system challenge of dual handsets and frequency sharing with terrestrial cell towers using the same frequencies is avoided.

Nevertheless, significant challenges remain. Some were encountered and addressed during the development of the current systems. More are being addressed to enhance the existing systems and in the development of the next generation systems. These challenges stem primarily from the need to provide each user of the system an acceptable and desirable experience, while providing the same level of performance to enough user customers for the system to be a commercial business success. The challenges can be categorized into three areas:

1. Bandwidth and Capacity: Individual User data rates and total capacity of the system
2. Latency over the Satellite links and the users' Internet experience
3. Interfaces and Standards

Figure 19.16 illustrates a typical Broadband Satellite Internet Access System. These systems consist of Gateways, Satellites, and User Terminals. Multiple user spot beams are provided to maximize the total capacity of the system through frequency reuse. The architecture presented in Fig. 19.16 is referred to as a hub spoke architecture. Each user terminal in a spot beam corresponds to spokes connected to a central gateway hub. The Gateways are typically interconnected via the Internet in a private network as well as serving to access the public Internet for the users. The user links and Gateway links consist of forward links from the gateway through the satellite to the user terminal, and return links from the user terminals through the satellite and to the gateway.

In addressing challenge 1, broadband satellites must take maximum advantage of their frequency assignments to maximize the total capacity. This is accomplished by using polarization diversity and frequency reuse. As the user beam count is

increased, the need for multiple gateways arises as each user signal must pass through a gateway in the hub spoke architecture. Other architectures provide for user beam to user beam connectivity as will be discussed later when considering an Internet router in space.

The frequency bands used for broadband access are currently C, Ku, and Ka. While C and Ku are typically limited to 500 MHz of spectrum, Ka can provide 1,000–1,500 MHz of licensed bandwidth. The industry is also looking forward to V-band for increased capacity.

In 1998, the FCC took another step to address the increasing needs for satellite bandwidth and to ensure technological development with preliminary allocations at Q-/V-band. The current allocations provide at least 4 GHz of bandwidth which is considerably greater than the satellite systems in use today. RF components for both ground and space are already in development, and its likely broadband access satellite systems will make use of this increased capacity in the next 5–10 years.

The total capacity of the satellite is limited by the number of beams, the frequency reuse scheme, the bandwidth available to each beam, and the information data rate of the selected modulation and coding for each beam. Simply, 60 spot beams with 500 MHz of bandwidth and 2 bits/Hz signaling provides 60 Gbps of available capacity. The total capacity of a given satellite network is typically defined by adding the forward and return capacities, with the current systems targeting hundreds of Gbps.

Nevertheless, the maximum capacity of a satellite Internet access system is much less than that provided by terrestrial fiber infrastructure due to the RF bandwidth constraints. Commercial business success for the satellite system, then, must carefully assess the infrastructure build-out plans and market needs for services and applications.

O3b Networks Limited founder, Greg Wyler, is not trying to compete against fiber infrastructure. Instead, he has defined the company mission to make the Internet accessible and affordable to those who remain cut off from the information highway (3B Networks). O3b is short for the Other Three Billion, referring to the nearly half of the world's population that has little or no access to the World Wide Web. Many investors are on board, including: SES S.A., Google Inc, HSBC Principal Investment, Liberty Global Inc, and North Bridge Venture Partners.

O3b considers their mission a social responsibility. "Every member of the O3b team is passionate about bringing affordable, state-of-the-art broadband services to the three billion people who have been denied them for reasons of geography, political instability and economics" (O3B Web).

O3b Networks plans to launch a constellation of satellites in an orbit only 8,000 km from earth. This is four times closer than geosynchronous satellites. Consequently, the O3b satellites will continuously move relative to the user as they circle the Earth. The Internet traffic will be provided by the satellite over the region. Before the satellite passes out of range, the traffic will be routed over the next satellite coming through.

The MEO orbit and architecture reduces the round trip latency of a geosynchronous orbit from over 500 ms to approximately 100 ms. O3b believes this is important and will allow their system to provide the user with a Web experience significantly closer to terrestrial systems such as DSL or Optical Fiber. The satellites are also equipped with steerable antennas and support a reconfigurable RF payload to adapt to changing needs and market evolution.

Nevertheless, they recognize the demand for connectivity goes beyond the bandwidth they can provide with the satellite system. Consequently, their plan includes partnering with PC and device manufacturers to eliminate bottlenecks in the system to provide the most capacity possible.

Hughesnet, WildBlue, and Viasat are continuing with their Geo Satellite orbit approach to providing Internet access systems. These systems eliminate the need to develop efficient satellite hand-off methods, and the developers concentrate on improving the network efficiency with advances in modulation and coding, and reduced overhead and latency in the network protocols.

Protocols and Network Performance

A significant challenge in the integration of satellites and terrestrial networks involves the development and acceptance of the ways digital messages are formatted and the rules governing how these messages are interchanged – in short, the protocols. Protocols describe the syntax, semantics, and synchronization of communications between the wide range of entities that are part of an integrated satellite-terrestrial network. Protocol developers must ensure effective communication performance and guarantee that entities that develop products that follow the protocol standards are able to communicate.

This section will present some general development frameworks in which protocols have been and are being developed. These frameworks, called Protocol Reference Models, map a process by which protocols can be developed and products implementing these protocols can be commercialized.

Some of the protocols currently in use in both terrestrial and satellite-terrestrial integrated systems will be discussed along with the challenges in protocol development to improve network performance in integrated satellite-terrestrial architectures.

Reference Models

A protocol reference model provides a generally agreed-upon framework in which communication protocol development can take place. To do this a model categorizes the role for protocols in terms of functionalities in the overall execution of the end-to-end communication of message data.

The types of reference models discussed here are called layered reference models. Figure 19.17 illustrates the general structure of a layered reference model, the International Standards Organization (ISO) Seven Layer Reference Model (RM)

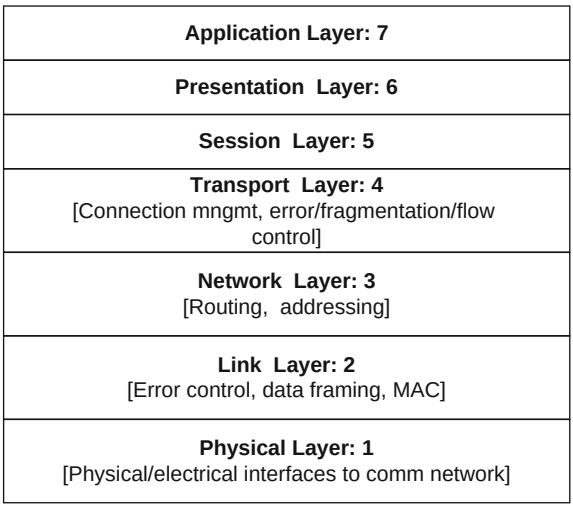


Fig. 19.17 ISO reference model

for Open Systems Interconnect (OSI) (Gibson 1997, page 570). Each layer of the RM defines a set of well-defined functions in the context of overall communication. The layer operates by sending messages to the corresponding layer (called the peer layer) of the remote entities with which it is communicating. Each layer also provides services to the layer above and uses the services of the layer below. Peer layer messages can consist of both user and control data.

Figure 19.18 shows the total communication process. Note that conceptually, each layer communicates with a similar peer layer; however, in practice, the resulting protocol message units of a particular layer are passed by means of the services of the lower layer. As illustrated in the figure, node entity A sends a message to node entity B, with all bits flowing over the physical link. The digital message from node A employs the protocol specified services of the descending layers (7 thru 1) until generating message bits with the peer layer protocol message unit overhead included from each layer. Through this process, protocols at each layer encapsulate the inputs from the layers above. Data received at entity B is processed in reverse. Each peer layer message is extracted from the encapsulation and the resultant is passed to the next higher layer until the message is received by the final destination application at Layer 7.

Two examples illustrate some of the different ways the peer layer messaging can work. For example, consider layer functions that perform data translation type of operations, like data compression, data encryption, and forward error correction (FEC) encoding, the transformed data may be simply messaged to their corresponding remote receiving peer layers without any additional control or instruction exchange. Upon receipt by the remote peer layer, the message data are decoded per the definition of that layers function, e.g., decompression.

At another extreme are the many variations of the Transport layer protocol called Transport Control Protocol (TCP), which is discussed in more detail below. The TCP protocol sends a variety of messages to the remote receiving layer including

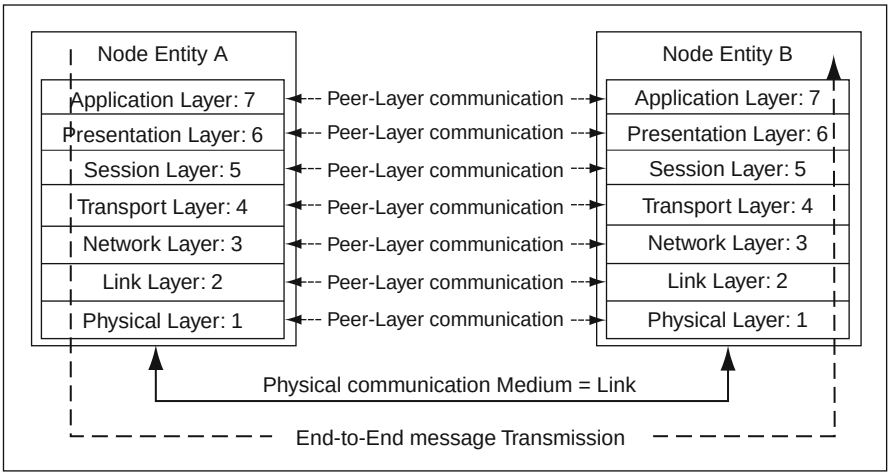


Fig. 19.18 ISO reference model communications process

set up signaling messages, return receipt acknowledgments, and segments of the entity-to-entity digital message that is being transmitted.

Actual protocols that conform to a RM must be defined by specification standards following guidelines of the RM. There are two types of standards for each layer: the protocol standards, defining how peer entities at this level communicate (the rules, conventions, and message formats for peer layer interactions) and the service standards with interface specifications for the services that are provided to the higher layer immediately above. In order of an RM-conforming set of protocols to support community of interacting communication entities, it is necessary to have at least one protocol specified at each layer of the RM. The resulting set of protocols is called a *protocol profile*, a name that is often prefixed by the name of the RM (e.g., a OSI protocol profile). Each level of a protocol profile can have several protocols, each providing a different service to the higher layer.

There are several communication protocol RMs in use that have relevance to the challenges related to the integration of satellites into terrestrial networks. The main three RMs that are discussed below are the ISO seven layer Protocol RM (also known as the OSI RM), the TCP/IP Protocol RM (also called the Internet Protocol RM or Suite) ((Gibson 1997, page 575) and (Stevens 1994), and the Broadband Satellite Multimedia (BSM) RM (architecture) (ETSI TR 101 984 V1.2.1 (2007)).

Role of Reference Models

A principal factor in the successful integration of satellites with terrestrial networks is the development and wide range acceptance of protocols that not only enable an effective and transparent (seamless) working system combining these two elements,

but also enable this relationship to continually evolve as these two elements advance technologically. The reference models and the development methodology they require are essential for the ongoing process.

The RM methodology provides for the development of layered protocol profiles where each layer's functionality is well-defined and its interfaces with the two adjacent layers rigorously specified. This open systems approach allows for the independent development and implementation of proposals independently for those of the other layers and enables the interchangeability of protocols meeting the layer's specifications. Thus continued innovation is made possible as is the wide range acceptance of products that implement these innovations. Potential producers can focus on manageable parts of the overall system with reasonable assurance that products resulting from this focused work will be compatible with existing implementations and thus, if superior, have the potential of gaining wide range support. And, of course, wide range acceptance both makes possible and is made possible by increased affordability of protocol implementations.

ISO 7-Layer Reference Model for Open System Interconnect

The ISO 7-Layer RM is illustrated in Fig. 19.17. The development and issuing of standards for protocols complying with the ISO RM are the work of the International Standards Organization. The ISO RM is also widely called the OSI RM.

The ISO RM yields protocol profiles that are to be implemented in communication networked entities that are linked by some physical medium (radio links, fiber optic links, cable links, IR links, and multiple networks of such links). The highest ISO layer, Layer 7, is closest to the user interface (or application programs); the lowest ISO layer, Layer 1, provides the connection to the physical link. Below is a brief description of the general functionality of each of the ISO layers (Note: For the purposes of satellite/terrestrial network integration, the focus is on the lower four ISO layers):

Layer 1 – Physical Layer: defines the means of transmitting/receiving raw *bits* rather than logical data packets over a physical transmission medium connecting network nodes – the bit stream may be grouped into symbols that are converted to signals for transmission (and conversely, for reception); provides an electrical, mechanical, and procedural interface to the transmission medium, including (as appropriate to the medium) the specifications for:the shapes and properties of the electrical connectors, the frequencies, the modulation scheme to use, and similar low-level parameters.

Link Layer 2: provides the functional and procedural means to transfer data, in *frames*, between peer Link Layers, using hardware addresses and focusing on local delivery, addressing, and media arbitration; and may also provide for the detection and correction of physical layer errors. In the computing environment, the Link Layer is often implemented in software as a "network card driver."

Often the Link Layer is subdivided into the two following sublayers:

Link Control Sublayer (data multiplexing/decoding, error control – rebroadcast of damaged frames, acknowledgments, error notifications, flow control – what to do when a frame is received)

Medium Access Control (MAC) Sublayer (multiple access control, framing. synchronization, data queuing/scheduling, hardware addressing, QoS control, level 2 switching)

Network Layer 3: This layer provides the functional and procedural means of transferring variable length data sequences (called *packets*) from a source to a destination host via one or more networks while maintaining the quality of service functions; includes addressing, packet queuing, and packet forwarding.

Transport Layer 4: This layer provides transparent transfer of data between end users, providing reliable data transfer services to the upper layers. The Transport Layer controls the reliability of a given link through flow control (of *segments*), segmentation/desegmentation, and error control. There are a range of connection and reliability service options available to the upper layer.

OSI Layers 5–7 – Collectively, Application Layers (Session Layer, Presentation Layer, Application Layer): Collectively (and almost universally) referred to as the Application Layers, – for complete descriptions of OSI layers 5–7, refer to (Gibson 1997)): establishes, manages, and terminates the connections between the local and remote application, manages the quality of the connection (reliability, security, full-duplex, half-duplex, or simplex operation), and connection setup – checkpointing, adjournment, termination, and restart procedures. These layers contain the functionality that supports user software applications.

Note the vocabulary lesson presented in these above layer descriptions relative to the naming of the *protocol data units* that are used for peer layer-to-peer layer communications: at Physical Layer, *bits*; at Link Layer, *frames*; at Network Layer, *packets*; and, at Transport Layer, *segments*.

TCP/IP Protocol Suite

Figure 19.19 illustrates the TCP/IP or Internet Protocol Suite or Reference Model. The development of protocol and standards based on this RM is under the aegis of the Internet Engineering Task Force (IETF), which is now a standards organization for the protocols for use on the Internet. The TCP/IP Protocol Suite is not quite the top/down comprehensive design reference for networks as is the ISO RM. It was formulated for the purpose of illustrating the logical groups and scopes of functions needed in the design of the suite of Internet working protocols of TCP/IP, needed for the operation of the Internet. In general, direct or strict comparisons of the OSI and TCP/IP models should be avoided. TCP/IP Protocol Suite does not have specific protocols for each ISO RM layer; however, loose comparisons are often made between the two. The Internet Protocol Suite is a layered model widely in use for computing systems networked on the Internet and the IETF has established standard development procedures (Brandner 1996), which lead to open system, vendor-independent standards based on the TCP/IP Protocol Suite.

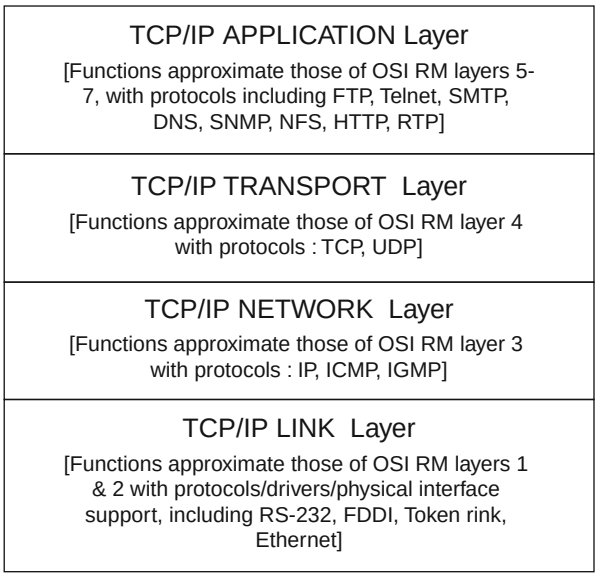


Fig. 19.19 TCP/IP protocol suite

Figure 19.19 provides the names of each layer of the TCP/IP RM and provides a general description of each layer's functionality by reference to the ISO RM (again, these are not strict comparisons). More details of the layer functionality of the Internet RM are found at (CH-CRC 1997, page 575).

Broadband Satellite Multimedia Protocol Architecture

Figure 19.20 introduces a RM based on the needs of the integration of satellites with terrestrial networks, the BSM Protocol RM (ETSI TR 101 984 V1.2.1 2007). The BSM RM is developed and maintained by the third international standards group, the European Telecommunications Standards Institute (ETSI). ETSI has issued a number of standards applicable to the integration of satellite links into terrestrial networks. The BSM RM is a layered model that has been partitioned into a satellite-dependent set of layers (the lowest two layers) and the satellite-independent layers. This RM has an explicitly defined interface, the Satellite-Independent Service Access Point (SI-SAP), between the satellite-dependent and satellite-independent layers. The intention is to focus the ETSI's satellite link work on the specification of protocols for the lowest 2, satellite-dependent layers and to work on the standardization of the SI-SAP. Thus, once the Network layer has been enhanced to interface with the satellite-dependent layers (the Satellite Independent Adaptation Functions), layer 3 and above development can proceed independently of satellite concerns. In practice, this approach has some difficulties that are discussed in the section below, TCP Challenges.

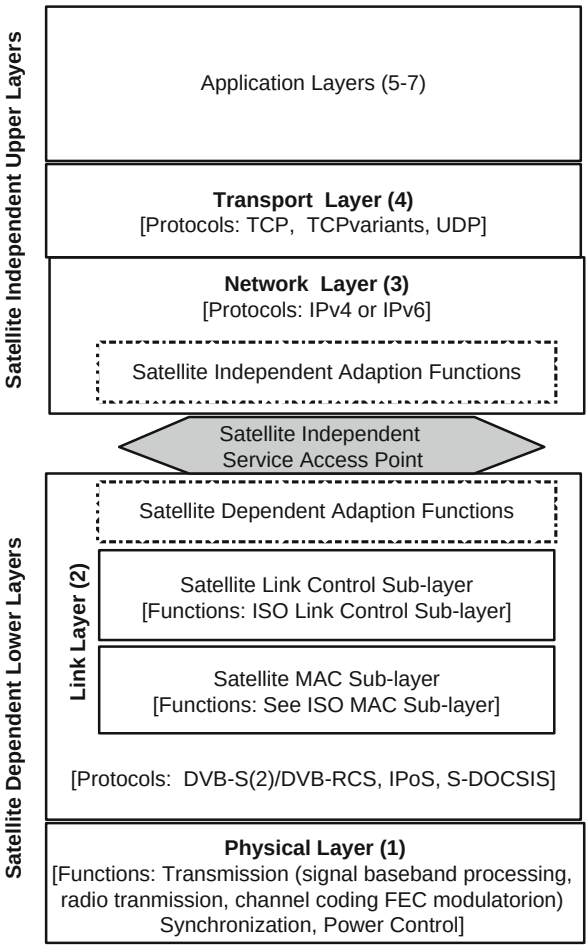


Fig. 19.20 ETSI-BSM protocol architecture

BSM RM in its lower two layers can be compared rather strictly to the corresponding layers of ISO RM (again loosely, as in the case of the TCP/IP RM). A number of protocol profiles conforming to the BSM RM at these lower layers have been developed, specified, and are in use in satellite links carrying Internet traffic. Three will be mentioned here:

- DVB Standards: Digital Video Broadcast over Satellite (DVB-S2) (ETSI-DVB-S2 2009), for the forward link and Digital Video Broadcast-Return Channel over Satellite (DVB-RCS) (ETSI DVB-RCS 2009), for the return link. These standards have as their base video broadcasting and are ETSI's principal standards for BSM RM layers 1 and 2.
- IP over Satellite (IPoS) (ETSI TS 102 354 V1.2.1 (2006)) – mostly a pointer to TIA Document TIA-1008-A, and (Hughes Network Systems 2007).

These protocols are used by Hughes Network Systems in its offering of Internet access via satellite.

- Satellite Data Cable Service Interface Specifications (S-DOCSIS) – the satellite version of DOCSIS (2010) and (DOCSIS 3.0 2010), the Physical and Link Layer protocols that are used for the widely service for Internet accesses via cable modem.

The Satellite Independent layers of the BSM RM above layer 2 can be compared largely with the TCP-IP RM, where the Network Layer (3) uses Internet Protocol version 4 (IPv4) or version 6 (IPv6). And at the Transport Layer (4), User Datagram Protocol (UDP) and a version of TCP are used. In the following section, it is shown that challenges arise to keeping the development of BSM-layer 4 protocol independent from the implementation of the lowest two DSM layers, the Satellite-Dependent layers.

TCP Challenges

One of the ongoing challenges facing the integration of satellites and terrestrial network concerns the protocol implementation of the Transport Layer protocol, TCP (Stevens 1994). The importance of this protocol layer is emphasized by the amount of traffic currently on the Internet directly related to TCP. Analysis of the traffic on the Internet backbone between 2006 and 2009 has shown more than 80% of the bytes and the IP packets at the Network layer level carrying TCP traffic (Dusi and John 2010). Thus, satellite links and systems cannot be well integrated with Internet and other terrestrial system unless traffic resulting from TCP protocols can be effectively and transparently transported by the satellite link.

Basic TCP

The original TCP protocol was specified in September 1981 (Transmission Control Protocol 1981) to provide a reliable transmission of data over the Internet, where "reliable" means that the service provides a guarantee that all the user data will be received without error. To accomplish this, TCP must establish a dialogue with the receiving peer layer and then retransmit segments of data that have been corrupted or have been lost. This process must continue until the receiving TCP has accurately received all sent data. The mechanism to determine if a data segment was successfully received by the receiving TCP is for the receiving entity to acknowledge with a message, an ACK, that tells the sender that the segment was successfully received.

Thus, the TCP at the sending node must retain a copy of a transmitted segment until an ACK for this segment is received. Consequently, the TCP of the sending node must buffer all sent segments that remain unacknowledged. The size of this buffer is called the congestion window. The receiving TCP must also buffer received segments that are missing at least one segment that should have already been received. TCP partitions the total data to be transmitted into an ordered set of segments, which correspond to the order originally sent. The receiving TCP passes

the full, accurately received data message to the higher layer; it cannot send up the data of segments out of order.

Segments may be missing for several reasons. They may have been lost to some collision or due to queue overflow at some intermediate router – both of these losses are considered to be traffic congestion related. There may also have been loss of a segment due to an uncorrectable data corruption due to a bit error in the network links. Another network phenomenon that may give the appearance of a segment loss is the reordering of segment arrivals from the original transmission order; this is due to the possibility that the transmitted segments may follow different network routes to the receiver, each of different transit times and of different congestion levels.

When the receiving TCP experiences an out-of-order segment arrival (when a segment arrives before one of its predecessors), it cannot immediately know whether this is because of a segment loss or due to a delayed arrival of a prior segment. The arrival of an out-of-order segment is ACK'ed (called a *duplicate ACK*) not with the segment number of the arrived segment, but with the segment number of the last segment received in order (all preceding segments already having arrived). Of course, the sender upon receiving an ACK or a duplicate ACK knows that all segments preceding the one whose number is in the ACK have been successfully received and, with the receipt of a duplicate ACK, that the next segment has not yet been received.

The simple description of the original and basic elements of the TCP protocol algorithm is:

- *Initiation*: There is a starting exchange of messages to initiate the TCP session, initiated by the sender.
- *Slow Start*: The sending TCP starts with one segment and increase the *rate* of segment sending based on the number of (success) ACKs received (until a parameterized threshold is reached, at which point the rate increases linearly until a limiting rate is reached).
- *Retransmission*: If an ACK for a segment has not been received within a given time frame based on a computation of latency for a round-trip, a *timeout* occurs and the segment is retransmitted.
- *Recovery*:
  - If a timeout occurs, resume retransmission with slow start.
  - If a duplicate ACK is received, halve the current transmission rate.
- *Termination*: The TCP session ends with a message handshake, initiated by sender.

A full description of the basic TCP protocol is given in (Stevens 1994).

Network Environment and TCP Optimization

The original TCP provided a basic mechanism that was discovered to not be very efficient in terms of data throughput. Its effectiveness is very dependent on the network environment of the TCP session. Among the important factors affecting TCP efficiency include:

- Network latency (delay time to receive the ACK, the round-trip-time, (RTT))

- Size and symmetry of the bandwidth available to the sender/receiver
- Rate and the causes of segment loss
- Topology of the network (single path or multiple path possible for segment transmission –affects the variability and ordering of segment arrival)
- Network traffic conditions in which the TCP-based communication takes place

From the basic form of TCP originated in 1981, a number of varieties have been and continue to be developed. Each variety is usually developed to enhance the effectiveness of the protocol in a specific range of network environments: The performance of each is optimized for this specific environment. At least three major environmental network types can be distinguished:

- Terrestrial Internet – characterized by small latency (short RTT) links, symmetric links, segment loss due to congestion not to corruption (small BERs) and a topology allowing for multiple routes; the TCPs for Internet are discussed in the following section
- Satellite links – characterized by large latency (long RTT), generally single links (to/from), often unequal bandwidths for user in the forward and return directions, links with high BERs, and often single link communications
- Long, very high-speed/capacity links – characterized by large latency (large RTT), symmetric with high bandwidth links which are of low-loss (low congestion and BER) (it is noted that these "long, fast" links are of growing importance for Internet and there are several TCP variants developed to optimize performance in networks with these type of links: TCP CUBIC, TCP HighSpeed, HTCP (Hamilton TCP), TCP Illinois, and TCP Scalable – for all these TCP variants, refer to (Caini and Firrincieli 2009))

Given the complexity of the TCP algorithm, it has not been feasible to provide analytic verification of the superiority of one TCP protocol over another. Judgments of this type are normally based on end-to-end network simulation results where a valid representation of the relevant factors modeled.

In fact, modeling, simulation, and emulation are extremely important tools in all phases of a networked satellite-terrestrial system: architecture, design, deployment, and operation. Due to the complexities of the interaction between the protocol layers, including the resource allocation algorithms, these tools are required if the system developer wants to understand the true achievable network performance, efficiency, and capacity with heavily loaded traffic. The tools may also be used to test and beta-license new applications without consuming the actual satellite network resources (Lindberg-Walker).

Classical TCP

The basic TCP protocol was soon found to be very efficient in terms of segment throughput. This realization led to a sequence of improvements for the Internet environment that culminated in the early 1990s in what is called here the Classical TCP, which is best represented by a TCP protocol called New Reno. The sequence of protocols leading to New Reno were: TCP Tahoe, which speeded up the

retransmission process, an improvement called Rapid Retransmit (Stevens 1994); TCP Reno, which speeded up the recovery process in a process called Fast Recovery (Stevens 1994) and (Allmana and Paxson 1999); and finally, TCP New Reno, which improves on TCP Reno's Fast Recovery (and still called Fast Recovery) (Floyd and Henderson 1999). New Reno is probably the most widely adopted TCP variant (Caini and Firrincieli 2006a); it has been one of the TCP available in the Linux kernel, where it is called Reno (Caini and Firrincieli 2009).

Satellite Link TCP

During the 1990s, as satellite links use was growing, it became evident that the classical TCP variants would not function well in this new environment. The satellite link environment is characterized as: High latency (with RTTs often more than 550 ms), high segment loss due to channel error and not congestion, usually a single pipe topology (between the Gateway (GW) and the Satellite Terminal (ST)), and often asymmetric in that the forward link (GW to ST) is allocated more bandwidth for user traffic than the return link (ST to GW).

By the end of the 1990s, numerous possible solutions had been proposed (Allman et al. 1999). Many of these proposed approaches were developed and simulated over the subsequent decade and this development process is continuing ((Joing et al. 2009), (Roseti et al. 2010). Below is a list of some of the recent or often cited TCP variants or options that have been developed to improve TCP performance on satellite links (Note those TCP variants that accelerate NewReno's fast recovery rely on the usually valid assumption that in satellite links, segment loss is not an indication of congestion but to channel error – so there is no reason to slow down transmission to avoid congestion):

- SACK (Mathis et al. 1996), an option that can be used with other TCP protocols, allows TCP receivers to inform TCP senders exactly which packets have arrived. SACKs allow TCP to recover more quickly from lost segments, as well as avoid needless retransmissions. SACK is a TCP option available in recent Linux kernels (Caini and Firrincieli 2009).
- *TCP Hybla* (Caini and Firrincieli 2009), a TCP protocol that augments the classical TCP by the use of the SACK option and by increasing the New Reno's fast recovery rate by making the rate of increase proportional to RTT (7-Caini, C; Firrincieli, R. – Hybla, West). TCP Hybla is a TCP variant found in the Linux kernel.
- *TCP Westwood* (Caini and Firrincieli 2009) is a TCP variant that augments the classical TCP by increasing New Reno's fast recovery rate by, instead of halving the segment transmission on segment loss detection, setting it to the available bandwidth, thus avoiding harsh slowdown (Westwood's fast recovery algorithm is called *Faster Recovery*). This TCP variant is found in the Linux kernel.
- *TCP Peach* (Akyidldiz et al. 2001) (Joing et al. 2009) (Luglio et al. 2004) is a TCP variant that used dummy segment transmission to access available

bandwidth. TCP Peach speeds up the classical slow start algorithm, calling it *sudden start*, and speeds up the classical fast recovery algorithm, calling it *rapid recovery*.

- To improve TCP over satellite links many other transport layer protocols have been proposed, including recently, TP-Satellite (Joing et al. 2009); SCPS-TP (CCSDS 2006) introduced as part of CCSDS's Space Communications Protocol Specifications protocol suite; and TCP Noordwijk (Roseti et al. 2010).

TCP Performance-Enhancing Proxies (T-PEP)

As mentioned previously, a very major portion of Internet traffic is TCP traffic and the satellite network must integrate with the terrestrial network and handle TCP efficiently for commercial viability. TCP, however, is a connection-oriented protocol, and defines the protocol dialogue between the Transport layer peers at the two end points of the communication, whereas end-to-end communications in general traverse the links of several different network environments (classical Internet links, satellite links, long-fast links, and, perhaps mobile links). Different TCPs have been developed to optimize the data throughput performance of each of these link types separately. But, how can the overall end-to-end performance of TCP sessions be optimized?

One solution is to divide the end-to-end TCP session into several shorter ones, each of which crosses a single link type where each dedicated TCP session can be optimized. An approach for implementing this is through the use of Performance-Enhancing Proxies (PEPs) which when applied to the TCP layer are called TCP PEPs (T-PEPs).

Figure 19.21 provides an example of the use of T-PEPs and is based on an ETSI standard for PEPs (ETSI EN 302 307 V1.2.1, DVB-S2 2009). The protocol architecture includes the Internet (TCP/IP) Protocol Suite and the protocol profile based on the BSM RM. The general model is that of a satellite access system for Internet.

In Fig. 19.21, an Internet-connected node on the Internet is delivering data content to a Satellite-connected user; this needs to be via a reliable communication service. Without TCP session splitting, there would be one end-to-end TCP session; however, in this BSM PEP architecture (38-ETSI PEP), the data content delivery is via three TCP sessions, each occurring in a different network environment:

- TCP Session A (Internet node TCP to the GW PEP Standard TCP): This TCP session is in the Internet environment.
- PEP-to-PEP TCP Session (GW PEP Satellite TCP to the ST PEP Satellite TCP): This TCP session is principally in a satellite link environment (with perhaps two Ethernet LAN at each end).
- TCP Session B (ST Standard TCP to the Sat-connected node TCP): This TCP session is via Ethernet LAN or perhaps some private local terrestrial Internet work.

In this example, TCP sessions A and B would most likely be a classical TCP variety such as TCP NewReno. However, the PEP-to-PEP TCP session would be

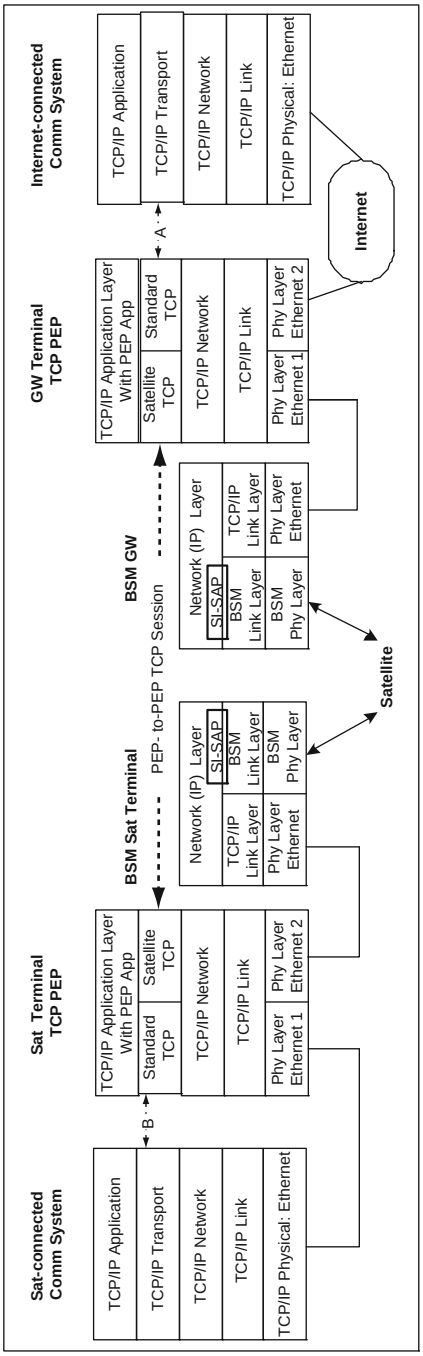


Fig. 19.21 Distributed TCP-PEP architecture for BSM internet access BSM satellite communication system

a TCP variant that has been optimized to the satellite link – and this would depend on the network environment of this satellite link (GEO/LEO for RTT, bandwidth size and symmetry, link BER, satellite link topology – one or a network; this would also depend on the protocols in use in the Link Layer). Note that in both the GW T-PEP and the ST T-PEP, there would be PEP applications that receive and direct TCP traffic from/to the appropriate TCP protocol.

Figure 19.21 illustrates one type of satellite T-PEP architecture, called distributive. The Transport Layer TCP communications between two communication systems are split into three separate TCP sessions: A – Internet-connected communication system-to-GW PEP; PEP-to-PEP TCP session; B – Sat Terminal PEP-to-Sat-connected communication system.

Alternatively, an integrated architecture is obtained by eliminating the Sat Terminal PEP and using the Sat TCP protocol in the Sat-connected communication system. Other TCP splitting architectures have been studied, including that of splitting the end-to-end session onboard the satellite (Luglio et al. 2004). Additional T-PEP models and techniques are described in (38-ETSI PEP).

TCP PEP architectures offer the opportunity to split the end-to-end TCP session and enable the selection of optimal TCP protocols based on the network environment of each section. However, this approach has some problems and development challenges, three of which are discussed below:

- System Complexity. Again referring to Fig. 19.21, the T-PEP architecture necessitates the addition of one (Integrated architectural case) or two (Distributed case) PEP, each of which must provide the TCP processing and the related PEP application layer processing. The PEP on the GW (Internet) side must provide for each user full TCP session processing and buffering for the two TCP sessions: Session A and the PEP-to-PEP TCP session. This can require a large amount of GW processing and memory capacity, which becomes even more difficult for onboard PEPs. For Integrated PEP architectures, the receiving ST, user terminals, would be required to maintain duel protocol stacks.
- Network Security: The T-PEP is not compatible with the growing use of IP encryption, IPSec. End-to-end IP encryption is not possible because the TCP headers, which are encapsulated as data in the IP packets, are encrypted and thus not accessible for the intermediate processing needed to support TCP splitting using T-PEPs. Thus end-to-end secure communications would necessitate an application level encryption scheme.
- Mobility: The main challenges with mobility and the T-PEP architecture are those related to the GW PEP or, if onboard, the onboard Satellite PEP when a satellite user needs to handoff to either a different Gateway or a different satellite. This can occur when the satellite user moves from one beam to another beam that is in a different subnet (see Fig. 19.16) or with non-Geo Satellite constellations where the satellite passes through the coverage area. Changes in the middle of an end-to-end communication could introduce unacceptable disruption to the user communications as the entire complex TCP state of the GW PEP might need to be managed by another GW. Issues and solutions to T-PEPs and mobility are discussed in (18—Dublis PEPissues).

Cross-Layer Signaling

The importance of the challenge of TCP optimization has led to some solutions that violate the rules of the layered protocol models. One set of such solutions are those employing means to exchange information between non-adjacent layers, called cross-layer signaling. For example, a cross-layer protocol could provide a means for the Transport Layer to receive information from the Link Layer via message exchanges that are either in-band (using normally passed headers) or out-of-band (passing the information by other means, (e.g., SNMP messages)). The goal is to better understand the behavior of the satellite link in order to accurately control the Slow Start and Fast Recovery mechanisms and improve the overall network efficiency. Refer to (IET-Comm) for more on cross-layer signaling methods.

TCP and Web Acceleration: Turbopage®

As discussed, when using a geo-stationary satellite network for Internet access one concern is the latency due to the delay caused by satellite path propagation time. Typical Web page layouts require multiple hops to various servers in order to fetch the required data and graphical objects. HTTP Pre-fetch, which is marketed by Hughes as TurboPage, is an application acceleration feature that enhances the performance of the HTTP (Web browser) protocol by pre-fetching the HTML (Hypertext Markup Language) objects embedded in a Web page into the remote terminal before the application requests them. This feature uses a hub traffic server to request HTTP objects in advance and on behalf of browsers at the remote terminal. The proxy server pushes these objects to the remote VSAT where it is cached and, therefore, immediately available when the client browser requests the object. This concept is illustrated in Fig. 19.22. This feature provides significantly better user experience as Web pages are "painted" more quickly.

HTTP Pre-fetch is primarily aimed at improving Web page response time rather than at improving bandwidth efficiency. However, it does contribute to the latter in two ways. First, because objects are pre-fetched, GET requests for the objects do not need to be sent across the return channel, thus conserving bandwidth. Second, HTTP Pre-fetch applies V.44 compression to HTTP objects as they are forwarded, reducing forward channel bandwidth requirements.

IP Routers in Space

The challenge of reducing the latency and maximizing the transport efficiency has led to the consideration of placing an Internet router in space. The question remains if the IP router functionality supporting full IP satellite communication networks is best implemented in a hub on the ground or on board the satellite.

Claimed benefits of a spaceborne router include increased security, increased network manageability, increased capacity, and decreased delay.

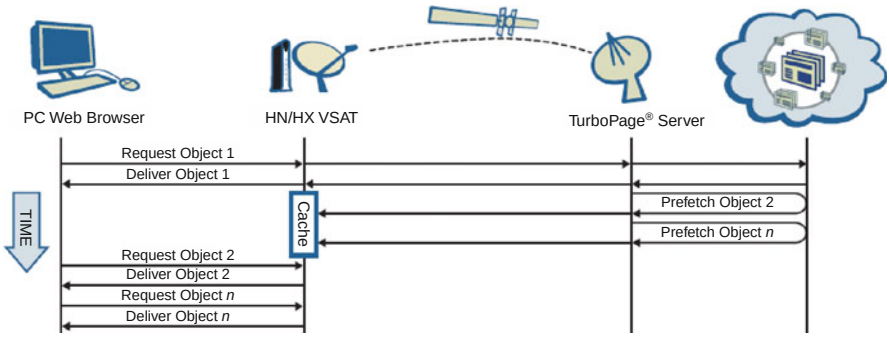


Fig. 19.22 HTTP turbopage pre-fetch

The actual benefit of implementing an IP router on a satellite depends on the service and comparable satellite architecture being implemented. These benefits, which in most cases are marginal, do not come without a cost as the modems and routers consume satellite resources that have direct relationship to satellite capacity.

The potential benefits of spaceborne processing for Internet routing derive from two functional areas: resource allocation algorithms and beam-to-beam switching. These two functions may be separated (one in Space, one at the Gateway) or combined.

The use of on-board routing for beam-to-beam switching permits a user in one beam to communicate directly with a user in another beam in "one-hop" through the satellite. The hub spoke architectures discussed previously are "two-hop" links. A user in one beam, say beam A, sends a message to the satellite which relays that message to a "gateway" Earth station. The gateway station recognizes the message is intended for a user in beam B, re-transmits the message up to the satellite, and the satellite downlinks the message to beam B. Providing beam-to-beam switching addresses two architectural points. It reduces the user-to-user delay, and reduces the Gateway bandwidth requirements.

Although Performance-Enhancing Proxies (PEPs) have and continue to minimize the impact of delay on the network efficiency, on-board routing has a real and significant advantage to applications requiring low delay. In situations where satellite users communicate directly with other satellite users through beam-to-beam routing, the delay is reduced by 250 ms since the RTT to the Gateway is eliminated. Applications that are real-time interactive, such as fast-paced gaming, are most affected by the added 250 ms delay if beam-to-beam routing is not available.

Interactive voice communication also receives much discussion regarding delay. Current MSS satellite networks have accepted the two-hop for mobile voice communications; however, there is a perceptible difference from terrestrial service, since users generally start recognizing delay when the round trip delay exceeds 480 ms.

As an aside, there are methods of achieving beam-to-beam connectivity in one-hop through a satellite that does not require an Internet Router on the spacecraft. For example, one or more transponders on a C-band uplink beam can be connected on a satellite to a Ku-band downlink beam, with the corresponding Ku-band uplink beam transponders connected to the C-band downlink beam. Then, a user in beam A who wants to send a message to beam B simply transmits the message in the appropriate uplink frequency which is routed through the satellite transponder directly to beam B. Such a connection, although effective in applications such as a VSAT network, is fixed, and represents a loss of usable capacity if traffic patterns do not match the fixed assignments.

However, a full on-board processing architecture makes the satellite literally a "switchboard in the sky." The satellite traffic is formatted into packets and each uplink packet includes the customized satellite system "address" of the intended downlink beam. This system has the advantage that, even with many beams, any uplink beam can be connected to any downlink beam, and the system can easily respond to changing traffic patterns. However, the on-board digital processing necessary to correctly route every packet requires a large fraction of the satellite resources (mass and power), leaving less resources for satellite RF transmit power which results in less overall capacity for the satellite network.

Since capacity is directly related to the satellites radiated power, satellite designers tend to optimize their architectures to maximize the power dedicated to EIRP. Any resources directed at processing remove resources from transmit power, thus having a negative effect on the total capacity.

In addition to the beam-to-beam interconnectivity advantage of an IP router, there is also a potential advantage to placing the resource allocation mechanism in space. A dynamic multiple access allocation algorithm reduces wasted bandwidth while maintaining service quality by allocating bandwidth to users only when data is available for transport and by leveling traffic by prioritizing access based on the application or user authenticated priority.

This process is generally performed in the Gateway Network Operations Center (NOC) today. However, two-hops are required for a user to establish a connection and receive their frequency and time slot assignment. Further, any changes to that allocation require another two-hops. If this function is placed in space, there will be less delay before the resource assignments can be made or changed. Consequently, more efficient use of the capacity is possible, by packing the available time, frequency, and/or code slots more fully. The smaller the data to transmit the greater efficiency on-board allocation algorithms provide. Since, with large amounts of data, the time to perform the allocation is a small percentage of the overall connection time.

In November 2009, Intelsat 14 (I-14) was launched with a payload that includes a "demonstration" Internet router in space (IRIS), as shown in Fig. 19.23 (Cisco_Website), with C-band and Ku-band beam coverage. This payload permits IP data over a portion of the bandwidth in any of three beams. A single transponder of 36 MHz bandwidth for each of the three beams is connected to the router.

Experiments will continue with the I-14 payload to help assess the true benefits of implementing a full payload with Internet routing. Nevertheless, two major



Fig. 19.23 Internet router in space launched on intelsat-14

hurdles must be overcome: justifying the reduction in the total satellite capacity by directing limited satellite resources from transmit power to IP processing, and having sufficient flexibility and robustness in the spaceborne router to keep up with technological advances and the evolution of traffic profiles.

Conclusion

There are many technical challenges in the integration of satellite and terrestrial networks. Both MSS and Broadband Satellites for Internet access provide service today through innovative solutions to many of these challenges. Nevertheless, enhancements to the current solutions, as well as new solutions will be needed to keep up with the rapid advances in the terrestrial marketplace.

The long life of satellites (in excess of 15 years), when compared to terrestrial products, will continue to demand integrated solutions which design the satellite node of the architecture to be as flexible, robust, and transparent as possible. While some limitations are unavoidable in the satellite node, such as the frequency band, orbit, and round trip time, clever implementations in the satellite payload design and in the ground node such as PEPs and GBBF can provide the flexibility needed for satellite networks to continue to play an important role in the rapidly advancing network services and applications throughout the world.

References

(Lindberg-Walker) The emulation of satellite communication systems in load-stressing conditions, in *The Proceeding of the 2003 Conference of the American Institute of Aeronautics and Astronautics*, Monterey, May 2003

I.F. Akyidldiz, G. Morabito, S. Palazzo, TCP-peach: a new congestion control scheme for satellite IP networks. IEEE/ACM Trans. Network. **9**(3), June 2001

M. Allman, D. Glover, L. Sanchez, Enhancing TCP over satellites channels using standard mechanisms. RFC 2488, Jan 1999

M. Allman, W. Paxson, TCP Congestion Control. RFC 2581, April 1999

S. Brandner, The Internet Standards Process. RFC 2026, Version 3, Oct 1996

C. Caini, R. Firrincieli, End-toEnd TCP enhancements performance on satellite links, in *Proceedings of the 11th IEEE Symposium on Computers and Communications* 2006a, pp. 1031–1036

C. Caini, R. Firrincieli, D. Lacamera, PEPsal et al.: a performance enhancing proxy designed for tcp satellite connections, in *Vehicular Technology Conference*, 2006b. VTC 2006-Spring, IEEE 63rd, vol. 6, May 2006, pp. 2607–2611

C. Caini, R. Firrincieli, Comparative performance evaluation of TCP variants on satellite environments, in *Proceedings of the IEEE International Conference on Communications* 2009, pp. 1–5

CCSDS, Recommendations for Space Data System Standards, CCSDS 714.0-B-2. Blue Book, Issue 2, Washington, DC.: CCSDS, Oct 2006

Cisco Website: http://www.cisco.com/web/strategy/government/space-routing.html

D.C. Cox, H.W. Arnold, R.P. Leck, Phase and amplitude dispersion for earth-satellite propagation in the 20 to 30 GHz frequency range. IEEE Trans. Antennas & Propag. **AP-28**(3), 359–366 (1980)

L. Cui, S.J. Koh, X. Cui Enhanced wireless TCP for satellite networks. *International Conference on Wirelsss Communications, Networking and Mobile Computing*, 2007, pp. 1813–1816

DOCSIS 3.0, MAC and Upper Layers Protocol Interface Specification. CM-SP-MULPIv3.0-I14-101008, Oct 2010

DOCSIS 3.0, Physical layer specification, CM-SP-PHYv3.0-I09-101008, Oct 2010

E. Dubois, et al., Enhancing TCP based communications in mobile satellite scenarios: TCP PEPs issues and solutions, in *Proceedings of the Fifth Advanced Satellite Multimedia Systems Conference and the 11th Signal Processing for Space Communications Workshop*, 2010, pp. 476–483

W. Dusi, W. John, Estimating routing symmetry on single links by passive flow measurements, in *(draft presentation for) the 1st International Workshop on TRaffic Analysis and Classification (TRAC) colocated with the 6th International Wireless Communications & Mobile Computing Conference (IWCMC 2010)* in June 2010

ETSI EN 302 307 V1.2.1, DVB-S2, 2009

ETSI TR 101 984 V1.2.1 (2007-12): Satellite Earth Stations and Systems; Broadband Satellite Multimedia; Services and architectures

ETSI TR 102 676 SES, BSM, PEPs. V1.1.1, Nov 2009

ETSI TS 102 354 V1.2.1 (2006-11): Satellite Earth Stations and Systems; Broadband Satellite Multimedia; Trnasparent Satellite Star; IP over Satellite Air Interface Specification

Fall 2010 Global internet phenomena report. Sandvine Corporation, Oct 2010 (Available online: www.**sandvine**.com/amp;/**2010**%20Global%20Internet%20Phenomena%20Report.pdf)

FCC, Notice of proposed rulemaking (In the Matter of Flexibility for Delivery of Communications by Mobile Satellite Service Providers in the 2 GHz Band, the L-Band, and the 1.6/2.4 GHz Band), FCC 01-225, IB Docket No. 01-185, Aug 2001

FCC, Memorandum Opinion and Order and Second Order on Reconsideration, FCC 05-30, IB Docket No. 01-185, Feb 2005

S. Floyd, T. Henderson, The NewReno modification to TCP's fast recovery algorithm. RFC 2582, April 1999

J.D. Gibson, *The Communications Handbook* (CRC Press, Boca Raton, 1997)

Y.F. Hu, M. Berioli et al., Broadband satellite multimedia. Inst. Eng. Technol. Commun. **4**(13), 1519–1531 (2010)

Hughes Network Systems, IP over Satellite (IPoS) – The Standard for Broadband over Satellite, June 2007

L.J. Ippolito, Propagation effects handbook for satellite systems design. NASA Doc. 1082, Feb 1989
ISO 13537:2010 Space data and information transfer systems – reference architecture for space data systems, 2010
ITSI EN 301 790 V1.5.1, DVB; Interaction Channel for Satellite Distribution Systems, 2009
R.C. Johnson, H. Jasik, *Antenna Engineering Handbook*, 2nd edn. (McGraw Hill, New York, 1984)
L. Joing, C. Zhigng, M.J. Khan, TP-Satellite: a new transport protocol for satellite IP networks. IEEE Trans. Aerosp. Electron. Syst. **45**(2), 502–515 (2009)
M. Luglio, M.Y. Sanadidi, M. Gerla, J. Stepanek, On-board satellite "split TCP' proxy. IEEE J. Sel. Areas Commun. **22**(2), 362–370 (2004)
M. Mathis, et al., TCP selective acknowledge options. RFC 2018, Oct 1996
Network Simulator (NS-2) (Online). Available: http://www.isi.edu/nsnam/ns/. Accessed Nov 2010
O3B Networks, http://www.o3bnetworks.com. Accessed Nov 2010
E. Rendon-Morales, et al., Cross-layer architecture for TCP splitting in the return channel over satellite networks, in *Sixth International Symposium on Wireless Communications Systems*, 2009, pp. 225–229
C. Roseti, M. Luglio, F. Zampognaro, Analysis and performance evaluation of a burst-based TCP for satellite DVB RCS links. IEEE/ACM Trans. Netw. **18**(3), 911–921 (2010)
I-PEP Specifications, Satlabs Group Recommendations, Issue 1a., Oct 2005 (on-line: www.satlabs.org) Specific profile of SCPS-TP
W.R. Stevens, *TCP/IP Illustrated*, vol. 1 (Addison-Wesley, Reading, 1994)
Terrestar Networks, http://www.terrestar.com. Accessed Nov 2010
Transmission Control Protocol, DARPA Internet Program Specification, RFC 793, Sept 1981
J.L. Walker, B. Day, S. Xie, Architecture, implementation and performance of ground-based beam forming in the DBSD G1 mobile satellite system, in *AIAA 28th ICSSC Conference Proceedings* Aug 2010

Satellite Communications Regulatory, Legal and Trade Issues

20

Gérardine M. Goh Escolar

Contents

Introduction 508
International Regulation of Satellite Communications 509
Satellite Communications in the International Legal System 509
International Standards 511
International Satellite Operators: Regulatory and Legal Issues 511
Regional Regulation of Satellite Communications 514
Arab Satellite Communications Organization (Arabsat) 514
Association of Southeast Asian Nations (ASEAN) 515
Asia-Pacific Satellite Communications Council (APSCC) 515
European Regulatory Authorities 516
NAFTA 519
Regional African Satellite Communications Organisation (RASCOM) 521
National Regulation of Satellite Communications 521
Satellite Communications in Global Trade 522
The General Agreement on Trade and Tariffs (GATT) and the World Trade Organization (WTO) 522
The General Agreement on Trade in Services (GATS) 523
The Annex on Telecommunications 524
The Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPs) 525
Dispute Settlement, Responsibility, and Liability Issues 526
Dispute Settlement 526
Responsibility and Liability 527
Cross-References 529
Notes 529
References 531

G.M. Goh Escolar
International Court of Justice, United Nations, c/o Vuistbijl 77, 2497 DT, The Hague, The Netherlands
e-mail: gerardine.goh@gmx.net

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_23, © Springer Science+Business Media New York 2013

Abstract

This chapter focuses on the regulatory, legal, and trade issues related to satellite communications. The chapter first examines the regulatory and legal issues on three levels: the global or international arena, regional regulatory institutions, and national regulatory frameworks. Next, the chapter focuses on a discussion on the role of satellite communications in global trade, and will review the regulatory, legal, and trade issues of satellite communications on the global scale. The last part will discuss issues related to the settlement of disputes that may arise from satellite communications, as well as the legal principles of responsibility and liability for any damage caused by satellite communications. Note that this chapter does not deal with the legal and other issues related to the International Telecommunications Union and radio frequencies, which have been dealt with in previous chapters of this book.

Keywords

Arabsat • Asia-Pacific Satellite Communications Council (APSCC) • Association of Southeast Asian Nations (ASEAN) • Dispute settlement • EUMETSAT • European Space Agency • EUTELSAT • General Agreement on Trade and Tariffs (GATT) • General Agreement on Trade in Services (GATS) • INMARSAT • Intellectual property • INTELSAT • International law • INTERSPUTNIK • Legal issues • Licensing • NAFTA • National legislation • Outer Space Treaty of 1967 • RASCOM • Regional organizations • Regulatory framework • United Nations Committee on Peaceful Uses of Outer Space (COPUOS) • World Intellectual Property Organization (WIPO) • World Trade Organization (WTO)

Introduction

The satellite communications industry is a complex multi-billion dollar global industry that involves a diverse amount of regulatory, legal, and trade issues. This is because the technology involves international and national allocations of frequencies and safety and environmental provisions. The patterns of international regulation of satellite communications has shifted greatly in past decades with a movement toward privatization of satellite communications systems and the use of regulated competitive systems to bolster the range of service offerings and to reduce the cost of service. The pattern of this regulatory shift is reported in some detail.

Satellite communications is also an international undertaking that involves international treaty and regulatory provisions relating to strategic and defense-related consideration, the discrimination between air space and other space, and, in particular, to the activities of various United Nations entities and interpretation of a number of space-related treaties in the international court system. Many of the subjects discussed in this chapter involve complicated legal interpretations of international law, treaties, and conventions. Rather than seeking to address each

of these areas in great depth, the approach has been to provide an overview of the issue and then indicate by reference where even more fine-grained information is available.

International Regulation of Satellite Communications

As a component of international space-based telecommunications, satellite communications are bound by the rules of international law generally[1], and international space law specifically (Salin 2000). Crucially, satellite communications are not primarily governed by the principles of international air law. Despite the academic debate on the delimitation on airspace and outer space[2], satellites clearly fall into the category of "space objects." As such, international space law applies.

Further, regulations and standards pertaining to satellite communications have been promulgated by international standards setting organizations. These standards, which are derived from a recognized technical competence, deal with the nuts-and-bolts of launching, maintaining, and operating a satellite communication system. Further nuances of the international regulation of satellite communications are added by the activities of international satellite operators and multinational satellite consortia.

Satellite Communications in the International Legal System

The 1967 Outer Space Treaty incorporates many basic principles of international law, and also includes regulations specific to activities in the outer space environment.[3] These include the following principles:

- States are obliged to act for the benefit and in the interests of all countries in conducting their space activities.
- Outer space shall be the province of all mankind.
- Outer space shall be free for exploration and use by all States.
- Outer space shall not be subject to national appropriation by claim of sovereignty, use or occupation, or any other means.
- International law shall apply to activities in outer space.
- The placement of nuclear weapons or other weapons of mass destruction in orbit, installation of such weapons on celestial bodies, or the stationing of such weapons in outer space is prohibited.
- States bear international responsibility for national activities in outer space, whether such activities are carried on by governmental agencies or by nongovernmental entities.
- Activities of nongovernmental entities shall require authorization and continuing supervision by the State.
- A launching State is internationally liable for any damage caused by the space object it launched or procured the launch of.
- A State has jurisdiction and control over space objects that it has registered.

- States are to cooperate and mutually assist other States with activities in outer space, and shall conduct all space activities with due regard to the interests of other States.
- States are to avoid harmful contamination and adverse changes in the environment of the Earth in the conduct of their space activities, and must consult with other States where harmful interference may occur.

The Outer Space Treaty is supplemented by several other international treaties applicable to satellite communications, including the following:

- 1971 INTELSAT Agreement
- 1971 INTERSPUTNIK Agreement
- 1972 Liability Convention
- 1974 Brussels Convention
- 1975 Registration Convention
- 1975 European Space Agency Convention
- 1976 INMARSAT Agreement
- 1976 Intercosmos Agreement
- 1976 Arabsat Agreement
- 1982 EUTELSAT Convention
- 1983 EUMETSAT Convention
- 1994 International Telecommunications Union Convention

It must be noted that much of the basic framework of public international space law was created through the organs of the United Nations, and in particular through the work of the Committee on the Peaceful Uses of Outer Space (COPUOS).[4] Aside from the negotiation and drafting of the treaty framework applicable for activities in outer space, work originating from the COPUOS has also led to the adoption of various relevant United Nations General Assembly (UNGA) Resolutions. Although such UNGA Resolutions are not binding at international law, they are indicative of State practice and *opinio juris*. As such, some principles framed in the UNGA Resolutions relating to activities in outer space are considered to have become customary international law. In the context of satellite communications, the two UNGA Resolutions that are of particular significance are as follows:

- The Declaration of Principles Governing the Use by States of Artificial Earth Satellites for International Direct Television Broadcasting
- The Declaration on International Cooperation in the Exploration and Use of Outer Space for the Benefit and in the Interest of all States, taking into account the Needs of Developing Countries

Certainly, customary international law dealing with activities in outer space,[5] as well as other soft law provisions, also applies to satellite communications.

It has been lamented by highly qualified publicists that the rate of law-making at the United Nations in relation to activities in outer space has considerably decreased in recent years. However, it is interesting to note that this deceleration has not affected issues related to space-based communications. Instead, there is a burgeoning increase in the development of technical standards and codes of conducts in recent years that are directly related to satellite communications (Jasentuliyana 1995).

Of particular interest is the interaction of public international law with the evolution of two characteristics of modern satellite telecommunications: commercialization and liberalization[6]. Certain principles inherited from the public international law roots of the law relating to satellite communications will necessarily have bearing on the more recent developments in the field. In particular, the principle that activities in outer space should be undertaken for the benefit and in the interests of all countries, irrespective of the fact that million- or billion-dollar investments may have been made by private commercial entities, appears to be untenable (Christol 1991).

Other legal principles jostle for attention in the increasingly liberal and commercial marketplace for satellite communications. Among these are the international responsibility and liability imposed on nation States for space activities,[7] restrictions on military activities in outer space,[8] and principle of non-appropriation.[9] Certainly, as some highly qualified publicists have pointed out, there is "some kind of naïveté" in the idea that commercial entities that have invested vast sums in the construction, infrastructure, launch, operation, and maintenance of a satellite communication system would be amenable to sharing the benefits of this system with those who have invested nothing.

International Standards

Technical and policy standards impacting satellite communications are also established by other international and professional organizations. The harmonization of technical and other protocols has necessitated the establishment of clear standards to ensure interoperability. Professional and technical organizations such as the International Standards Organization (ISO) and the International Electrotechnical Commission (IEC) have established standards approved and in use by international, national, and private organizations dealing with satellite communications.

Standards have also been set by some regional and national organizations. These include the European Telecommunication Standards Institute (ETSI), the Standards Committee T1 Telecommunications (North America), and the Japanese Telecommunications Technology Committee. To avoid any possible overlaps or contradictions, the International Telecommunications Union has ensured interoperability and cooperation through the Interregional Telecommunication Standards Conferences (ITSC), which held its inaugural meeting in 1990.

International Satellite Operators: Regulatory and Legal Issues

What is readily apparent from the structure and framework of these international satellite operators (and their regional siblings discussed in the next section) is the cross-cutting nature of their operations. These organizations straddle the line between public regulatory agency and private commercial enterprise. While regulating satellite communications in line with their respective constituent agreements,

many of these organizations are also satellite owners and operators. This leads to the interesting paradigm of the dual-nature public/private entity that promulgates the policies and regulations its commercial activities are then bound by.

INTELSAT

INTELSAT was established in August 1964 with the signature by 11 Member States of an Interim Agreement and a Special Agreement. A Multilateral Agreement and an Operating Agreement signed in 1971, followed by a Headquarters Agreement signed in 1976, set up its organization. Initially an international intergovernmental government, its membership complement consisted of States that were members of the ITU. Each Member State had one vote and subscribed to the capital of INTELSAT in proportion to its use of the INTELSAT space segment. As such, it combined the characteristics of a public intergovernmental organization and those of a private commercial entity. The case of INTELSAT is particularly noteworthy as a sterling example of trends toward globalization and commercialization.

INTELSAT was composed of its Assembly of Parties, the Meeting of Signatories, the Board of Governors, and the Executive Organ. Member States are represented by the national telecommunications agencies. Historically, only Member States had access to the INTELSAT system, although they were also free to have their own satellites, as long as these were compatible with the INTELSAT system. They were also free to resell capacity on the INTELSAT system to end users. Nonmember governments may become "Duly Authorised Telecommunications Entities" (DATEs), and contract with INTELSAT for the use of the system. Direct Access Customers, which are not Member States and which may also be investors in INTELSAT, came onto the scene slightly later.

INTELSAT exercised some regulatory control over its own, as well as other separate, satellite systems. INTELSAT's mission is to ensure, on a commercial basis, the provision of the space segment necessary for international public telecommunications services of quality and reliability. As a result of Articles III and XIV of the Multilateral Agreement, it mandates compulsory coordination for separate satellite systems. Where Member States use INTELSAT's services for the provision of domestic specialized telecommunications services, these must be coordinated with the INTELSAT system. It will be noted that Article XIV of the Multilateral Agreement in particular specifies the rights and obligations of Member States, including compliance with the principles of INTELSAT. Pursuant to Article XIV, this means that, for its domestic public telecommunications services, Member States must consult with the Board of Governors for recommendations, and for its international public telecommunications services, Member States must consult with the Assembly of Parties. This is to ensure technical compatibility of their facilities and operations with the INTELSAT space segment.

In the 1990s, it became clear that INTELSAT was successful in establishing a "global commercial telecommunications satellite system and Organisation . . . to provide expanded telecommunication services on a non-discriminatory basis to all areas of the world." As a result of the changes in the economics of international telecommunications, the deregulation of the space market, and the intent of INTELSAT to expand its activities and business concerns, INTELSAT was

restructured and privatized on July 18, 2001, becoming Intelsat Ltd. and is now internationally headquartered in Bermuda. It was sold in January 2005 to four private equity firms, and then subsequently resold to other equity firm holdings and also it merged with and took over the Panamsat system. Intelsat continues to generate revenue through satellite usage fees. A small International Telecommunications Satellite Organization was also established as of 2001 to address international concerns related to satellite services as safety and the provision of satellite services to developing countries.

INMARSAT

INMARSAT was founded in 1979 by the Inter-Governmental Maritime Consultative Organization. The INMARSAT Convention and Operating Agreements of 1976 were originally aimed at the provision of transponders for ship and maritime communications via satellite. In 1982, INMARSAT became an international maritime satellite organization. Under the name INMARSAT-II, it later became a global maritime satellite network system in 1990, providing mobile communications for users at sea, on the ground, and in the air.

The original INMARSAT Convention binds States, while its Operating Agreement had States and other entities as members. It combined, as did INTELSAT above, the characteristics of public service and private commercial activity. The Preamble of its Convention refers to Article I of the Outer Space Treaty. INMARSAT comprised an Assembly of Parties, a Council, and a Directorate. The Assembly of Parties was composed of all Parties, which made recommendations to the Council in regard to "the activities, purposes, general policy and long-term objectives of the Organisation." This Assembly included States as well as non-State entities. The INMARSAT Council comprised 22 members, representing its 18 largest contributing members and 4 geographical members elected by the Assembly. It managed the INMARSAT space segment. The Directorate, which reported to the Council, administered the Organization.

It must be noted that INMARSAT took a different approach than INTELSAT in allowing parties to contract as between themselves, while using INMARSAT facilities. As such, regulatory policies promulgated by INMARSAT may or may not have had an impact on inter-Party (extra-organization) dealings. Significantly, however, INMARSAT had quasi-public authorities, such as the authority to compel private competitors to provide early warning of any projected activities that would be in competition with INMARSAT.

With the pressure of increasing commercialization and competition, INMARSAT was privatized in 1999. It was thus the first of the international satellite treaty organizations to make this transition. INMARSAT was divided into a commercial entity, Inmarsat plc (headquartered in London), and a regulatory body, the International Mobile Satellite Organisation (IMSO). IMSO continues to provide public regulatory services such as the administration of the Global Maritime Distress Safety System (as established by the International Maritime Organization), and the public service component established through the Standards and Recommended Practices established by the International Civil Aviation Organization (ICAO).

INTERSPUTNIK

Intersputnik International Organization of Space Communications was established on November 15, 1971, in accordance with the Intergovernmental Agreement on the Establishment of the Intersputnik International System and Organization of Space Communications. It is an international organization, with its seat in Moscow, and vested with the right to execute contracts, acquire, lease, and alienate property, and to institute proceedings. It develops and maintains contacts and cooperation with other global, regional, and commercial satellite organizations. Intersputnik was registered with the United Nations on March 27, 1973. It has the status of permanent observer at the UN Committee on the Peaceful Uses of Outer Space, the ITU, and UNESCO. It is a member of the Asia-Pacific Satellite Communications Council, and the Global VSAT Forum.

Today, Intersputnik has 25 Member States, including countries in Latin America, the Arab Peninsula, and Asia. Its core business is the lease of satellite capacity to telecommunications operators, broadcasters and other clients, as well as services for the establishment, engineering, and operation of satellite networks. The organization procures and deploys spacecraft in orbit. It is also an international organization that participates in international and intergovernmental fora. The Board and Operations Committee work in tandem with the Directorate to cover issues such as policy-planning, technical standards promulgation, determination of the amount of share capital, compliance with its international obligations, and establishment of its financial and commercial policies. Through its fully owned subsidiary Intersputnik Holdings, it also conducts commercial satellite communications activities through three daughter companies: Isatel Russia, Isatel Kyrgyzstan, and Isatel Tajikistan.

Regional Regulation of Satellite Communications

Aside from the international satellite operators mentioned above, there are several regional organizations using and regulating space-based satellite communications. These include:

1. The Arab Satellite Communications Organization (Arabsat)
2. The Asia-Pacific Satellite Communications Council (APSCC)
3. The Association of Southeast Asian Nations (ASEAN)
4. Various European institutions such as EUTELSAT, EUMETSAT, and the European Space Agency (ESA)
5. The grouping of the North Atlantic Free Trade Agreement (NAFTA)
6. The Regional African Satellite Communications Organisation (RASCOM)

Arab Satellite Communications Organization (Arabsat)

Arabsat was founded by the 21-member Arab League in 1976. It is an intergovernmental organization with a paid capital of USD $500 million, paid up in various proportions by its Member States. Its establishment in 1976 initially aimed to design, execute, and operate the first Arab space system for satellite communications. The

first of its first-generation satellites was launched in 1985. Today, Arabsat is manufacturing, launching, and operating its third-generation satellite network.

Arabsat carries more than 400 television channels and 160 radio stations, reaching an audience of more than 164 million in the Arab Member States alone. It offers broadcast, telecommunications, and broadband services, including data network solutions, telephony, and Internet provider trunking backbone connectivity, as well as broadband Internet access.

Arabsat comprises a General Assembly of Member States, a Board of Directors, and a Management Committee. The General Assembly negotiates and approves satellite communications policies for the region and the Arabsat network. The Board of Directors holds periodic meetings to secure, invest in, and maintain the satellite assets of the Arabsat Organization, as well as to implement policies passed by the General Assembly. The Management Committee is responsible for the day-to-day operations of the Arabsat network.

Association of Southeast Asian Nations (ASEAN)

The Association of Southeast Asian Nations (ASEAN) is a ten-member grouping of countries founded in 1967. Its declared aim is to "accelerate the economic growth, social progress and cultural development in the region through joint endeavours" and to "collaborate more effectively for . . . the expansion of their trade, including the study of the problems of international commodity trade, the improvement of their transportation and communications facilities and the raising of the living standards of their peoples."

In the context of satellite communications, the Report of the meetings of the Sub-committee on Posts and Telecommunications is of pertinence. This Sub-committee has purview over the Integrated Work Programme in Posts and Telecommunications, including domestic and regional satellite communications and radio frequency coordination in ASEAN. The ASEAN Radio Frequency Coordination Committee has also been established to coordinate the use, the administration, and the management and planning of radio frequency in the ASEAN region especially in the border areas. The Secretariat of the Committee is situated in Jakarta. The Sub-committee proposes broad policies that are considered by the ASEAN Member States and that, if and when approved, are implemented by the Secretariat.

Asia-Pacific Satellite Communications Council (APSCC)

The Asia-Pacific Satellite Communications Council (APSCC) grew out of a proposal by the government of the Republic of Korea, presented at the United Nations Workshop on Space Communications for Development in the Asia-Pacific held in November 1992. The APSCC Constitution was adopted in 1994. Today, APSCC represents all sectors of satellite and space-based industries in the region under the purview of the United Nations Economic and Social Commission for Asia and the Pacific (ESCAP). It aims to promote satellite communications and broadcasting for the socioeconomic and cultural development of the ESCAP region.

APSCC provides the forum at which its members may exchange views on policies, technologies, systems, and services in relation to satellite communications. It counts satellite manufacturers and operators, launch service providers, satellite risk management companies, telecommunication carriers, and broadcasters among its members. It aids in the formulation of recommendations on policies, regulations, and technical standards within the ESCAP region, and in relation to the global satellite communications arena. APSCC also works with Member States to minimize technical and regulatory barriers to the deregulation of the satellite communication industries in ESCAP. To that end, it cooperates with other international institutions such as the United Nations Office for Outer Space Affairs, ESCAP, the ITU, and the Asia-Pacific Economic Cooperation (APEC).

European Regulatory Authorities

The particular case of Europe merits special attention. The central regulation exercised by the European Union, together with its structures and institutions, cannot be examined to the detail they deserve within the scope of this chapter. Reference instead should be made to the plethora of learned writings on the topic.[10] This section will only look at the institutions that have a direct impact on the legal and regulatory framework of Europe on satellite communications.

The European Legal Framework on Satellite Communications

The European Commission (EC) has been very active in promulgating Community law pertaining to telecommunications, space activity, and satellite communications. The various EC Directives, Resolutions, Decisions, Recommendations, White Papers, Green Papers, Guidelines, and Communications issued since 1986 go a long way to show the EC policy on satellite communications.

In particular, the EC has focused on regulating the following issues:

- The harmonization of domestic regulations on satellite communications
- The liberalization of the Earth segment and related market sectors
- The competition between satellite operators
- The application of EC law and the development of the single European internal market

Further, the EC has also sought to ensure the separation of regulatory and operation functions in Member States, granting free and unrestricted access to the space segment capacity, while ensuring commercial freedom for space segment providers. It is significant to note that the European authorities may intervene in cases of restrictive or anticompetitive business practices, including corporate strategy development and monopolistic actions.

The harmonization of domestic regulations on satellite communications was established by the Telecommunications and the Satellite Green Papers first issued in 1987 and 1990. Measures undertaken include framework Council Directors on:

- *The Open Network Provision (ONP)*: All ONP conditions must be objective, transparent, and published, and should guarantee equal access under

nondiscriminatory principles of Community law. Access to public telecommunications networks or services cannot be restricted except for reasons of essential requirements, and such restrictions must comply with EC law.

- *The Protection of the Satellite Communications*: This includes the protection of all forms and content of satellite communications, including copyright, the protection of databases and encrypted services, as well as the protection of neighboring rights. This established the author's exclusive right to authorize the diffusion of his/her work via satellite, and elaborates upon the applicable EC law relating to trans-boundary retransmission.
- *The Standardization of Satellite Communications Equipment*: This ensured harmonized procedures for certification, testing, quality assurance, and product monitoring for satellite communications equipment throughout the Community. It also harmonized domestic laws, regulations, and standards pertaining to safety, compatibility, interoperability, and interworking of telecommunications network equipment and terminal equipment.
- *The Reciprocal License Recognition for Satellite Services*: This ensured that licenses for satellite services issued by one Member State would be recognized in other Member States of the Community, replacing the system by which a satellite service provider would have to apply for different licenses in different domestic legal systems.

These measures were aimed at allowing the development of the European internal market, lifting trade and technical barriers across borders, while ensuring the freedom of access, use, and provision of satellite communications services. They were also aimed at the provision of a competitive environment for the progress of the European satellite communications industry as a whole.

The liberalization of telecommunications and satellite communications in the Community was aimed primarily at the following:

- *Separation of regulatory and operational functions*: This was meant to correct the monopoly that most national telecommunications operators had in the domestic market of Member States, and to ensure free market access for private operators.
- *Freedom of competition of space segment capacity providers*: This objective of this measure was to ensure open and free competition among providers of the space segment equipment and services.
- *Liberalization of infrastructure*: This intended to ensure a common approach to the telecommunications infrastructure in the Community. Liberalization was provided for so as to ensure competition not only in the equipment and services sector, but also in infrastructure provision and operation. Further, the application of the EC's competition rules allowed for fairness and transparency.
- *Access to the space segment*: This measure aimed to ensure direct access throughout all Member States, ensuring the joint reform of international satellite organizations and joint management of the space segment as an essential resource of the Community and the European Union. It also aimed at providing a competitive environment for European industry players.

Regulation as to competition in the satellite communications industry was also given special attention by the Commission. In particular, action has been

undertaken in regard of restrictive business practices, monopolies, dumping practices, and anticompetitive practices. This is achieved through the review of strategic alliances, service arrangements, as well as mergers and acquisitions of satellite communications entities. The domestic application of EC law in the context of the internal market also undergoes periodic review by the Commission to ensure full compliance with European standards.

EUTELSAT

The European Telecommunications Satellite Organization (EUTELSAT) was established in 1982, upon the opening of its Convention at the Paris Convention. Prior to the entry into force of the Convention, it operated under interim agreements entered into on a provisional basis. The establishment of EUTELSAT came as a regional impetus from European States to increase European participation in space and satellite-related activities, and also as a European response to INTELSAT. Sixteen Member States of the European Conference of Postal and Telecommunication Administrations came together to establish EUTELSAT, which aimed to become a global leader in fixed satellite services.

EUTELSAT consisted of an Assembly of Parties, a Board of Signatories, and an executive organ. Its structure is similar to that of INTELSAT's, with the Board of Signatories in this case undertaking the responsibilities of INTELSAT's Board of Directors. Additionally, EUTELSAT's Board of Signatories was charged with ensuring compliance with the ITU Radio Regulations rules. Procurement was by international invitation to tender, with consideration given to the Parties and Signatories. EUTELSAT's practice in regard of coordination differed from that of INTELSAT. EUTELSAT recognized separate systems – although acknowledging that coordination was essential so as to ensure technical compatibility and interoperability, EUTELSAT recognized the separation between systems launched by its Member States and those launched by private European entities.

With the liberalization of the European satellite communications industry, the public operations and activities of EUTELSAT were transferred to a private company, Eutelsat S.A., in July 2001. In April 2005, Eutelsat Communications, the holding company of Eutelsat S.A., was established. Eutelsat continues to serve the space, telecommunications, and audiovisual industries of Europe, the Middle East, Africa, and some parts of Asia and the Americas.

EUMETSAT

The European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) is an intergovernmental organization established in 1983. At the time of writing, it had 26 European Member States and 5 Cooperating States. EUMETSAT's mission is to establish, maintain, and exploit European systems of operational meteorological satellites. Member States make mandatory contributions to the organization, which are proportional to their gross national income.

EUMETSAT works closely with the European Meteorological Infrastructure (EMI) to respond to the national meteorological needs of its Member States. It ensures the European contribution to the space-based component of the World

Meteorological Organization (WMO), while continuing to establish bilateral and multilateral agreements from non-EUMETSAT countries and projects. It also conducts environmental monitoring and disaster management activities.

EUMETSAT comprises a Council, an Executive body, and a Director. The Council is composed of two representatives for each Member State. EUMETSAT's data policy has been harmonized to ensure equal and nondiscriminatory access for both paying customers and public authorities, in line with the European legislation.

The European Space Agency (ESA)

The European Space Agency (ESA) was established in 1975 and is an intergovernmental organization with 18 Member States. It aims to promote for peaceful cooperation among European States in space research and technology, by elaborating a European space policy and harmonizing the national space policies of Member States. Headquartered in Paris, ESA coordinates the European space program, and integrates national space programs of Member States especially in regard of the development of applications satellites. It also elaborates and implements a coherent industrial policy in regard of space activities of its Member States.

ESA is not an agency or institution of the European Union (EU). Non-EU countries such as Switzerland and Norway are members of ESA. However, it must be noted that there has been work on defining the legal status of ESA in regard of the EU. The EU and ESA corporate on some satellite projects such as the Galileo satellite navigation system. The 2007 European Space Policy commits the EU, ESA, and their respective Member States to increase coordination in their space activities. Further integration of ESA into the framework of the EU may mean greater coordination and harmonization in the policies, regulations, and legislations relating to satellite communications.

It must be noted that ESA also comprises a Directorate of Telecommunications and Integrated Applications. ESA activities in regard of satellite communications aim to

1. Develop advanced technologies for both the space and ground segment
2. Develop and implement full systems for new capacities
3. Implement new missions in partnership with industry
4. Develop and implement new applications for satellite communications

In order to achieve these goals, the Directorate has executed a programmatic framework known as the Advanced Research on Telecommunication Satellite Systems (ARTES). Commercial entities within ESA Member States may submit proposals to the ARTES program, which provides for strategic analysis, funding, technical development, product demonstration, and deployment of the proposed satellite systems.

NAFTA

The North American Free Trade Agreement (NAFTA)[11] was signed by Canada, Mexico, and the United States on December 17, 1992. Upon receipt into the

national legal system of the signatory States and the fulfillment of all necessary procedures, NAFTA entered into force on January 1, 1994. This section will only address chapters of NAFTA that pertain to satellite communications:

- Chapter 12 on Cross-Border Trade in Services
- Chapter 13 on Telecommunications
- Chapter 15 on Competition Policy, Monopolies, and State Enterprises
- Chapter 17 on Intellectual Property

Chapter 12 deals with cross-border trade in services, which refers to any kind of service in all their various phases. It does not apply to financial services, public procurement, public subsidies or grants, and air services. Chapter 12 stipulates that the following principles apply to satellite communications:

- National treatment of nationals of the other two parties
- Most Favored Nation treatment of nationals of the other two parties
- Best standard of treatment among the above

Significantly, licensing, certification, and legal consulting requirements must be coordinated between the parties.

Chapter 13 applies in particular to the transfer of data, electronic exchange of information, and the maintenance of intra-corporate networks. For each of these categories, basic services are distinguished from value-added services. The provisions of Chap. 13 provide for open access and nondiscrimination for public networks. It also provides for common regulations on technical standards that operators must conform to. Monopolies are allowed insofar as these do not engage in anticompetitive actions. Generally, the chapter provides for harmonization to ensure interoperability and cooperation, as well as transparency to ensure competitive fairness. Chapter 13 does not apply to the regulation of direct broadcasting services and non-public network-connected telephony satellite services.

Chapter 15 stipulates that monopolies may not engage in anticompetitive conduct. While it does not prevent the designation of a monopoly or a State enterprise, prior notification must be given to the other two parties, and the monopoly in question may not behave in a discriminatory or anticompetitive manner.

Chapter 17 on intellectual property reflects the agreements achieved in the World Trade Organization as discussed below. In particular, it provides that unauthorized decoded and distribution of program-carrying satellite signals is a violation of NAFTA. Further, copyright protection is afforded for computer programs and data compilations. Parties are allowed to grant even more extensive protection, as far as these are not inconsistent with NAFTA provisions.

It should be noted that Part 8 of NAFTA provides for certain exceptions to the obligations outlined briefly above. Among these exceptions are national security and certain exceptions based on cultural industry. It should also be noted that NAFTA includes a set of reservations and exceptions, which may also impact upon satellite communications.

Regional African Satellite Communications Organisation (RASCOM)

The Regional African Satellite Communications Organisation (RASCOM) was founded on the growing awareness of the importance of telecommunications for economic development and productivity. Following several international consultative meetings, African leaders decided to pool their efforts so as to provide the continent with a satellite telecommunications infrastructure. A feasibility study conducted in some 50 African countries from 1987 to 1990 found that continental satellite telecommunications was the best system by which to meet Africa's telecommunications needs. The findings of this study were adopted by the African States in February 1991 in Abuja. In May 1992, the African States met in Abidjan and established the RASCOM.

RASCOM is an international intergovernmental organization run with commercial capital. Its mission is to design, implement, operate, and maintain the space segment of the African telecommunications satellite system. It aims to further African integration through the appropriate policies and technologies so as to provide an affordable infrastructure on a large scale to both urban and rural areas of the continent.

In order to achieve its objections, RASCOM relies on three basic organs: The Assembly of Parties, the Board of Directors, and the Executive Organ. The Assembly of Parties is made up of the 44 African signatory governments to the RASCOM Convention. It negotiates and promulgates strategic and policy-oriented regulations, and meets in ordinary sessions biannually. The Board of Directors has responsibility for the design, development, operation, and maintenance of RASCOM's space segments, as well as other activities that the Assembly of Parties authorizes it to undertake. It comprises representatives of the signatory governments and representatives of non-signatory shareholders. The Executive Organ of RASCOM implements the decisions of the Board and is responsible for the daily operations of RASCOM.

National Regulation of Satellite Communications

International law is implemented through the receipt of its obligations into the domestic legal system. Further, State practice through national acts or the passage of national legislation plays an important role in the development of customary international law relating to satellite communications. It is therefore important to remember that the national regulation of satellite communications must be in line with the State's relevant international legal obligations, and yet States' national legislation may shape the future landscape of the international regulation of satellite communications.

A point of significance is that international law is not self-executing in every State. The dualist legal systems in countries such as the United Kingdom and Canada require the adoption of international law through the passage of national legislation. These systems are different than those of countries such as France and the United States, where the international law is self-executing due to the monist legal systems in those countries.[12]

It is beyond the scope of this short chapter to discuss individual pieces of national domestic legislation that deals with satellite communications. In view of the increasing commercialization of satellite communications, there is an increasing need at the national level to implement international standards through domestic legal regulations that provide for corporate responsibility, technical qualifications and standards, ethical practices, liability insurance, and licensing regimes. Some States have chosen to enact the national space legislation,[13] which inevitably impacts upon satellite communications through its space segment. Other States have elected to pass specific legislation on a plethora of issues that acutely impact upon satellite communications, such as telecommunications, direct television broadcasting, spacecraft operations, and database management. Aside from these sector-specific pieces of legislation, other more peripheral regulations, such as copyrights and trademarks, labor laws, national security provisions, and engineering standards will also impact upon satellite communications.

Satellite Communications in Global Trade

The 1994 General Agreement on Tariffs and Trade (GATT) texts (GATT Secretariat 1994) included telecommunications as trade in services. Since then, regulation of international telecommunications and satellite communications has increasingly been recognized as governed by market access, competition, and deregulation. This is a marked departure from the public law protectionist framework that used to dominate. Through international and regional trade conventions, principles such as reciprocity, the Most Favored Nation (MFN) principle, national treatment, and transparency now operate to regulate the satellite communications industry. Satellite communications now play an instrumental role in trade-in services, foreign direct investment, and transnational trade policy, requiring transnational and international regulatory structures to cope with the expansion of the global commercial satellite communications industry and its attendant needs (Drahos and Joseph 1995).

This section will discuss the legal and trade frameworks in the context of the satellite communications industry.

The General Agreement on Trade and Tariffs (GATT) and the World Trade Organization (WTO)

The 1994 GATT texts, also known as the World Trade Organization (WTO) agreements, were the result of the 1986–1994 Uruguay Round negotiations, and were signed at the April 1994 ministerial meeting in Marrakesh.[14] Among the 60 documents, the documents of particular relevance to the satellite communications industry are

1. The Agreement Establishing the WTO (Marrakesh Agreement 1994)
2. The 1994 GATT
3. The 1994 Uruguay Round Protocol to GATT

4. The General Agreement on Trade in Services (GATS)
5. The Agreement on Trade-Related Aspects of Intellectual Property Rights, including Trade in Counterfeit Goods (TRIPs)
6. The Understanding on Rules and Procedures Governing the Settlement of Disputes

One of the most significant decisions of the Uruguay Rounds was the conclusion of the Agreement establishing the World Trade Organization, which establishes a single institutional framework encompassing the GATT, all agreements concluded under it, and the results of the Uruguay Rounds. The 1994 GATT amended the original 1947 GATT, which was focused mainly on goods, to include tertiary sector markets.

The WTO structure comprises a Ministerial Conference mandated to meet biannually, a General Council, a Dispute Settlement mechanism, and three Councils on services, goods, and intellectual property, respectively.

The General Agreement on Trade in Services (GATS)

The GATS mandates WTO Member States to liberalize and deregulate trade in services through continuing negotiations. Ministers returned to the round table for more negotiations on services pursuant to the November 2001 Declaration of the Fourth Ministerial Conference in Doha with follow-up ministerial discussions taking place in Cancún (2003), Hong Kong (2005), and Geneva (2004, 2006, and 2008).

Satellite communications are directly concerned by the GATS and two of its Annexes.

Article I of the GATS provides the scope and definition for the Agreement. Essentially, trade in services refers to cross-border or transnational supply, foreign consumption, commercial presence abroad, and presence of a natural person abroad. Satellite communications may concern one or more of these modalities of trades. It is significant to note, however, that satellite communications provided by government authorities in the domestic national context is excluded from the GATS.

The second part of GATS concerns the applicable obligations, principles, and rules in the trading context of satellite communications services. It is important to consider these principles in light of the relevant obligations under public international law and international space law that was discussed earlier in this chapter. Of particular significance to the satellite communications industry are the following obligations:

- The application of the Most Favored Nation treatment (Article II)
- Transparency in publishing all relevant measures (Article III)
- No forced disclosure of legitimate confidential information (Article III bis)
- Increasing participation of developing States (Article IV)
- Free economic integration (Article V)
- Free labor markets integration (Article V bis)
- Compliance of domestic regulations with GATS principles (Article VI)

- Recognition of authorization, licenses, and certification granted abroad (Article VII)
- Compliance of monopolies and exclusive service suppliers with the Most Favored Nation treatment (Article VIII)

In the context of satellite communications, Article VIII is of particular import. While WTO Member States are not prohibited from regulating monopolies, each member is obliged to abide by the MFN treatment, and must grant the same status to all other competitors, including foreign operators. This implied obligation to deregulate the domestic satellite communications market does not enjoy widespread observance today.

Certain exceptions are acceptable under GATS, including national security, and whatever measures Member States deem necessary to protect its security interests and act in compliance with its obligations under the Charter of the United Nations. This ensures that a Member State's application of GATS to the satellite communications industry will comply with its public international space law obligations.

Part IV of GATS focuses on progressive liberalization, mandating "successive rounds of negotiations," to reduce or eliminate adverse effects on trade in services as a means of providing effective market access. In the present context, this again points to an obligation on WTO Member States to progressively liberalize the satellite communications industry and to eradicate hurdles to market entry through negotiations.

Two Annexes to the GATS are specifically applicable to the satellite communications arena. The first is the Annex on Telecommunications, and the second is the Annex on Trade-Related Aspects of Intellectual Property Rights (TRIPs).

The Annex on Telecommunications

Satellite communications have progressively been integrated with commercial telecommunications services. The Annex on Telecommunications recognizes the integration of space-based communications with commercial telecommunications services, both as a tool for other activities, and as an activity in its own right. The Annex specifically focuses on the "transmission of signals by any electromagnetic means," restricting its application to public telecommunications as opposed to radio, television, and cable broadcasting. Telecommunications transport services is defined as the "real-time transmission of customer-supplied information between two or more points without any end-to-end change in the form or content of the customer's information," bringing satellite communications within the purview of the Annex. Satellite-based communications using the low Earth orbit systems, including voice- and data-only systems such as mobile telephony and broadband data transfer systems, respectively, fall entirely within the framework of the GATS provisions.

In practical terms, this means that States are obliged to be transparent in the promulgation of domestic tariffs, technical standards, terminal connections, and licensing criteria. Further, access to public network or services should be nondiscriminatory, meaning that information must be free and unrestricted

unless in clear contravention of security or confidentiality concerns. More particularly, special measures may be taken to increase the participation of developing States in programs of organizations such as the ITU, the United Nations Development Program (UNDP), and the World Bank. The harmonization of international standards to ensure global interoperability is also to be the linchpin of individual States' interaction with the ITU and other international organizations. It must be noted that to that end, a Decision on Negotiations on Basic Telecommunications have been undertaken by some States in order to promote progressive liberalization and deregulation of telecommunications markets and networks.

These standards and regulations aim to build a borderless global network of international telecommunications, as well as the liberalization of commercial communications networks and markets. While noteworthy strides have been taken by GATS in the area of value-added services in satellite communications and telecommunications, it must be noted that the GATS regulations do not affect "basic" telecommunication services. This means that competition is better provided for in value-added services rather than basic services, and that the deregulation of basic telecommunication services still has some way to go toward true liberalization. Moreover, the exceptions provided for by GATS in the face of security and confidentiality concerns may provide governments with a way out of allowing competition into the domestic telecommunications market. These issues raise interesting concerns as well for the operations of satellite communication networks. Pursuant to the agreements made in Doha, negotiations are ongoing to agree on removing regulatory barriers and opening domestic markets to foreign competition.

The Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPs)

The TRIPs Agreement is an elaborate standalone agreement. This section will address only those provisions of the Agreement that impact upon satellite communications. It will be noted in the preamble that the TRIPs Agreement recalls the conflicting interests of the Member States – action against counterfeit goods, as opposed to the special needs of "least-developed country Members in respect of maximum flexibility . . . to create a sound and viable technological base."

In the scope of this chapter, it must be noted that the basic provisions in relation to intellectual property rights are to be found in international agreements concluded prior to TRIPs and supplemented by it. These are:

- 1883 Paris Convention for the Protection of Industrial Property (as amended in 1967 and 1979)
- 1886 Berne Convention for the Protection of Literary and Artistic Work (as amended in 1971 and 1979)
- 1961 Rome Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organizations

- 1989 Washington Treaty on Intellectual Property in Respect of Integrated Circuits

TRIPs, together with the World Intellectual Property Organization (WIPO), oblige Member States to extend the National Treatment and the MFN Treatment to all other members.

Some points are of note about the application of TRIPs and these Conventions to satellite communications. First, Article 10 of TRIPs indicates the applicability of the 1971 Berne Convention for compilations of data without extending to the actual data itself. Such compilations receive protection for 50 years from the date of production of the work. Second, the 1967 Paris Convention applies for trademarks in respect of services, as per Article 16 of TRIPs. Third, in regard of industrial designs, the provisions of the 1883 Paris Convention are of significance in regard to situations of commercial exploitation of satellite communications. Fourth, section 7 of TRIPs enunciates an approach akin to the trade secret concept, by which "undisclosed information" is protected from unauthorized disclosure without consent.

Section 1 of TRIPs clarifies that the object of the Agreement is to achieve effective action, fairness and equity, decision on merits, judicial review, and Member States' freedom in the enforcement of their own domestic laws. Section 2 lists a full range of dispute settlement procedures and remedies, with provisional measures provided for in Section 3. These three sections provide for the enforcement of intellectual property rights under TRIPs, and are applicable to satellite communications where such activities fall under the TRIPs regime.

Dispute Settlement, Responsibility, and Liability Issues

What happens in the event that the regulatory framework does not prevent disputes from arising, or damage from occurring? This section will address issues related to the settlement of disputes arising from satellite communications, as well as the location and substantive obligations raised by legal principles related to responsibility and liability for such disputes and damage.

Dispute Settlement

In the specific case of satellite communications, there are several parties to consider when a dispute arises. These include:

- The satellite operator
- The satellite owner
- The satellite user
- The satellite manufacturer
- The satellite launcher
- The satellite insurer
- The regulatory agency of the country linked to the satellite
- The third party

Perhaps unsurprisingly, there are very few instances of claims filed in the case of damage caused by a satellite, whether domestically or internationally. This may be due to various reasons, but in particular because of the prevalence of cross-waivers of liability and exculpatory provisions, which act as bars to litigation.[15] Litigation or other adversarial claims take a long time to resolve, which is counterproductive for business. Further, a closed field such as satellite communications tend to attempt the preservation of good relationships between parties, which may be threatened by such adversarial processes.[16]

Aside from a dearth of publicly settled disputes in satellite communications, there is also to date no international case brought before open court. The only provision for a dispute settlement mechanism is found in the 1972 Liability Convention, which provides for the formation of a Claims Commission to determine compensation in the case of damage caused by a space object. It will be noted that there is no provision for a dispute settlement mechanism in the case that there is a dispute as to the responsibility or wrongdoing, or indeed, any merits issue relating to the claim. Many highly qualified publicists have, however, noted that arbitration appears to be the favored method of dispute settlement in space activity-related disputes.[17]

Responsibility and Liability

Satellite communications is considered first and foremost an activity in outer space. This means that, by virtue of Article VI of the Outer Space Treaty, States bear international responsibility for the acts of their nationals, whether these activities are by public organizations or private commercial entities. Article VII of the Outer Space Treaty stipulates that States are liable for damage caused by activities in outer space. The 1972 Liability Convention stipulates a detailed system for liability and claims for damage caused by space activities.

The international responsibility of a State may be invoked only if the State commits an internationally wrongful act, and if the act is attributable to it. In respect of international responsibility, States have generally ensured that they are in compliance with Article VI of the Outer Space Treaty by enacting domestic legislation requiring governmental authorization in the form of LICENSES for space activities by its nationals or on its territory. The structure provided by these domestic legislations and licensing schemes has several advantages for the State in question. First, it ensures that the relevant information about the space or satellite undertaking is provided to the regulatory authority. Second, it allows for the requirement of license renewal, which allows ongoing supervision of the space activity as required under Article VI of the Outer Space Treaty. Third, it provides the framework on which liability may be passed or shared between the government and entity involved. This last point is of particular significance in the case where private entities undertake commercial space activities, as is the case in many satellite communications projects. In many cases, the license which

grants permission for the space activity usually carries with it a requirement for the private entity to prove the undertaking of sufficient insurance in the case of damage caused.

When damage occurs due to the space activity, a dual system of international liability is envisaged by the 1971 Liability Convention and imposed upon launching States. Absolute liability is prescribed for damage caused by the space object on the surface of the Earth or to aircraft in flight. Exoneration is only provided for in the case where there was an act of gross negligence or omission on the part of the claimant State, unless the launching State itself was in violation of international law. On the other hand, fault liability is prescribed for damage caused elsewhere than on the surface of the Earth to a space object, or persons or property on board the space object in question. In both cases, damage refers to personal injury, or damage to property, whether natural or juridical.

Two points are of particular interest in the Liability Convention. First, Member States of international organizations are jointly and severally liable for damage caused if that organization has acceded to the rights and obligations of the Convention. Second, the Liability Convention addresses liability claims as between States – and not between private individuals or entities. Of particular significance is the fact that the Convention does not apply to nationals of the launching State, or to foreign nationals of the participating States. Only three categories of States are envisaged as in a position to make a claim: the State of nationality, the State of the territory on which damage occurred, and the State whose permanent residents have suffered the damage. The claim for compensation should be presented within a year after the damage or identification of the liable State, through diplomatic channels, another State, or the Secretary General of the UN. The exhaustion of local remedies is not necessary for a claim to be presented. Where claims do not produce a settlement, the Convention provides for the establishment of a Claims Commission.

In addition to the international liability outlined above in this section, another issue that may arise in relation to satellite communications is in relation to disputes involving private commercial entities and claims arising therefrom. Situations which may incur claims may include actions by such an entity in violation of international legal limitations (including the use of nuclear power), actions in violation of regional or national regulatory regimes (ranging from anticompetitive practices to labor laws and licensing requirements), and the acquisition, operations, modification, or termination of the enterprise (such as financing, shareholders' rights, dividends, bankruptcy, and so on).

Conclusion

This chapter aims to provide a concise overview of the many regulatory, legal, and trade issues that concern satellite communications. The overarching public service concerns, in light of growing commercialization and liberalization, has led to a complex labyrinth of rights and obligations on

entities involved in the satellite communications field. Owners, operators, end users, regulatory agencies, and public institutions must keep this dual nature of satellite communications in mind when considering the regulatory and legal matters that arise. The increasing role of satellite communications in the field of global trade and services exchange also compounds the intricacy of the applicable regulatory frameworks. A heartening development is the global inclination toward harmonization and coordination. A coherent and practicable legal, regulatory, and trade framework for satellite communications can only be put into place with mutual collaboration and innovative foresight.

Note All opinions and any errors in the text remain entirely those of the author and do not engage the United Nations or any other organizations with which the author is affiliated.

Cross-References

- ▶ International Committee on GNSS
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends
- ▶ Regulatory Process for Communications Satellite Frequency Allocations
- ▶ Satellite Communications and Space Telecommunications Frequencies
- ▶ Space Telecommunications Services and Applications
- ▶ Trends and Future of Satellite Communications

Notes

1. Article III, *Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies* entered into force 10 October 1967. 610 U.N.T.S. 205 (1967) [hereinafter "Outer Space Treaty"]. See text below.
2. Vast literature has been devoted to the subject of the delimitation of airspace and outer space. See among others Dodge, M.S., Sovereignty and the Delimitation of Airspace: A Philosophical and Historical Survey Supported by the Resources of the Andrew G. Haley Archive. J. Space Law. **35**(1), 5–36 (2009); K.M. Gorove, Delimitation of Outerspace and the Aerospace Object – Where is the Law. J. Space Law. **28**(1), 11–28 (2000); S. Gorove, Aerospace Object – Legal and Policy Issues for Air and Space Law. J. Space Law. **25**(2), 101–112 (1997).
3. It is beyond the scope of this chapter to provide an elaboration that would do justice to this field of the law. For an excellent overview, see F. Lyall and P.B. Larsen, *Space Law: A Treatise*. Ashgate, Surrey/Burlington, 2009). A legal commentary on the Outer Space Treaty can be found at S. Hobe, B. Schmidt-Tedd, K.-U. Schrogl (eds.), *The Cologne Commentary on Space Law*, vol. I. Carl Heymanns, Cologne, 2009.
4. United Nations Committee on the Peaceful Uses of Outer Space (COPUOS), website online at http://www.oosa.unvienna.org/oosa/COPUOS/copuos.html. Accessed 1 March 2012. The website also provides a directory with links to all the relevant international documents, treaties, agreements and declarations.

5. On the topic of customary international law, see the judgment of the International Court of Justice in the *North Sea Continental Shelf Cases*, (Federal Republic of Germany v. Denmark, Federal Republci of Germany v. the Netherlands), Judgment, (1951) ICJ Reports 116.
6. See for example N.M. Matte, *Space Activities and Emerging International Law* (Centre for Research of Air and Space Law, McGill University, Montreal, 1984).
7. See, among others, B. Cheng, Revisited: International responsibility, national activities and the appropriate State. J. Space Law, 1998, **26**, p. 7; F. von der Dunk, Liability versus responsibility in space law: Misconception or misconstruction?, *Proceedings of the 34th Colloquium on the Law of Outer Space*, **34**, 363 (1991); A. Kerrest, Remarks on responsibility and liability, *Proceedings of the 40th Colloquium on the Law of Outer Space*, **40**, 134 (1997); N. Jasentuliyana, R.S.K. Lee *Manual on Space Law* vol. 3 (Oceana, New York, 1979).
8. L. Haeck, Aspects juridiques de certaines utilisations de léspace. Ann. Air and Space Law, **XXI**, 65–104 (1996). See also B. Cheng, Military use of outer space: article IV of the 1967 outer space treaty revisited, in C.-J. Cheng, D.H. Kim (eds.), *The Utilization of the World's Air Space and Free Outer Space in the 21st Century*. (Kluwer Law International, The Hague, 2000), p. 305; G.M. Goh, Keeping the peace in outer space: a legal framework for the prohibition of the use of force, *Space Policy*, **20** 259 (2004); N. Tannenwald, Law versus power on the high frontier: the case for a rule-based regime for outer space. Yale J. Int. Law, **29**, 363 (2004).
9. See among others, S. Gorove, Interpreting Article II of the outer space treaty. Fordham Law Rev. **37**, 349 (1969); M. Williams, The principle of non-appropriation concerning resources of the moon and celestial bodies, *Proceedings of the 13th Colloquium on the Law of Outer Space*, **13**, 157 (1970).
10. See among others, T. Hartley, *The Foundations of European Union Law*, (Oxford University Press, 2010); A. Kaczorowska, *European Union Law*, (Routledge-Cavendish, 2010); D. Chalmers, others (eds.), *European Union Law: Cases and Materials*, 2nd edn. (Cambridge University Press, 2010); A. Dashwood, others, *Wyatt and Dashwood's European Union Law*, (Hart, 2012); B. Bishop, *European Union Law for International Business: An Introduction*, (Cambridge University Press, 2009).
11. For more information on NAFTA, see http://www.fas.usda.gov/itp/Policy/NAFTA/nafta.asp. Accessed 1 March 2011. Concise information on NAFTA can also be found online at http://www.naftaworks.org/. Accessed 1 March 2011.
12. See also generally T. Buergenthal, Self-Executing and non-self-executing treaties in national and international law. *Collected Courses of the Hague Academy of International Law*, 1994, p. 295.
13. A list of the national space legislations, and other pieces of domestic legislation that may have an impact on satellite communications, can be found at the Website of the United Nations Office for Outer Space Affairs, online at http://www.oosa.unvienna.org/oosa/SpaceLaw/national/def-delim/index.html. Accessed 1 March 2011. A concise backgrounder on legislation on the most active country in satellite communications, the United States of America, can be found online at http://www.unoosa.org/pdf/limited/c2/AC105_C2_2008_CRP09E.pdf. Accessed 1 March 2011.
14. For a summary of the texts resulting from the Uruguay Round of multilateral trade negotiations, see the Website of the World Trade Organization, online at http://www.wto.org/english/docs_e/legal_e/ursum_e.htm. Accessed 1 March 2011.
15. On the topic, see Salin at p. 39 and P. B. Larsen, Cross-waivers of liability, *Proceedings of the 35th Colloquium on the Law of Outer Space*, **35**, 91 (1992).
16. Fewer than twenty lawsuits, for example, have been filed in the United States of America, arguably one of the most active countries in satellite communications. For an overview, see P. Meredith, Spacecraft failure-related litigation in the US: Many failures, but few suits, *Proceedings of the 38th Colloquium on the Law of Outer Space*, **38**, 22 (1995).
17. See generally G. M. Goh, *Dispute Settlement in International Space Law: A Multi-Door Courthouse for Outer Space* (Martinus Nijhoff, Leiden, 2007).

References

Agreement Relating to the International Telecommunications Satellite Organization (INTELSAT) with annexes and Operating Agreement. Entered into force 12 Feb 1973. 23 UST 3813, 4091 (1973)

Arabsat, http://www.arabsat.com/. Accessed 1 March 2012

Association of South East Asian Nations (ASEAN), http://www.aseansec.org/. Accessed 1 March 2011

R. Bender, *Launching and Operating Satellites* (Utrecht Studies in Air and Space Law/Nijhoff, Dordrecht, 1998)

C.Q. Christol, The sharing of access and resources by states of varying capacities, in C.Q. Christol, *C.Q. Christol Law: Present, Past and Future* (Kluwer, Deventer, 1991), p. 289

R.R. Colino, The possible introduction of separate satellite systems: international satellite communications at the crossroad. Colum. J. Transnat'l Law **24**, 13 (1985)

Convention Establishing the European Telecommunications Satellite Organization (EUTELSAT). Entered into force 3 July 1985. Misc. No. 25, Cmnd, 9069 (1985)

Convention for the establishment of a European organisation for the exploitation of meteorological satellites (EUMETSAT). Entered into force 19 June 1986. Bundesgesetzblatt, Federal Republic of Germany (1987) Teil 11, p. 256

Convention of the European Space Agency and Rules of Procedure of the ESA Council. Entered into force 30 October 1980. ISSN 1010–5697

Convention on Registration of Objects Launched into Outer Space (Registration Convention). Entered into force 15 Sept 1976. 28 UST 695 (1976)

Convention on the International Liability for Damage Caused by Space Objects (Liability Convention). Entered into force 1 Sept 1972. 4 UST 2389 (1972)

Convention on the International Maritime Satellite Organization (INMARSAT). with annexes and Operating Agreement. Entered into force 16 July 1976. 31 UST 1, 135 (1976)

Developing the high-speed telecommunications links (electronic highways) for the Community's 1992 market, European Commission COM (88) 341. Brussels, 21 June 1988

P. Drahos, R.A. Joseph, Telecommunications and investment in the great supranational regulatory game. Telecommun. Policy **19**(8), 619–635 (1995)

E. Ducasse, L'Europe des Télécommunications par Satellite: entre Libéralisation et Coopération. (European Centre for Space Law, Paris, 1993)

EUMETSAT, http://www.eumetsat.int/. Accessed 1 March 2011

European Commission Decision (International Private Satellite Partners). 15 Dec 1994. OJEC L 354/75

European Council Resolution on the Development of the Common Market for Satellite Communication Services and Equipment. Entered into force 19 December 1991. (92/C8/01) (OJ C8/1) 14 Jan 1992

European Space Agency, Directorate of Telecommunications and Integrated Applications, http://www.esa.int/esaTE/index.html. Accessed 1 March 2011

EUTELSAT, http://www.eutelsat.com/. Accessed 1 March 2011

GATT Secretariat, *The Results of the Uruguay Round of Multilateral Trade Negotiations – the Legal Texts* (World Trade Organization, Geneva, 1994)

S. Hobe, B. Schmidt-Tedd, K.-U. Schrogl, G.M. Goh, *The Cologne Commentary on Space Law*, vol. I (Carl Heymanns, Cologne, 2009)

Inmarsat, see its website, http://www.inmarsat.com/. Accessed 1 March 2011

Intelsat, http://www.intelsat.com/. Accessed 1 March 2011

International Electrotechnical Commission (IEC), http://www.iec.ch/. Accessed 1 March 2011

International Organization for Standardization (ISO), http://www.iso.org/iso/home.html. Accessed 1 March 2011

International Telecommunications Union (ITU), http://www.itu.int/en/pages/default.aspx. Accessed 1 March 2011

Intersputnik, http://www.intersputnik.com/about/information/. Accessed 1 March 2011
R. Jakhu, International Regulations of Satellite Communications, pp. 78–101, in *Legal Aspects of Space Commercialization*, ed. by K. Tatsuzawa (CSP Japan, Tokyo, 1992)
R.S. Jakhu, Safeguarding the concept of public service and the global public interest in telecommunications. Singap. J. Int. Comp. Law **5**, 71–102 (2001)
N. Jasentuliyana, Recent developments in the United Nations activities relating to outer space. J. Space Law, **23** (2), p. 172 (1995)
G. Lafferranderie, *Les telecommunications par satellites, aspects juridiques*. (Editions Cujas, Paris, 1967)
F. Lyall, *Law and Space Telecommunications*. (Aldershot, Dartmouth, 1989)
F. Lyall, P.B. Larsen, *Space Law: A Treatise* (Ashgate, Surrey/ Burlington, 2009)
GATT Secretariat, *The Results of the Uruguay Round of Multilateral Trade Negotiations – the Legal Texts*. June 1994, Geneva
Marrakesh Agreement Establishing the World Trade Organization of April 15, 1994, *The Legal Texts: The Results of the Uruguay Round of Multilateral Trade Negotiations* 4 (1999), 1867 U.N.T.S. 154, 33 I.L.M. 1144 (1994) (hereinafter "WTO Agreement"), online at http://www.wto.org/english/docs_e/legal_e/04-wto.pdf. Accessed 1 March 2011
Ministerial Declaration on Trade in Information Technology Products. 18 Dec 1996, World Trade Organization. WT/MIN/(96)/DEC. 9–13 Dec 1996
North American Free Trade Agreement (NAFTA), http://www.fas.usda.gov/itp/Policy/NAFTA/nafta.asp. Accessed 1 March 2011
Regional African Satellite Communications Organization (RASCOM), http://www.rascom.org/. Accessed 1 March 2011
M.A. Rothblatt, The impact of international satellite communications law upon access to the geostationary orbit and the electromagnetic spectrum. Tex. Int. Law J. **16**, 207–244 (1981)
P.-A. Salin, *Satellite communications regulations in the early 21st century* (Utrecht Studies in Air and Space Law/Nijhoff, Dordrecht, 2000), pp. 11–12
I.H. Shefrin, The NAFTA: telecommunications in perspective. Telecommun. Policy **17**(1), 14–26 (1993)
M.L. Smith, *International Regulation of Satellite Communication* (Utrecht Studies in Air and Space Law, Nijhoff/Dordrecht, 1990)
Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies (Outer Space Treaty). Entered into force 10 Oct 1967. 610 UNTS 205 (1967)
United Nations Committee on the Peaceful Uses of Outer Space (COPUOS), http://www.oosa.unvienna.org/oosa/COPUOS/copuos.html. Accessed 1 March 2011
R.L. White, H.M. White Jr., *The Law and Regulation of International Space Communication* (Artech House, Boston/London, 1988)
World Trade Organization, http://www.wto.org/. Accessed 1 March 2011

Trends and Future of Satellite Communications

21

Joseph N. Pelton

Contents

Introduction 535
The Path Forward 536
Advanced Spacecraft Antenna Design 536
Improved Transmission Systems and Onboard Processing Systems 541
Improved Satellite Power Systems 542
More Effective and Reliable Spacecraft Design 545
Satellite Orbital Configurations, and Improved Spacecraft Orientation and Pointing 545
New Ground User Systems 546
Integrated Satellite and Terrestrial Services and New Market Demand 547
Telemetry, Tracking, Command, Monitoring, and Autonomous Operations 548
Future Trends for Markets and Regulatory Systems 549
Other Drivers and Opportunities 551
New Initiatives in Space Communicaions 551
Keys to the Future of Satellite Communications 552
Off-World Communications 552
Smart Satellites and Advanced Encoding 552
Integrated Satellite and Terrestrial Networks (i.e., The "Pelton Merge") 553
Advanced Launch Capabilities, In-Orbit Servicing, and Advanced Platforms 554
Space Safety and Orbital Debris 555
Cross-References 556
Notes 556
References 557

Abstract

Satellite Communications technologies have achieved remarkable breakthrough efficiencies and increases in performance in nearly a half century. These developments, however, have occurred in parallel with large gains in performance by

J.N. Pelton
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: joepelton@verizon.net

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_24, © Springer Science+Business Media New York 2013

other IT and telecommunications systems. Thus, these dramatic gains are not as apparent to the general populace as might have been the case if this explosion in performance had happened in isolation.

In many ways today's satellites are digital processors in the sky and specialized software defines how they perform and defines their communications capabilities. In fact, the innovations in satellite communications as well as the progression in all forms of telecommunications and computer processes have followed similar courses. In short, Moore's law that predicted a doubling of performance every 18 months has generally held true for all fields involving digital processing whether it be computing, communications, video games, or even digital entertainment systems. What had been past is thus likely to be prologue. It is reasonable to anticipate continuing gains in terms of overall processing power, digital communications, and "intelligent" space communication systems.

In short, there are remarkable new technologies still to be developed in terms of space-based satellite communications systems, more powerful processors, new encoding capabilities, and new user terminal capabilities that can make user systems more mobile, more versatile, more personally responsive, more powerful in terms of performance, and yet lower in cost (Pelton 2005; Also see Iida et al. 2003).

As the world national economies become more global and as all parts of the globe, the oceans, and the atmosphere are exploited by human enterprise, the need for effective wireless interconnection via terrestrial wireless and satellite communications will expand. Further the increased utilization of space systems – manned, unmanned, and planetary bodies – will evolve the need for improved space communications systems. Clearly foreseeable technologies suggest that several decades of continuing innovations are now possible. But technology will not be the only source of change for the satellite communications industry. Other drivers of change will include: (a) New service demands in both civilian and defense-related markets; (b) Restructuring of commercial satellite organizations through acquisition, merger, and regulatory change; (c) New allocations or reallocation of frequencies; (d) Convergence between the various satellite applications markets – both in terms of technology and structural integration; (e) Constraints in orbital configurations; (f) Orbital debris; and even (g) Growth of human activities in outer space may prove to be significant shapers of the growth of satellite systems in the next 20–30 years (Pelton 2005; Iida et al. 2003).

Keywords

Application specific integrated circuits • Bits per Hz • Codec • Encoding • Extremely high frequencies (EHF) • Integrated space and terrestrial systems • Millimeter wave frequencies • Modems • Onboard processing • Orbital congestion • Phased array antennas • Quantum computing • Rain attenuation • "Smart" antennas • Super high frequencies • Transmission efficiency • Turbo coding

Introduction

Today's communications satellites represent a very impressive gain in performance when compared to those first deployed almost a half century ago. Contemporary satellites solar arrays can generate well over 100 times more power, and advanced satellite antenna systems can deliver the equivalent of up to a 1,000 times usable bandwidth than that of the Early Bird Satellite – the world's first commercial satellite spacecraft. Deployable solar arrays have become larger in size, photovoltaic cells have improved in performance, and improvements in design have allowed the arrays to achieve maximum exposure to the sun. Battery systems have also improved greater in power density and longevity. Satellite antenna systems have evolved and improved in many different ways. These have included better pointing and focusing of radio frequency (RF) energy, multi-beam antennas, frequency reuse strategies, and improved large-scale antenna manufacturing. Overall, an ongoing series of technological improvements have increased the performance and lifetime of satellite systems in space and have made the user equipment on the ground easier to use, more accessible, and lower in cost (Pelton 2006).

The future suggests that many of these powerful trends will continue. There are, however, key challenges. One of these challenges that has been discussed in the previous section is that of integrating satellite communications systems with terrestrial wireless and broadband fiber and coaxial cable systems. The other challenge is to adapt satellite technology to a changing world. This could mean many things. It means effective use of satellite systems not only to communicate across the Planet but to points beyond throughout the Solar System (European Space Agency 2008). It could mean more integrated space applications so that user devices could provide not only voice, data, and video signals, but space navigation and location services, Earth imaging, weather and meteorological data, and other desired information on demand.[1]

The future will thus be shaped more by new service and market demands than new satellite technologies. Indeed regulatory shifts, industrial consolidation, constraints imposed by orbital debris, and even a change in financial and insurance markets could also dictate major shifts in the satellite communications industry.

Some might suggest that now satellite communications systems have gone from giant 30 m, multi-ton Earth stations to handheld transceiver devices there is little further room for further innovations. But history has proven forecasters wrong many times in the past. Forecasters such as Thomas Watson, Chairman of IBM, once thought that the world would need only a few dozens of computers to be used by elite scientists. Others thought that trains would travel no faster than a 100 km an hour (or about 60 miles/h) because of wind resistance. It was suggested in the nineteenth century that patent offices could be closed because all important inventions had already been registered. Demand for new services and new capabilities in human society constantly gives rise to new technologies which in turn generate new applications and the process regenerates itself again and again. Sometimes enthusiasm for technology overestimates future trends as

well. For the field of satellite communications projecting demand for new services can outweigh technological innovation in achieving accurate forecasts. Indeed predictions based on technological innovation cannot only often be wrong, but frequently greatly overstated (Schnaars 1989).

The Path Forward

The world of satellite communications is quite complex and technical, but the dynamic range of physical systems within which the new satellite networks are defined is remarkably small. Antenna systems focus "power" and "radio frequencies." To increase performance antennas must focus power more effectively or have access to more power or find a way to utilize available frequencies more effectively, either in higher frequency ranges or by more effective "reuse of the frequencies" or by both. These are the range of tools available to make satellite communications more effective. Of course one can invent more effective ways to send more "usable information" via a communications channel whether that be a fiber optic link, a terrestrial wireless link, or a satellite. The way forward essentially lies along one of these pathways. This section thus explores the future in terms of more effective antennas, improved power, more effective spacecraft design – including improved lifetime, reliability, and pointing systems, satellite orbital configurations, improved transmission capabilities, and finally improved user transceivers.

Advanced Spacecraft Antenna Design

The key to a satellite antenna's performance involves how well it can focus a RF beam toward the designed reception or "catchment area." This characteristic of the antenna is called antenna gain. A larger aperture antenna can create a narrower beam and thus there is less path loss due to the spreading of the signal between the satellite and Earth. Today's largest aperture satellite antennas with diameters on the order of 20 m or more can be used in conjunction with a multi-feed system to create many hundreds of highly focused and narrow beams that allow intensive frequency reuse. This type of advanced multi-beam, large aperture communications satellite antenna can be observed in such spacecraft as the Terrestar, Skyterra, and Inmarsat 4 satellites. This is because beams that are geographically separated from one another can use the same frequencies over and over again. The question naturally arises as to just how large can satellite communications space antennas grow without their being structural or cost barriers to their future expansion? (Pelton 1998; Iida and Suzuki 2001)

The answer to this question turns out rather complicated in that there are a variety of ways that one can create narrow beams for the purposes of reducing path loss and allowing intensive frequency reuse. These strategies can often be applied in parallel and thus are not necessarily mutually exclusive. The "best

design" for the space antennas of the future might thus involve a combination of these various approaches.

- *Use of Higher Frequency Antennas with a Smaller RF Wavelength*: If one moves to higher frequencies and thus smaller wavelengths the effective capability of a satellite antenna and its "gain" changes. With a higher frequency the antenna's aperture can be smaller. Since the spacecraft antennas are transmitting and receiving smaller wavelength signals, a smaller antenna can achieve the same result as another larger spacecraft antenna operating at a lower frequency and thus a longer RF wavelength. Indeed, since the "gain" of an antenna is inversely proportional to the square of wavelength, this makes a dramatic impact on the required size of the antenna needed to achieve the same effective performance. The largest satellite antennas today are for mobile satellite communications and the aperture size of this type of antenna is driven to a larger dimension because the antennas for down-linking signals to mobile users are in the range of 1,700 MHz to 2,500 MHz. These "lower radio frequencies" are used in part because the signals do not need to have direct line of sight to the satellite and can complete the link without necessarily having to "see" the user terminal that might be partially blocked by the top of a car or a telephone pole (Wakana 2003. Also see Hoeber).

  The down side of this consideration is that the antennas operating in these mobile satellite frequencies in the L band and UHF frequencies need to be larger to shape for transmission of the longer wavelengths. In contrast, the satellites that use higher frequencies such as the Ka band (30 and 20 GHz), for instance, require direct line of sight, but the spacecraft antenna needed to achieve similar gain at these frequencies allows an aperture size that is on the order of 100 times smaller. This is true since the frequencies are ten times higher and thus the wavelengths are ten times smaller. Since the aperture size is driven by the square of the wavelength the aperture size calculates to be 10^2 or 100 times smaller. The shortage of available bandwidth in the lower bands, however, is pushing satellite systems operation toward these higher frequencies for services other than mobile satellite communications – namely – fixed or broadcast satellite services.

  Unfortunately, there are problems with rain and other types of precipitation attenuation at these higher frequencies. This requires higher link margins translate into requirements for either higher power or larger and higher gain antennas. Also the electronics technology is much more demanding in terms of requiring the generation of quite precise, tiny wavelengths and quite high frequencies. This tends to drive costs higher.

  Although the antenna's aperture may be smaller, its contours must be much more exactly shaped to direct the smaller wavelength beam in exactly the correct way. Again this exacting contour also drives antenna fabrication costs higher. Further, the satellite must also be much more exactly pointed toward Earth so that the beams can be more precisely targeted. In short, while the higher frequencies and shorter wavelengths allow the spacecraft antennas to be smaller, the complications just cited can more than offset the advantages of the smaller aperture size and result in higher manufacturing costs. New electronics

technologies for the EHF bands, strategies to address precipitation attenuation, and exacting manufacturing techniques all combine to drive up the costs. Eventually, however, these difficulties of migration to utilize these new and higher frequency bands are overcome. As more and more satellites and ground systems are manufactured and deployed in these new bands the costs tend to go down. As the Ka-band systems are deployed the next horizon will be the next frontier, which are the frequency bands in the 48 and 38 GHz bands.

- *Phased Array Antenna Technology*: There is a new technology that is well suited to creation of larger-scale spacecraft antennas to improve satellite performance in the future. This technology is called phased array antenna systems. With this type of antenna an array of electronic components are combined to create "virtual" high performance and highly focused antenna beams. This type of antenna can support highly efficient multi-beam transmissions. The result is an antenna system that allows effective reuse of available RF frequencies many, many times. There are two ways that this technology can be used. One way is simply by directly extending today's satellite technology. This approach continues to deploy a very large high-gain conventional antenna reflector, but uses a phased array multi-beam feed system to generate a very large number of beams by reflecting off of a large parabolic reflector (see Fig. 21.1 below) (Iida and Pelton 2003).
- As shown above this could be a "tethered" large-scale reflector or "untethered." The feed system could use a phased array feed or horn feed array using more conventional technology.
- The second and more technically advanced way to proceed would be to design a "phased array antenna where the various electronic components actually 'electronically form' a beam for transmission or reception from a 'virtual reflector'." With this more advanced technology one can create a virtually shaped electronic beam of any shape and of large-scale dimension. Since the beam in this case is "virtually created" the effective size of the beam can grow quite large by going from something like a "six by six" phased array to a "twelve by twelve" or even a "hundred by hundred" phased array. A hundred by hundred phased array could, in theory, generate 10,000 different "pencil thin" beams and allow more than a 1,000-fold reuse of the same spectrum band available for satellite communications. The problem with this approach is that the technology is still at a very early stage and creation of a phased array antenna of this type is quite expensive.

The extension of phased array antenna technology could potentially go quite far. One concept is to deploy a large number of phase array components into space as a free-flying cluster. The micro-elements of a "virtual antenna reflector" could form a cluster of distributed "pico satellite array components" covering perhaps square kilometers and create beams that would create "pico cells" on Earth (see Fig. 21.2 below) (Iida and Pelton 2003, pp. 188–190).

With such a device the ability to reuse RF frequencies might climb to perhaps hundreds of thousands of times. At this time such concepts are just that. There would be many technical problems to be solved. These would include the issue of how to recollect all of the phase array components – perhaps with magnetic attraction – so as to minimize orbital space debris. There are also issues of

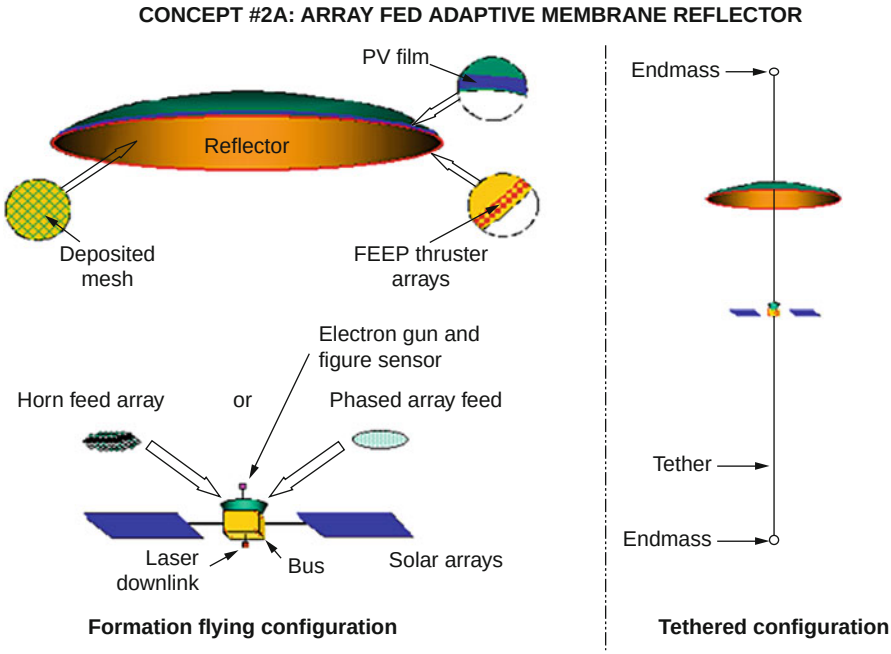


Fig. 21.1 Array fed adaptive membrane reflector – tethered or untethered (Graphic courtesy of Joseph Pelton and Ivan Bekey)

interference with other satellite systems and terrestrial communications systems that utilize the same frequencies.

There are other technologies that might be employed to boost future satellite communications performance as well. One future approach that might be used is what is called either "Scanning beam" or "Hopping beam" technology. This is a type of technology that allows a number of spot beams on an advanced satellite to work dynamically in the time domain using Time Division Multiple Access (TDMA) or Code Division Multiple Access (CDMA) multiplexing. In this type of antenna configuration, beams can be directed to different locations and bursts of data of varying durations (measured in milliseconds) can be sent to various destinations depending on levels of traffic demand. This technology allows streams of broadband traffic (voice, data, and video) to be sent in a burst to a particular location covered by a spot beam and then "hop" to the next location, and then "hop" to another location, and so on at extremely short time intervals. The advantage of this type of "hopping beam" is that the duration of the broadband blast of the digital data stream can be adjusted to the times of day as peak loads vary from time zone to time zone. Also if the satellite is operating in the higher frequency bands such as the Ka band (30 and 20 GHz), or in the future "Q" or "W" bands (48 and 38 GHz), then the dwell time of the data blast in a particular spot beam can be adjusted to compensate for a heavy rain story or other forms of rain attenuation. A dwell time of perhaps 10-ms for a data transmission in a particular beam might be doubled or even

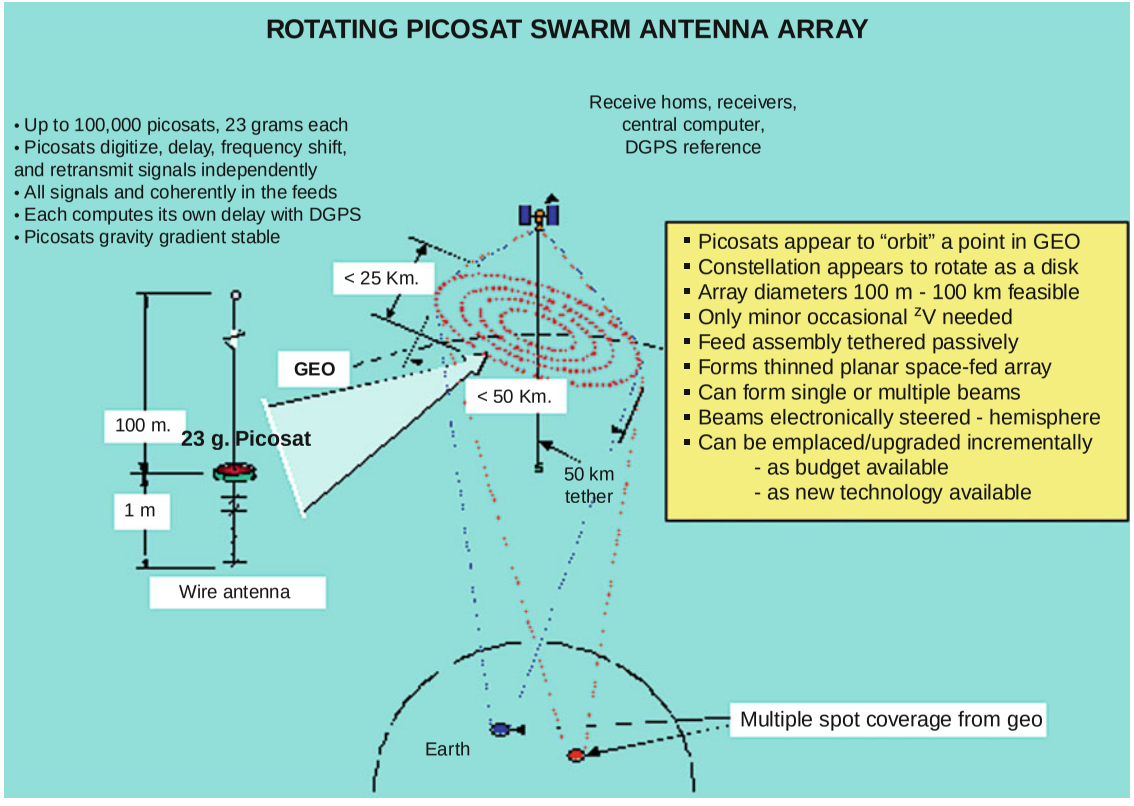


Fig. 21.2 Conceptual design of a Pico satellite array free-flying in space (Graphic courtesy of Joseph N. Pelton and Ivan Bekey)

tripled in duration in an area where there is heavy rain rate and thus severe rain attenuation. These types of concepts were tested in the US experimental communications satellite program undertaken by NASA known as the Advanced Communications Technology Satellite (ACTs).[2]

Regardless of the particular antenna designs of the future, the challenge will likely be to achieve the ability to generate more and more focused spot beams that can allow greater frequency reuse and lesser path loss by having transmission beams that "spread out" less between the satellite and Earth. The key will be to add intelligence along with the improved antenna systems to allow the antennas to work more efficiently. This could be increased "intelligence" that would be applied to "beam hopping" capabilities so that digital transmissions within particular beams could better be matched to actual overall demand for communications services, changing peak load requirements, atmospheric conditions, or other requirements.

Improved Transmission Systems and Onboard Processing Systems

The key to the efficient throughput of communication services hinges first and foremost on enhanced digital processing capabilities. The efficiency of satellite communications systems can be measured most directly in terms of digital throughput or simple in bits per Hertz. Techniques such as interconnection of geographically separate spot beams, polarization discrimination, and operation at higher frequency bands where wider spectrum bands are allocated to satellite communications allow a satellite to expand its amount of available bandwidth. Digital encoding (and in particular more efficient coders and decoders (Codecs)) and improved digital processing, modulation, and multiplexing techniques allow more bits to be delivered per Hertz. A few years ago a typical communication could provide 1 bit per Hz and could use a variety of techniques to reuse available spectrum by something like six to eight times. Today, through the use of advanced codecs communications satellites can derive something like 2.5 bits per Hz and can reuse spectrum by factors on the order of 20 times or more – particularly in the case of mobile satellite systems that operate at the lowest frequencies and thus have the smallest amount of useable spectrum. This trend will continue.

The trends for satellite communications and terrestrial fiber optic connections, however, will follow different patterns for specific technical reasons that separate how communications satellites and terrestrial cable operate. In the case of fiber optic networks, there are two specific advantages. The fiber optic networks operate, not in the RF frequencies, but in the much higher light wave region of the electromagnetic spectrum. There is an incredibly large amount of spectra available for communications in the optical wavelengths and (Dense Wave Division Multiplexing) currently operates in these bands by multiplexing signals only a quarter of a nanometer apart to achieve tremendous broadband throughput speeds. Further fiber optic cables have very little sources of external noise or interference and thus the quality of the signal remains very high over relatively long transmission distances. This means that fiber optic networks, due to broad available spectra

and low noise, do not have to be nearly as concerned with transmission efficiency as is the case with satellite networks. Further since satellites must interconnect uplinks and downlinks and increasingly interconnect different uplink and downlink beams, satellite systems need to use time-based multiplexing systems so there is time for digital processing associated with these switching operations. Fiber optic networks work almost exclusively with "wave division." This is in part because fiber transmissions do not require "time division" intervals for processing purposes. This is unlike the case with satellite networks that must cope with the problems of beam interconnection, time delay spoofing, etc. These differences in multiplexing techniques are important as these competing approaches tend to separate the fiber world and the satellite world. The terrestrial wireless world of telecommunications, however, has similar constraints, particularly with regard to beam interconnection and multiple reuse of the same frequencies over and over again. Thus, the ever increasing worldwide demand for mobile services and the similarity of satellite and terrestrial wireless networks (including their dependence on processing time for "beam" or "cell" interconnection) helps to tie the terrestrial wireless and the satellite networks together. This similar approach to time-based multiplexing, as used by satellites and terrestrial mobile, works to ensure that the future standards for interconnection of satellites, terrestrial wireless and fiber optic networks will keep compatible protocols for universal worldwide communications linkages. Clearly a challenge for the future will be to keep all forms of telecommunications transmission media to interconnect as "seamlessly and compatibly" as possible (Sachdev 2004).

Improved Satellite Power Systems

The story of improved satellite power systems has had several components. First of all solar arrays have greatly increased in size. Deployment systems for these arrays have grown more sophisticated to allow these very large-scale systems to unfold or otherwise be deployed from the compact configurations required to fit within the rocket fairings at launch. Secondly the "efficiency" of photovoltaic (PV) solar cells performance in terms of converting solar energy into the power required to generate RF signals has also increased. Solar cells have improved from amorphous silicon solar cells, to structured silicon to gallium arsenide cells. Further, the number of gates or junctures where solar energy is captured have increased and moved up into the ultraviolet part of the spectrum where the maximum amount of energy is obtainable. In short the efficiency of energy conversion has increased from around 7% to 30% and soon may become close to 50% in the most efficient systems. Not only have the solar cells improved in performance, but the ability of the satellites to display PV cells so as to achieve maximum solar radiation has also improved. The change from having solar cells mounted on the outside of cylinder-shaped satellites where the sun was "hidden" 40% of the time to three axis body stabilized arrays has made a large difference. Today the solar array can be constantly pointed toward the sun and even "angled" to get maximum illumination. Except when the satellite and

its solar arrays are in Earth eclipse, the arrays are now deployed with great efficiency to soak up the most solar power that is possible.

The third major trend has been the increase in battery performance. There has been an increase in the "energy density" of the battery systems and of their operational lifetimes. Batteries have increased in performance from Nickel-Cadium batteries to Polymer Lithium Ion cells. The lifetime of batteries and solar cell arrays – both of which deteriorate in performance in orbit – are critical components in the ability of the spacecraft to continue to operate over long periods of time in the harsh environment of outer space. The power of communications satellites have increased from around 100 W with the Early Bird (or Intelsat I satellite) to power systems that generate on the order of 15 kW. In short overall output power performance has increased by a factor of some 150 times. These power systems also have much longer lives. These systems, however, are also much more massive. When one seeks to measure the net performance for satellite power subsystems in terms of watt per kilogram per year in orbit the net increase is more on the order of 10–20 times.

The future seems to promise further improvements. So-called rainbow solar cells that have perhaps seven different PV junctions may be able to achieve a net efficiency over 60%. So-called quantum dot energy systems may also be able to achieve these levels of efficiencies. Both of these technologies are still in the laboratory, and thus not yet available to the commercial market (Ippolito and Pelton 2004).

The amount of solar power available in space is enormous if it could be effectively captured. One concept is to have some form of solar collector or concentrator to illuminate solar cells with higher intensity. There can be simple mirror surfaces associated with a solar array so that the cells can see the equivalent of two or three suns. There is, however, more advanced concepts that would deploy very lightweight film–covered collectors that could illuminate solar cells with great intensity. This technology is being developed in conjunction with plans to create solar power satellites so that solar cells could see the equivalent of hundreds of suns or even thousands of suns. This same technology, however, could be deployed in conjunction with a large-scale communications satellite platform. These large-scale and very low mass solar concentrators could be designed and deployed at much less cost than high performance solar cells (Fig. 21.3).[3]

There have been various studies to exam whether solar- and battery-powered satellites might "peak" in performance, in terms of maximum power and lifetime. These studies seek to compare large-scale solar arrays and battery systems versus other power sources such as nuclear energy, regenerative fuel cells, etc. These various studies have produced different optimization formulas and projected results, but many believe that for power systems above 20–40 kW nuclear energy or regenerative fuel cells might prove more cost-effective.

Certainly many space projects have used isotope-based SNAP generators for long-term and high-powered missions, but there is always the concern with safety (both at launch and reentry) with nuclear power sources. There are more ambitious longer-term research projects to develop nuclear reactors using thermal or ionized gases to provide propulsion systems. In these cases, nuclear power could also be

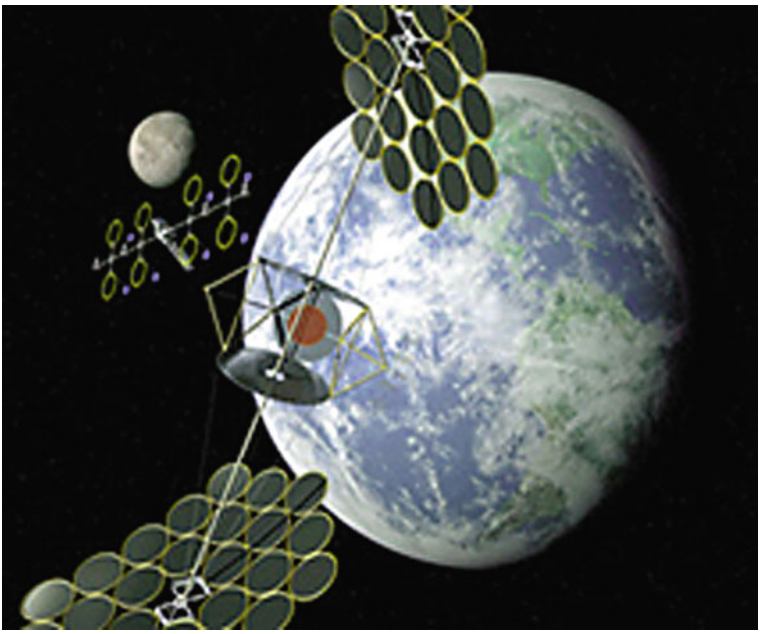


Fig. 21.3 Design concepts from Nasa for lightweight solar concentrators (Photo courtesy of NASA)

applied to generate electrical power to support a variety of missions in the space applications field. Such systems are still under development by most of the major space agencies. There are development issues related not only to mass-to-power performance ratios and usable lifetime, but most importantly nuclear generators in space give rise to concerns of the safety of radioactive materials and their safe disposal.

There are, however, certainly technologies with great potential that do not involve the risks of nuclear power. Two of the most promising are regenerative fuel cells that can produce reliable power for extended periods of time with high mass-to-power ratios and able to operate independent of sun exposure. Recent progress to use fuel cells for Earth-bound energy requirements for buildings and cars suggest that these may be effectively used in space within the next 10 years.

Even more immediate space-based energy systems involve systems that create further efficiencies with solar power. One innovation is that of very lightweight solar concentrators that be used not only with solar power satellites (SPS) but as a means to concentrate the equivalent of radiation of many suns on advanced solar cells or so-called quantum dots. Quantum dot technology which is still in the laboratory promises the ability to be perhaps three times more efficient in converting solar radiation into electrical power. A quantum dot can be defined as a special type of semiconductor whose "excitons" are confined from moving in any dimension. This constraint ends up giving quantum dot units properties that lay

between those of conventional semiconductors and the behavior of an individual molecule. Nanotechnology is seeking a wide range of applications for "quantum dots" in medical imaging, energy systems, and other fields.

Nearer term applications of quantum dot technology might involve using near infrared and perhaps higher frequency quantum dots as a retrofit to existing silicon solar cells to enhance performance.

More Effective and Reliable Spacecraft Design

The main drivers of satellite communications design and performance will likely be antenna design, increased power and digital encoding, and multiplexing techniques that will allow more throughput. Nevertheless, it is also important to have an effective design for the spacecraft to maintain reliable long-term operations and design features that can allow the satellite to be manufactured more quickly, at lower mass for lower cost launch or other sorts of improvements.

Satellite Orbital Configurations, and Improved Spacecraft Orientation and Pointing

The last few decades in satellite communications development can be largely summarized by the following trends. These have been to make spacecraft systems more powerful and to deploy larger aperture antenna systems with higher and higher gain and the ability to reuse available spectrum through the interconnection of cellular-like spot beams. These innovations plus the ability to launch and deploy larger spacecraft in space plus the gains that digital communications and digital compression techniques have brought to the satellite communications industry have dominated gains in satellite capacity and allowed many new applications to evolve, particularly in fixed, mobile, and broadcasting services. Although there have been various types of smaller satellites – variously described as micro-satellites, nano-satellites, pico-satellites, etc. – these spacecrafts despite innovative use of digital processing techniques represent far less than 1% of in-orbit capacity. The future expansion of satellite system capacity and the ability to provide consumers with low cost transceivers seem to require the deployment of even larger aperture, multi-beam antennas in space. The problem is that with current and projected launch capacity this may be quite difficult to achieve unless one starts to evolve toward one of several options. These would be: (a) the deployment of "parts" of systems that are either assembled in low Earth orbit and then flown into Geo orbit to create large-scale satellite platforms with antenna apertures in excess of 30 m; (b) the creation of "networked" antenna systems that fly in some form of formation or are cabled together to create a "virtual antenna system"; or (c) creation of a large-scale constellation operating the Extremely High Frequency (EHF) bands with a very large number of satellites in the constellation (i.e., similar to the proposed Teledesic satellite network). Some of these concepts were presented earlier in the section on spacecraft antenna.

In spacecraft design there are always various forms of trade-offs in terms of optimizing system capacity, lower cost and smaller user antennas, system lifetime, and strategies to cope with issues such as precipitation attenuation. One of the key constraints that would exist with very huge antennas with apertures in excess of 30 m is that the pointing accuracy of the space antenna would need to be very precise. Consideration would need to be taken of such aspects as thermal expansion of the antenna due to exposure to solar radiation, etc. In these design trade-off considerations, it is clear that much higher frequencies in the EHF (i.e., 30 GHz and above) allow the space antennas to be smaller in size, but on the other hand the problems of precipitation attenuation (especially rain) becomes much more severe and greatly complicates keeping user antennas on the ground small in size and low in cost.

New Ground User Systems

The predominant trend in all forms of digital communications, information processing, and digital entertainment (i.e., the ICE industries, or the information, communications and entertainment enterprises) have been to develop consumer-oriented, distributed systems that have moved closer and closer to the edge. This means that computers, entertainment systems, and communications devices have become smaller, more compact, more user-friendly, and lower in cost so that essentially consumers can own and operate these devices. Satellite communications have obviously followed this same trend. In the 1960s, computers were massive and highly expensive devices that required a team of experts to operate and highly centralized. Satellite earth stations were much the same. A typical Intelsat Standard A Earth Station of that era cost in excess of $10 million (US), required a team of 50 or more people to operate, and involved the precise pointing of a 30 m (93 ft) antenna that weighed many tons. Today there are handheld computers that have much greater computational power than the first digital computers. Likewise handheld satellite transceivers sold directly to the consumer can communicate directly to in-orbit satellites. Consumer video games, computers, and digital satellite phones cost in the hundreds of dollars and require a minimum of training for consumers to use these products.

This trend toward miniaturization of user-friendly consumer devices and digital instruments that can be acquired at low cost will undoubtedly continue. This means that the trend will continue toward “wearable devices” that are even more compact. Assuming that health-related issues involving RF radiation can be successfully overcome, it seems possible that embedded communications devices that are capable of linking to in-orbit satellites may represent the next tier of development.

There are clearly a number of technical challenges to overcome. There continues to be a rapid population growth with some 6.5 billion people on the Planet and, everyday more and more people are seeking access to broader band communications to support entertainment, video messaging, and voice and data

communications. Further more and more people are seeking broadband mobile connections either via terrestrial mobile services or satellite communications connections. This expanding demand for broadband mobile services requires new solutions. These can be in the form of more efficient digital compression techniques, new ways to reuse available frequencies more intensively within terrestrial or satellite wireless systems, migration to higher frequency spectrum bands, or some combination of these solutions. Smaller consumer devices with smaller antenna size make all of these efforts more challenging. Application Specific Integrated Circuit (ASIC) devices have contributed greatly to the ability to miniaturize consumer handheld communications devices, but new breakthroughs in quantum computing will be needed to achieve new levels of miniaturization. Such quantum computing–level breakthroughs may also help to reduce power requirements as well. This reduction of power would help not only with the problem of portable power supplies but also would assist with health-related concerns as well (Hagar 2007).

Integrated Satellite and Terrestrial Services and New Market Demand

Breakthroughs in quantum computing nano-material engineering, and the next generation of ASIC technology can all help to further the development of future communications devices used in terrestrial mobile systems as well as satellite-based networks, but the key may also be found in the integration of terrestrial cable and mobile systems with satellite networks in new and innovative ways. In short, the key way forward, given frequency constraints and the continuing expansion of broadband services would seem to involve using terrestrial wireless for localized services, fiber optic systems for intra-building risers, urban-wide area networks, and "trunking interconnections" and to use satellites for broadcast applications, for global and regional mobile services, and for services in rural, remote, and island areas.

The key to this type of integration of broadband services is seamless digital networking standards – most likely based in Internet Protocol standards. Terrestrial wireless, satellite communications networks and fiber optic cable systems all have their strengths, and seamless interconnectivity allows each of these transmission media to be optimized. The consolidation of IP-based protocols on a global basis will allow an intellectual platform upon which this integration of various types of transmission systems will be increasingly possible. Likewise, the Internet and TCP/IP will allow for all the space-based application satellites to be integrated as applications that be accessed on "smart phones" including, of course, "smart satellite phones" as well.

Originally satellite communication links, especially those to and from geosynchronous satellites, positioned almost a tenth of the way to the Moon at 35,870 km (or 22, 230 miles) above the Earth's surface encountered major problems with communications operating via TCP/IP because of transmission latency or delay. The original design of TCP/IP was to interconnect computers on the Internet.

Detected delays were considered to be the result of network congestion and the links automatically were time out to go into recover mode. Over time a number of changes have been made to optimize satellite operation using TCP/IP. There special IP over Satellite (IPoS) standards that employ "spoofing," reset of timers to accommodate satellite delay, and other techniques to allow satellite links to operate at increasingly high efficiencies. Also adjustments have been made to accommodate to IP Security (IPSec) and virtual private network (VPN) security measures when using satellite links. Part of the problem is that the Internet Engineering Task Force (IETF) and the International Telecommunication Union (ITU) have developed different methods to achieve efficient network interconnections. Part of the future challenge for the satellite industry is to be able to adapt to ITU and IETF standards and requirements rapidly, effectively, and economically.[4]

Telemetry, Tracking, Command, Monitoring, and Autonomous Operations

In coming years, it seems likely that satellite systems will continue to become larger and more sophisticated. As this evolution continues, communications satellites will tend to assume more complex roles with regard to onboard switching, onboard signal processing, and other functions that were once performed exclusively on the ground. This evolutionary movement, particularly toward onboard processing, will make satellites more capable and better able to cope with inter-beam connection, rain attenuation, and other advanced functions. This trend toward greater complexity in space will make the role of telemetry, tracking, command, and monitoring more difficult. Especially the software and engineering needed to achieve rapid fault detection when there might be something like a 1,000 beams and millions of possible inter-beam connections tends to become enormous. The most demanding role for the future in terms of designing advanced communications satellites will be the development of computer code to rapidly detect a particular fault in onboard link interconnections. For these future TTC&M roles, especially fault detection and detection of interference, there will be a trend to use artificial intelligence to assist with these functions. Likewise there will be increasing efforts to apply computer programs and artificial intelligence to control as many of the operations of the satellite over its 10–18 years lifetime through what is called "autonomous operation."

Communications satellites operate 24/7 throughout the year and are maintained by a team of human engineers and technicians to monitor all of these operations and to engage in rapid fault or interference detection which is increasingly uneconomic. One of the greatest technical challenges of the future for satellite system design thus will not only be the development of complex multi-beam antenna systems and onboard processing, but all of the new types of onboard intelligence that this will imply. Thus it seems that the communications satellites of the future will have largely automated tracking, telemetry, command and monitoring and fault detection systems and AI-based autonomous operations. This means that human-originated commands will be the exception and battery

discharges, activation of redundant receivers, shutdown of nonfunctioning switches, recording of billing information and hundreds of other operations that were once controlled by ground operators on a continuous basis will become activities that are increasingly assumed by onboard computers. Nevertheless, there will still be a need to monitor many satellite operations and maintain satellite operations centers. Autonomous operations and artificial intelligence will, however, serve to prevent these centers from growing exponentially in size and thus serving to make satellite operations uneconomic.[5]

Future Trends for Markets and Regulatory Systems

The future development of satellites is not dependent on technological development alone. Market demand and regulatory actions are likewise strong drivers of the satellite communications industry. Today's market trends suggest that new growth will be shaped by demand primarily in the three prime areas. These are: (a) entertainment and broadcasting services (i.e., video, high-definition and 3D digital television, plus broadcast radio that is also coupled with emergency vehicular services); (b) Mobile communications services in areas not well covered by terrestrial wireless services (i.e., air, maritime, and remote land areas); and (c) Gaps in communications services not well covered by fiber optic networks and terrestrial wireless broadband systems. Satellite broadband still remains key for Internet connections in many developing countries. New types of Internet-optimized satellite systems such as O3b (the Other Three Billion) are designed to bring broadband Internet and Voice over IP services to the parts of the world where effective terrestrial telephone and data networks are still lacking.

This is not to suggest that there may not be other market niches for satellite communications systems. Store-and-forward satellite systems that provide messaging services and business to business (B2B) services and are also often tied to space navigational services for trucks, trains, buses, and ships represent one such supplemental satellite service. One of the unknowns about future market demand relates to what might be called integrated space applications. In the world of "smart phone" applications, it seems increasingly likely that applications to support interactive navigation, immediate weather data updates, and remote sensing applications will evolve over time. Today remote system, Earth observation, meteorological and space navigation systems are delivered through separate space-based satellite systems and the delivery of information is largely through separate and "stove-piped" telecommunications networks. In the coming years, these systems can and likely will be integrated via Internet linkages to become just additional applications available via handheld "smart devices" or ultimately maybe via imbedded chip technology. All of these changes will serve to reshape the structure of the satellite communications industry. Mergers, acquisitions, and market integration via Internet applications will breakdown traditional industry divisions. This will mean at one level that companies will integrate across transmission technologies such as fiber, cable television, terrestrial wireless, and

satellites. On another level, companies in one space-related services such as satellite communications can and likely will be diversified into other space applications such as space navigation, space-based messaging, remote sensing, and real-time situational awareness.

In addition to market and service demand, regulations will also play a critical role. One of the most obvious areas will be that related to frequency allocations and the regulatory addition of new capabilities (i.e., new RF allocations that could interfere with satellites such as those provided for high altitude platform systems (HAPS). Such HAPS may be deployed over urban areas to provide television, remote sensing, or wireless broadband communications. Today one of the major constraints to the expansion of satellite communications services involves the lack of available spectrum and the lack of new orbital locations in GEO orbit for new satellite communications networks.

The demand for increasingly broadband services and expanded terrestrial wireless services to support mobile applications will only exacerbate this problem. This will push technology to develop larger multi-beam satellite systems to increase frequency reuse and improved ASIC (Application Specific Integrated Circuits) transceivers to operate more effectively in this spectrum-limited environment. Today the International Telecommunication Union (ITU) plays a key role in the allocation of spectrum and the establishing of recommendations to limit intersystem interference. The ITU is limited in role in many ways. Nations often establish through footnotes limitations on frequency spectrum allocations within their national borders. The ITU has no special enforcement powers such as fines or penalties for those who do not fully implement its recommendations. In the coming decades, this lack of enforcement powers and saturation of orbital locations and spectrum shortages could limit the growth of the satellite industry in a serious way.

Another serious regulatory issue involves the increasing spread of orbital debris. This spread is particularly troubling in the low earth orbit, but increasingly it is also a concern in the medium earth orbit and the GEO orbit as well. Controls designed through the UN Committee on the Peaceful Uses of Outer Space and ITU and national efforts at due diligence in these areas are starting to have some favorable impact, but the continued deployment of new systems may also make this issue a potential brake on the future development of the satellite communications industry and other space applications. With the latest high-definition tracking systems up to 30,000 objects the size of human fist can be tracked in Earth Orbit. The hope is that improved due diligence efforts to eliminate sources of orbital debris and decaying materials that drop from low earth orbit can bring the problem of "space junk" under control before cascading effects from orbital collisions can create a blizzard of hazardous materials in space.

There are currently efforts to create a global database to track the orbits of satellites starting with the Geo ring where the most communications satellites are currently in operation. To date Intelsat, SES Global, and Inmarsat have agreed to input data and Echostar, Telesat, and Eutelsat have indicated intent to participate (Chan and DalBello 2010).

Other Drivers and Opportunities

The future of satellite communications involves more than just spacecraft, launchers communications equipment, and new telecommunications equipment. The future success of the industry will require financial markets that will provide the capital to acquire new systems and technology. It will require a flexible insurance industry that can allow risks of various kinds posed by telecommunications markets and system failures to be overcome. It will also need an educational system that will produce the future engineers, financial analysts, regulators, and business people who can steer the industry forward.

Sometimes unsuccessful satellite ventures such as Orbcomm, Iridium, ICO, Globalstar, and Teledesic help to shape a more viable satellite market several decades later. This can be achieved by providing key data as to market demand, technological challenges to be overcome or better ways to achieve frequency efficiencies, interference reduction, or more appealing consumer products. Today launch systems remain one of the key obstacles to satellite communications economic efficiencies. Insurance and risk manager costs often represent over 10% of communications satellite system costs. The lack of new frequency allocations, orbital congestion, and orbital debris all serve as major barriers to industry growth. In the future environmental concerns could also serve as a brake to new system deployment. It is key to recognize these other factors that can pose as a key "limit to new growth" when considering the future of the satellite communications industry.

New Initiatives in Space Communicaions

Satellite communications technology and operations represent a still very rapidly evolving field. The telecommunications satellite industry is not only driven by a wide range of technologies to develop improved space systems and user equipment but also by competing communications systems – particularly fiber optic networking. Other technologies are also driving the curve. The future of satellite communications will thus also be shaped by artificial intelligence, robotics, terrestrial wireless systems, high altitude platform systems, quantum computing, laser communications, new multiplexing systems, and a host of other technologies. The future of humankind is ultimately based in space and thus space communications will one way or another continue to be a part of that future. If one considers the development of satellite communications in terms of throughput capabilities, power, antenna gain, lifetime, and costs it has improved by a factor of over a 1,000 times in the last half century and there is sufficient technology in the pipeline that it could improve another 1,000 times in another 50 years. The key to considering the future, however, is not by projecting the rate of technical innovation but to seek to understand basic market trends and to interpret what kinds of applications people will need to meet future societal, environmental, and economic needs in the decades ahead. Totally new markets could well evolve.

Keys to the Future of Satellite Communications

Predicting the future is always difficult but there are clear indicators as to trends and opportunities. The following "keys" are currently observable.

Off-World Communications

The most uniquely suited new market for satellite communications seems likely to be cislunar communications to Moon-based colonies as well as links to Mars or inhabited asteroids, artificial space colonies, or the satellites orbiting other planetary bodies in the solar system. The technology to support communications in support of today's scientific satellites already probing the Solar System will provide a head start in this direction. The most logical extension of capability in this respect would seem to be laser-based communications since the lack of an atmosphere on the Moon, asteroids, or the satellites of other planets makes light communications quite viable over the great distances of the Solar System. Laser beams are highly focused and thus are much less subject to path loss due to spreading. When the transmission distances are millions of miles the ability to reduce path loss is crucial.

Light attenuation within the Earth's atmosphere would suggest that laser communications would be directed toward Earth orbiting satellites perhaps in geosynchronous orbit. There have been other suggestions as to how to most efficiently establish such links. One suggestion is to have a solar sail–oriented satellite that could be positioned or "levitated" above one or both of the poles so that a signal could be relayed directly to anywhere in the Northern (or Southern latitudes). The various space agencies, and NASA in particular, have invested a good deal of research as well as state-of-the-art space communications hardware in intra-planetary and even interplanetary relays.

Smart Satellites and Advanced Encoding

The rapid advance in computer technology that was predicted by the so-called Moore's Law anticipated a doubling in capacity every 18 months and, in general, this exponential growth in computer performance has continued for some 30 years. The satellite communications industry has followed a similar curve of accelerated performance since the conversion from analog to digital satellite communications. In essence, today's communication satellites are, in fact, "software defined hardware." Although these are elaborate devices designed to operate in the harsh environment of space, the communications function is today essentially the result of very fast processing of digitally encoded information.

Advanced coding capabilities such as "turbo coding" allow the digital processing to be more and more efficient. Just a few years ago the most efficient communications satellites could process about 1 bit of information per 1 Hz of available spectrum. Today with more efficient coder/decoders (i.e., codecs) and more efficient modulator/

demodulators (i.e., modems) efficiencies of 2.4 bits per Hz of available spectrum are achievable. In future years, the efficiencies of digital satellite communications with onboard processing and improved codecs may be able to achieve efficiencies of 4 or 5 bits per Hz or even better throughput capabilities.

The limit on performance that derives from more efficient encoding of information is strongly determined by the amount of noise, interference, or more precisely in digital terms, the bit error rate, that prevents the use of more and more efficient codes. Particularly in the case of heavy rain and associated rain attenuation of signal clarity one might use 4 bit encoding such as QAM, but the use of even higher efficiency 8 bit or 16 bit encoding is not possible with today's communications satellites unless there is clear sky conditions and other forms of interference are not present. In the future, there is likely to be "onboard processing" that can restore up-linked signals to pristine quality and this ability to process signals onboard the satellite and again as they are received as down-linked transmissions can allow the transmission efficiency to rise.

It is significant to note that satellite systems and broadband wireless technologies are most keenly focused on processing and encoding efficiencies, as opposed to fiber optic networking systems. This is because in the fiber world, there is almost limitless spectrum with the use of dense wave division multiplexing (DWDM) and close to zero bit error rate (i.e., virtually no meaningful interference). Under these conditions, there are no strong incentives for the development of high efficiency encoding and increased bit/Hz throughput. With virtually unlimited available spectrum, there is little incentive for more and more efficient use of available bandwidth. In the world of satellites and broadband wireless communications, of course, the reverse situation holds true.

Integrated Satellite and Terrestrial Networks (i.e., The "Pelton Merge")

The world of communications today is driven toward integrating different types of transmission networks. The top objective is thus to combine "seamlessly" different transmission media to serve a wide range of consumer needs. The convenience of mobility drives demand for broadband wireless services, including communication satellites and in the future high altitude platform systems (HAPS) and Unattended Aerial Vehicles (UAVs) that provide platforms for communications and broadcast services. On the other hand fiber optic and coaxial cables can support very high efficiency and cost-effective services to fixed locations – particularly when heavy routes of traffic are involved. These divergent communications need ends in the need for operational and technical standards to allow these wire and wireless networks to link easily, at low cost, and with high quality.

The objective is easy to define and understand, but the ability to achieve "seamless" interconnection remains difficult. The difficulty stems from three key factors: (1) the world of fiber and coaxial cable multiplex signals in the wave division domain because of the vast amount of spectrum this provides for

broadband services, while wireless services, including satellites, operate in the time domain. This is because of the cellular type frequency reuse required to divide spectrum into small cells in order to allow multiple reuse of available spectrum. The interconnection of the signal used in these various cells requires digital processing time and thus multiplexing in the time domain. Interconnection is possible, but the differences add cost, complexity, and technical challenges in terms of smooth interconnection. (2) The second complicating factor between the world of fiber and satellites is related to the need to use high efficiency codecs to pack more information into available spectrum in the world of wireless services – as discussed above. Again the technical differences make "seamless interconnection" more difficult and thus more expensive. (3) The third factor comes from the world of Internet that was initially designed to run on terrestrially "wired" networks such as local area networks (LANS). In the original design, any significant delay in transmission was assumed to be network congestion rather than the transmission delay associated with geosynchronous satellites. Other aspects of the design such as IP Security (IP SEC) architecture were not designed to accommodate the transmission architecture of satellites. Initially where IP-based traffic was routed over satellites these problems tended to make satellite transmissions very inefficient and secure Virtual Private Networks (VPNs) difficult to establish because of the IP SEC process of "stripping off of header" information to preserve privacy led to confusion in satellite routing. Over time new standards related to IP over Satellite (IPoS) were developed and these issues of compatibility between satellite transmission and IP-based traffic were largely resolved, by resetting "clocks" to accommodate satellite transmission delay, and other measures to accommodate to IP SEC requirements associated with VPNs.

In the future, the design of compatible terrestrial and satellite networks to accommodate IP-based traffic will be a major challenge to make wire and wireless networks fully compatible, cost-effective, and of high quality. It is recognized by most planners today that a combination of wire and wireless systems will be needed to accommodate diverse networking and broadcasting needs and especially the demand for mobility. The achievement of the so-called Pelton Merge still remains a difficult-to-achieve goal that is dependent on improved interface standards optimized to meet the needs of IP-based traffic, low cost fiber optic cable networking design, and broadband mobile services provided via wireless and satellite systems.

Advanced Launch Capabilities, In-Orbit Servicing, and Advanced Platforms

For years the major brake on the development of the satellite communications industry has been with regard to cost and reliability of launch services. Although satellites have increased in capacity, power, lifetime, and cost efficiency, launch vehicle costs in terms of kilograms of payload to orbit have remained rather static. There is hope that breakthrough technology in terms of tether lift systems, advanced ion engines, and electrical propulsion or even space elevators might in

future years increase not only the reliability of launch systems, but dramatic decreases in cost. There have also been increased R&D on various systems to provide space tugs that might be able to move satellites from LEO to GEO orbit or perhaps even satellite service vehicles that could provide new batteries, positioning, and orientation fuel or even provide new satellite antennas or other components to in-orbit satellites. Such improved launch systems or space tugs could, of course, hasten the design and deployment of large-scale satellite platform or satellite clusters that could provide greatly expanded communications capability to orbit as well.[6]

Space Safety and Orbital Debris

For years, the focus of space safety systems and related technology was on human space flight. In recent years, however, there has been an increasing focus on how to make space activities in all its forms safer and the peaceful uses of outer space better guaranteed. The UN voluntary guidelines with regard to space debris are one important step in this direction. The recent UN COPUOS initiative on the "Sustainability of Space" is aimed at addressing this issue in broader terms that go beyond space debris and thus the effort is seeking ways to keep space from being militarized and finding new ways to ensure that access to space by all nations can better be assured.

Conclusion

The nature of satellite communications, in terms of technology, operations, institutions, finance, and markets, has changed greatly in the last 50 years. Satellite communications is by far the most successful satellite applications at least as measured by annual income. Part of the success of the satellite communications industry has been its constant evolution to find new markets. Today "submarkets" of satellite communications include fixed satellite services; direct to the home television satellite services; broadcast satellite services (for both television and audio); land, aeronautical, and maritime mobile satellite services; and store-and-forward or machine-to-machine satellite services. One could even suggest that satellite navigation services constitute a form of satellite communications. This constant expansion of such "submarkets" has served to expand global revenues in the field for nearly a half century.

The improvement of the technology has, of course, likewise expanded the success and the reach of satellite communications. So-called technology inversion that has allowed more powerful, capable, and longer-lived communications to work to ever smaller, more compact, and mobile user terminals at lower cost has expanded the reach of satellite communications. Instead of just a few large earth stations operating in the realm of satellite communications there are today literally millions of modest to quite small ground systems (including hand-held units) supporting a rich array of services. There is no reason to expect that technology innovative will slow as digital satellite communications systems

with increased levels of "intelligence" are deployed in future years. This chapter has examined a number of technologies of the future that might be anticipated.

Finally competition, innovation, and flexibility of institutional arrangements have allowed satellite communications markets and technology to grow and expand. This competitive environment has led to an active pattern of change. This change has seen–and continues to see – mergers, acquisitions, evolution of new entities, and new arrangements for financing and risk management in the field. This openly competitive environment has tended at times to give birth to new problems such as contending with orbital debris; interference between satellites, particularly in the geosynchronous (or Clarke) orbit; difficulties related to assignment of orbital locations; and the need for additional support to satellite communications in developing economies. One hopes that technical and operational innovation – and perhaps new economic and institutional arrangements – can provide required answers to such challenges in future decades.

Cross-References

- ▶ An Examination of the Governmental use of Military and Commercial Satellite Communications
- ▶ Economics and Financing of Communications Satellites
- ▶ Fixed Satellite Communications: Market Dynamics and Trends
- ▶ History of Satellite Communications
- ▶ Mobile Satellite Communications Markets: Dynamics and Trends
- ▶ Regulatory Process for Communications Satellite Frequency Allocations
- ▶ Satellite Antenna Systems Design and Implementation Around the World
- ▶ Satellite Communications and Space Telecommunications Frequencies
- ▶ Satellite Communications Antenna Concepts and Engineering
- ▶ Satellite Communications Modulation and Multiplexing
- ▶ Satellite Communications Regulatory, Legal and Trade Issues
- ▶ Satellite Communications Video Markets: Dynamics and Trends
- ▶ Satellite Earth Station Antenna Systems and System Design
- ▶ Satellite Orbits for Communications Satellites
- ▶ Satellite Radio Communications Fundamentals and Link Budgets
- ▶ Satellite Transmission, Reception and On-Board Processing Signaling and Switching
- ▶ Space Telecommunications Services and Applications
- ▶ Technical Challenges of Integration of Space and Terrestrial Systems

Notes

1. Op cit. J.N. Pelton, *Future Trends*, pp. 109–115
2. Advanced Communications Technology Satellite Overview, Nasa Goddard Research Center, acts.grc.nasa.gov and C.B. Cox, T.A. Coney, *Advanced Communications Technology Satellite*

(ACTS) Adaptive Rain Fade Protocol Performance, acts.grc.nasa.gov/docs/4thKa_Cox_Coney.pdf

3. NASA Research on Light Weight Solar Concentrators at Marshall Space Flight Center and Goddard Research Center, nmp.nasa.gov/ds1/tech/scarlet.html also see http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20090026381_2009013967.pdf
4. Tech Repubic White Paper, *IP Over Satellite*: *Optimization vs Acceleration End to End* (2010), whitepapers.techrepublic.com.com/abstract.aspx?docid
5. R. Sherwood, S. Chien, D. Tran, B. Cichy, R. Castano, A. Davies, G. Rabideau, *Next Generation Autonomous Operation on a Current Generation Satellite* trs-new.jpl.nasa.gov/dspace/bitstream/2014/7497/1/03-1398.pdf also see *JAXA Study A System Study for Satellite Operation and Control in Next Generation Systems* track.sfo.jaxa.jp/spaceops98/amp;/nfe_nakayama.html
6. Op cit, Iida and Pelton

References

J. Chan, R. DalBello, in *Data Sharing to Improve Close Approach Monitoring and Safety of Flight*, ed. by J.N. Pelton, R. Jakhu, Space Safety Regulations and Standards (Elsevier, New York, 2010)

European Space Agency, Interplanetary Communications Action 1: Requirements Analysis, Nov 2008, telecom.esa.int/telecom/www/object/index.cfm?fobjectid=29325

A. Hagar, in *Quantum Computing*. Stanford Encyclopedia of Philosophy (2007)

C. Hoeber, Annual satellite update, *AIAA Annual Review of Satellite Communications*

T. Iida, J.N. Pelton, in *The Next Thirty Years, ed. by T. Iida, J.N. Pelton, E. Ashford. Satellite Communications in the 21st Century: Trends and Technologies* (American Institute of Aeronauticcs and Astronautics, Reston, 2003), pp. 188–190

T. Iida, Y. Suzuki, Satellite communications R&D for the next 30 years, in *Proceedings of the 19th AIAA International Communications Satellite Systems Conference*, Toulouse, vol. 233, Apr 2001

T. Iida, J.N. Pelton, E. Ashford, *Satellite Communications in the 21st Century: Trends and Technologies* (American Institute of Aeronauticcs and Astronautics, Reston, 2003), pp. 1–15

L. Ippolito, J.N. Pelton, in Satellite Technology: *The Evolution of Satellite Systems and Fixed Satellite Services*, ed. by J.N. Pelton, R.J. Oslund, P. Marshall, Communications Satellites: Global Change Agents (LEA Associates, Mahwah, 2004)

J.N. Pelton, Telecommunications for the 21st Century. Sci. Am. 278 (4), 80–85 (1998)

J.N. Pelton, *Future Trends in Satellite Communication* (International Engineering Consortium, Chicago, 2005), pp. 1–19

J.N. Pelton, Basics of *Satellite Communications* 2nd edn. (International Enegineering Consortium, Chicago, 2006), p. 1–10

D.K. Sachdev, *Business Strategies for Satellite Systems* (Artech House, Norwood, 2004), pp. 1–15

S.P. Schnaars, MegaMistakes: *Forecasting and the Myth of Rapid Technological Change* (Free Press, New York, 1989)

H. Wakana, in *Mobile Service Update*, ed. by T. Iida, J.N. Pelton, E. Ashford. Satellite Communications in the 21st Century: Trends and Technologies (American Institute of Aeronauticcs and Astronautics, Reston, 2003)

Section 2

Satellite Precision Navigation and Timing Section

Introduction to Satellite Navigation Systems

22

Joseph N. Pelton and Sergio Camacho-Lara

Contents

Introduction 563
Technology of Satellite Navigation Systems 565
International Development of Satellite Navigational and Positioning Systems 566
Applications and Markets 567
International Coordination, Standards and Regulatory Issues 570
Cross-References 571
Notes 571
References 571

Abstract

The "youngest" of the major satellite applications is the field of satellite navigation. This field of measurement and ranging through the use of satellite positioning systems first started in the context of scientific research. This initial use of satellites was simply for positioning and location. These activities, that were first based on using Doppler frequency shifts as a satellite orbited above, were largely scientific and not strategic. These types of activities included geodetics (i.e., such as the measurement of continental drift over time) or the collection of scientific information and data from atmospheric land- or ocean-based sensors where the specific locations of the sensors were important.

The real strides in the development of satellite navigation however came when space systems were developed for the specific purpose of precise targeting of

J.N. Pelton (✉)
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: peltonjoe@gmail.com

S. Camacho-Lara
Centro Regional de Enseñanza de Ciencia y Tecnología del Espacio para América Latina y el Caribe (CRECTEALC), Luís Enrique Erro 1, Santa María Tonantzintla, Puebla, México
e-mail: sergio.camacho@inaoep.mx

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_10, © Springer Science+Business Media New York 2013

missiles and various other types of weapon systems. The Soviet Union/Russian Federation satellite navigation system, known as GLONASS, and the US-based Navstar system – with its Global Positioning Satellite (GPS) network – developed space-based systems that provided unprecedented capabilities to determine locations on the ground or oceans with great accuracy. Early systems such as Argos that relied on Doppler shift technology were only accurate within a precision of hundreds of meters. Today's advanced satellite navigation systems, however, are accurate for measurements that can be indicated with a precision only a few meters, and within centimeters if utilizing reference stations.

The fact that GPS and GLONASS satellite signals are freely available in space for all to use has spurred the development of low cost receivers for much more than strategic or military usage. Today, various types of civilian use of navigational and positioning satellites have become popular on a global basis. The latest development in application-specific integrated circuit (ASIC) chip technology has fueled the growth of applications based on the use of these precise space-based navigational and positioning systems. The availability of these highly capable but increasingly low cost chips – small enough that they could be included in handsets such as cell phones or included with the electronics available on various types of vehicles (i.e., cars, trucks, buses, trains, airplanes, ships, etc.) represented a real breakthrough. The development of specialized computer chips to perform satellite navigation calculations has increased the number of users of this technology from a relatively small population of several thousands to tens of millions. In the coming decade, the proliferation of international satellite navigation systems in orbit (i.e., US, Russian Federation, China, India, Europe, Japan) plus the continued development of ever lower cost of satellite navigation chips for receivers will continue the popular expansion of these increasingly "easy to use" and versatile systems.

Wide and easy access to low cost consumer receivers for satellite navigation services strongly suggests that this trend of expanded use will continue to surge. Thus, within a decade, use of these devices will increase to the hundreds of millions if not billions of people. Atomic clocks with incredible temporal precision today allow satellite navigation systems to be used by military organization, governmental geospatial scientists, and scientists to determine locations with great accuracy. But these applications are now just a small fraction of total usage. The deployment of the new international satellite navigational systems and smaller and lower cost ground receivers, will support an ever-increasing civilian consumer market for an ever-expanding range of everyday applications. These consumer applications include driving a vehicle to a desired location, safely sailing a boat, going mountain climbing or hiking without getting lost, or simply finding out where you are within a city.

The precise time keeping ability of today's satellite navigation systems also means that these spacecraft can also serve as a global timekeeper for computers and scientific experiments. The time stamp from a satellite navigation satellite can also be used not only for scientific or regulatory purposes but for other applications such as security and banking systems as well. These satellites are

also used to support the synchronization needs of communications satellites. In short, navigation and positioning satellites have also become in many ways the world's timekeeper.[1]

The official US discontinuation of the so-called "selective availability" feature for the Navstar satellite system has accelerated the use of the GPS network for highly sensitive applications such as assisting in the takeoff and landing of aircraft. The fact that selective availability could be reactivated has nevertheless been one of the concerns and invoked reasons why other countries have now proceeded to develop and launch their own satellite navigation systems. In short, when nations believe that certain space infrastructure represents a strategic asset, there is a strong motivation to deploy such a system rather than depending on other nations to own and operate such networks.

Keywords

Argos • Chinese BeiDou Navigation Satellite system • Chinese Compass Navigation Satellite • Doppler shift • European Galileo Navigation Satellite System • European Space Agency • Global Navigation Satellite System (GNSS) • Global Positioning Satellite System (GPS) • Global Navigation Satellite System (GLONASS) • Indian Regional Navigation Satellite System (IRNSS) • International Committee on GNSS (ICG) • Japan Aerospace Exploration Agency (JAXA) • Japanese Quasi-Zenith Satellite System • Japanese Regional Navigation Satellite System • National Aeronautical and Space Administration (NASA) • Navstar • Ranging • Selective availability • UNISPACE III

Introduction

Satellite navigation and positioning systems continue to evolve. The initial systems that relied on Doppler shift as a low-Earth orbit satellite moved across the sky was useful for such applications as collecting data from ground- or ocean-based sensors or tracking a ship as it made its way across an ocean. The development of extremely precise atomic clocks that were precise within picoseconds has allowed the development of a whole new type of space infrastructure that allows precise location and positioning that is orders of magnitude better than the first satellite systems. A network of over 20 such satellites in a well-designed constellation allows users of ground receivers to instantly calculate where one is located on the Earth's surface. The propagation time between various visible satellites can be calculated with great precision – especially when four or more satellites can be viewed at exactly the same time.

When such satellites using propagation time calculations were first developed in the 1980s, the intended purpose was strategic and, in fact, focused on the precise targeting of missiles and other types of weapons systems. This military application initially justified the billions of dollars that were involved in designing, manufacturing, and deploying scores of the GLONASS and GPS satellites. Once the satellites were up and transmitting signals back to Earth, people began envisioning a host of

new uses for these devices. Initially, these devices that were produced in low quantities were quite expensive, but as the demand for satellite navigation devices grew and their production increased to higher and higher levels, the costs began to fall. Today, the latest in computer chip technology has lowered the cost of production of receivers sharply and also let the devices to become very compact and even integrated with mobile phones simply by including an integrated circuit developed to process satellite navigation signals.

Initially, the satellite navigation systems designed for targeting were limited to the Soviet Union/Russian Federation GLONASS system and the US Navstar System that is composed of GPS spacecraft. Today, there are a growing number of satellites systems for navigation being developed and deployed by other space-faring nations. The navigation satellite systems in existence and those being developed are referred to collectively as Global Navigation Satellite Systems (GNSS). These satellite networks are being deployed in different types of orbits and some are oriented toward regional coverage as will be discussed in later chapters, but the combined use of these various satellite networks will allow for increasingly precise location and timing measurements. Most of these satellites will be providing their signals for free, although encrypted signals will sometimes be used to support strategic applications or to collect special value-added data for which clients will pay fees. From an economic perspective, the problem does become that of how can one system collect fees for its services when the same or similar information can be obtained from other countries' networks for free? In some cases the economic model is based on the collection of sales tax from the sale of receivers optimized for certain navigation satellites or the collection of value-added taxes for services associated with particular satellite navigation systems.

The proliferation of satellite systems for navigation and positioning has led to the recognition that some degree of coordination and standardization might be advisable to allow the various systems to work together efficiently. To this end, the Third United Nations Conference on the Exploration and Peaceful Utilization of Outer Space (Report of the Third United Nations Conference 1999) (UNISPACE III), held in 1999, called for the enhancement of, universal access to and compatibility of, space-based navigation and positioning systems. This call resulted in the development of a new international coordination entity, the International Committee on GNSS, (ICG) to exchange information on the various satellite navigational systems and sort out such issues as the optimum frequencies to use to not only support satellite navigation but also minimize interference among the satellite systems. The ICG, as discussed below, first met in 2006 and has now developed processes to coordinate the various satellite networks and to facilitate the development of units capable of receiving signals from multiple satellite networks.

Today, the global market for satellite positioning and navigation is the most rapidly growing satellite application in terms of percentage growth. Commercial applications of these satellites and their ground receivers, as estimated in statistics of the Organization of Economic and Commercial Development (OECD) has grown into a market that was nearly US$15 billion per annum as of year-end 2009 and may grow, in terms of total wholesale and retail sales of equipment and services – including new space systems to a global market of US$30 billion per

annum by year-end 2014 or 2015 (OECD 2012). Global statistics in this area are difficult to assemble, and others have put these numbers at a lower level, but all agree this is the most rapidly growing space-based business. Certainly, satellite navigation is now second only to satellite telecommunications in terms of a measurable commercial market.

Technology of Satellite Navigation Systems

The first satellite navigation system was Transit, a system deployed by the US military in the 1960s. Transit's operation was based on the Doppler effect: the satellites traveled on well-known paths and broadcast their signals on a well-known frequency. The received frequency will differ slightly from the broadcast frequency because of the movement of the satellite with respect to the receiver. By monitoring this frequency shift over a short time interval, the receiver can determine its location to one side or the other of the satellite, and several such measurements combined with a precise knowledge of the satellite's orbit can fix a particular position.

There are still satellites that use Doppler frequency shifts for position determination such as the polar-orbiting French Argos satellite positioning system. Argos transmitters' messages are recorded by a constellation of satellites carrying Argos instruments, and then relayed to dedicated processing centers. This system has been operational since 1978, and was initiated jointly by France and the USA. While Argos transmitters have become known to the public through their use in tracking yacht races, their primary mission remains the collection of data for the scientific community. By measuring temperature, pressure, humidity, and sea levels, Argos takes the pulse of the planet and its atmosphere.

The technology used in the major satellite navigation networks represented by the US GPS system, the Russian GLONASS system, the Chinese BeiDou/COMPASS system, the Indian Regional Navigation Satellite System, the European Galileo system, and the Japanese Quasi-Zenith Satellite System all hinge on ultraprecise atomic clocks and the relative transmission times from a network of satellites in a structured orbital constellation.

The satellite broadcasts a signal that contains orbital data (from which the position of the satellite can be calculated) and the precise time the signal was transmitted. The orbital data is transmitted in a data message that is superimposed on a code that serves as a timing reference. The satellite uses an atomic clock to maintain synchronization of all the satellites in the constellation. The receiver compares the time of broadcast encoded in the transmission with the time of reception measured by an internal clock, thereby measuring the time-of-flight to the satellite. Several such measurements can be made at the same time to different satellites, allowing a continual fix to be generated in real time.

Each distance measurement, regardless of the system being used, places the receiver on a spherical shell at the measured distance from the broadcaster. By taking several such measurements and then looking for a point where they

meet, a fix is generated. However, in the case of fast-moving receivers, the position of the signal moves as signals are received from several satellites. The basic computation thus attempts to find the shortest directed line tangent to three to four oblate spherical shells centered on the satellites emitting the signals. Satellite navigation receivers reduce errors by using combinations of signals from multiple satellites and multiple correlators, and then using techniques such as Kalman filtering to combine the noisy, partial, and constantly changing data into a single estimate for position, time, and velocity. This "estimate," however, is actually quite precise.

The intersecting "virtual spheres" represented by the transmission time from the network of satellites serve to create a positional location on the Earth. The "virtual spheres" from three satellites can identify a location but four or more satellites can determine a unique location even more accurately. More information on atomic clocks and the precision of today's navigation and timing satellites is included in later chapters in this section.

These various satellite networks represent very complex systems requiring an investment of billions of dollars to design, build, and launch a large constellation of very sophisticated satellites but the solid trigonometry involved is just a matter of rapid calculations called trilateralization. This process is depicted in Fig. 22.1 below where the location "A" on the Earth's surface has been identified.

There are two main technical problems with the satellite navigation systems today and they both hinge on the need to transmit radio frequency signals from the satellite to the ground receivers. The first problem is that interference from other satellites or terrestrial sources can affect performance. Specifically, jamming could create a major problem to these satellites as they do not emit high gain signals because the signal is very broadly spread across the Earth. The second problem is that a variety of frequencies are used in satellite navigation systems. This started with the GPS and GLONASS systems using different frequencies. This makes receiving units that are engineered to receive signals from multiple systems more expensive. The International Committee on Global Navigation Satellite Systems (ICG) and its Working Group A (WG-A) on Interoperability and Compatibility are, among other things, addressing these radio frequency issues.

As a final note it should be noted that satellites equipped with cesium atomic clocks can provide with great accuracy not only location, three-dimensional positioning, and navigational velocity, but also time synchronization. The cesium clocks, which are remarkably accurate "scientific instruments," are one of the major contributors to the cost of navigational satellites.

International Development of Satellite Navigational and Positioning Systems

The Navstar GPS Satellite system and the Russian GLONASS satellite system, now almost completely repopulated, already provide two comprehensive global navigation and position systems that have enough satellites to provide the capability to

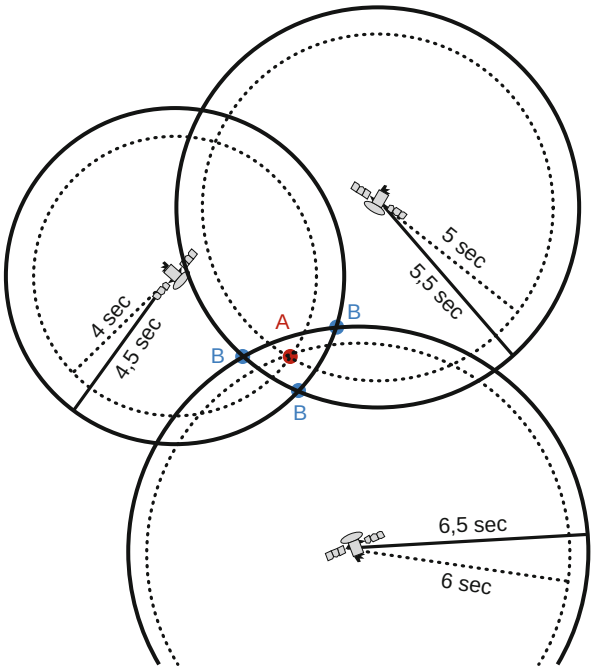


Fig. 22.1 Satellite-based determination of position via trigonometric trilateral calculation

determine a precise location over the entire surface of the world as well as incredibly precise timekeeping. The latest generation of GPS Block III satellites, and improved ground control system, represent a $5.5 billion investment that will make the system more resistant to interference, better able to support precision applications such as aircraft takeoffs and landings and even more precise time synchronization (Newman 2012). The GLONASS system had in the 1990s and the early 2000s lost a number of its original satellites, but a new generation of satellites has now been launched and the network has some two dozen satellites, fully functional spacecraft in the restored network.[2]

Despite the complete global coverage of these two systems, a rather amazing array of additional satellite navigation and positioning satellites, costing many billions of dollars have either been launched or new spacecraft designed for this purpose are in final aspects of design, engineering, manufacture, or launch. These additional systems from China, Japan, India, and Europe plus the US and Russian systems are discussed in detail in later chapters.

Applications and Markets

The ability to locate, navigate, and position using satellite technology – essentially in real time – has turned out to be a very valuable capability. Initial applications in

such field as geodetics – measurement of motion of the Earth's tectonic plates – or being able to know the location of sensors sending up data about ocean or weather conditions, were very much oriented toward scientific, geological, or ecological conditions. The possibility of targeting missiles using space systems led the governments of both the USA and Soviet Union to make the multibillion investment in global constellations of positioning and navigational satellite systems. It also led the USA to design the Navstar system to include a "selective availability" capability that would in times of warfare prevent anyone without specific codes to accurately target weapon systems using this constellation. Selective availability, in essence, was a degradation of the signal that was received by a user without the secret codes. President Clinton, however, signed an Executive Order to confirm formally a pledge not to use selective availability. This Executive Order was in part based on the fact that increasingly vital applications involving aviation traffic and airport safety depended on GPS accuracy and in part based on the fact that new capabilities involving calculations based on known locations (reference stations) could actually defeat the "selective availability" of spurious information in any event.

The scientific and defense-related applications of satellite navigation systems continue to increase but the civil applications for industrial and consumer services have skyrocketed in recent years. The following listing of ways that these satellite networks can be used today is not complete, but is nevertheless indicative of the uses that are now being made. Specialized data bases have been developed for very specific applications that consumers are now purchasing as "online applications."

- Cross country skiing
- Driving instructions to unfamiliar or unknown addresses
- Emergency recovery and restoration operations – repair crews, emergency vehicles, etc.
- Fire-fighting assistance in coping with unfamiliar terrain and with forest fires
- Golf course layouts and specific dimension of course hazards
- Hazardous marsh, fen, and difficult terrain identification
- Hiking and mountain and rock climbing information
- Location determination for automotive breakdowns for tow trucks
- Location reference points and "icemarks" for explorers in arctic regions and on glaciers
- Mapmaking
- Ocean hazards and location of pollution zones
- Oil spill containment and coping with pollutants, red tide, etc.
- Police, Narcotics enforcement, and illegal immigrant enforcement activities
- Precision agriculture
- Rescue operations in mountains, wilderness, arctic regions, etc.
- Routing of trucks
- Sailing and boating locations on lakes, rivers, canals, and on ocean voyages
- Synchronization services for computer networks, communications, and energy systems
- Security systems that are based on codes and real-time positioning of users

- Surveys and establishment of legal boundaries
- Timekeeping for scientific experiments
- Transit system maps and locations of stations for virtually every city in the world
- Transportation design, construction, repair, and maintenance
- Utility repair crews
- Urban planning

The navigation, position, and timing satellites can be used for knowing where you are, where you are going, and the best way to get there. These satellites can also be used not only for map making and surveying and the like, but also to know the exact time and to synchronize computer networks or communications systems. The real importance of this indicative list is that almost any activity is enhanced by flexible and reliable broadband telecommunications and the instant ability to know one's location and the ability to navigate to where one needs to go safely and effectively. The gradual development of improved software for satellite positioning and navigation systems adds precise knowledge as to the location of roads, highways, railroads, utilities, canals, rivers, bridges, shoals, reefs, cliffs, marshes, and the layout of towns, cities, and even amusement parks, golf courses, colleges, and industrial and governmental complexes.

The constantly expanding utility of satellite navigational systems and the software that provides the location and current timing of practically everything has made these space systems very valuable. First responders and emergency aid workers; military troops; all forms of transportation services; workers involved with deployment, repair, and maintenance of utility systems; urban planners and map-makers, and indeed civilian consumers going about their daily lives – either at work or at play – increasingly find that satellite navigation is a sine qua non. Each day the number of people who find satellite navigation almost as important as mobile communications continues to grow. As these functions are wedded together in cell phones, the applications only increase.[3]

Satellite navigation, as an identifiable economic activity, has thus grown rapidly in recent years and continues to lead growth in the space industry in terms of a percentage increase in revenues. The remarkable growth of space navigation services all trace back to the ability to make the technology available to global consumers at a very low cost and in a "demystified way."

The development of specialized computer chips that have embedded in them the ability to perform the specialized calculations needed to allow position location on a near instantaneous basis was a key step. The next important step in the satellite navigation, positioning and timing industry was the inclusion of this technology in consumer devices and particular in cell phones. Today, third and fourth generation cell phones come with embedded chips that allow consumers to know instantly where they are and to be synchronized to the Internet. Most recently, it has been the development of applications (i.e., particularly GPS-related "apps" on the latest generation of cell phones) that can be downloaded to mobile devices that has contributed to an explosion of satellite navigation and positioning uses not only in the USA but all across the globe (Fig. 22.2).

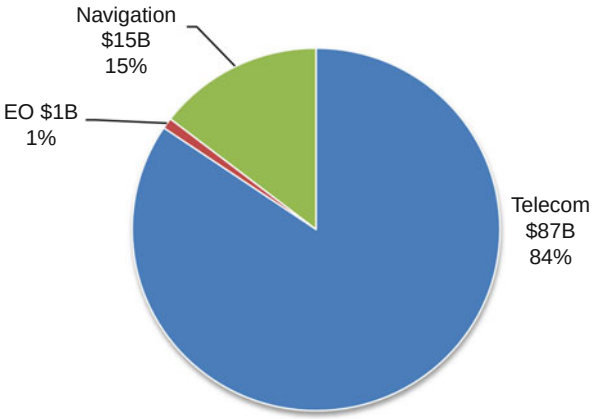


Fig. 22.2 Distribution of satellite services revenues (in US$ billions) between the three main types of satellite applications in 2009 (Source: *Op cit*, The Space Economy, OECD 2011)

International Coordination, Standards and Regulatory Issues

At the UNISPACE III conference, one of the clear trends in space development that emerged was the rapid proliferation of space navigation and positioning systems. The work of the "Action Team" established by the United Nations Committee on the Peaceful uses of Outer Space resulted in what is now known as the International Committee on GNSS (ICG). The ICG was formed through the voluntary participation of all the GNSS operators in order to create a useful forum for technical coordination and interference reduction in the operation of satellite navigation systems. The ICG is also now assisting in formulating internationally accepted satellite navigation and positioning standards and to stimulate key new civil applications around the globe.[4] The functioning of the ICG and the specific activities of its Working Group A and its three other working groups are detailed in ▶ Chap. 24, "International Committee on GNSS."

Conclusion

One of the great commercial successes in the global space economy that has emerged primarily in the last decade is that of the dual-use global satellite navigation, timing and positioning services. At first, military and strategic reasons provided the financial and economic justification to develop the technology and to launch very expensive and complex spacecraft with onboard super accurate cesium atomic clocks. These clocks are essential to the capability to determine locations and times, and thus speeds, with a very high accuracy. This new space technology allowed a large number of new applications to emerge where determining one's location and transit vectors can be used in a remarkably diverse number of ways. In the meantime, advances in software and hardware technology have combined with the improved signals from the GNSS to provide

more accurate measurements at lower costs. One can expect that, as the GNSS operators make the satellite systems interoperable, a single receiver will be able to process the signal from the satellites of any of the GNSS. This will result in even lower costs for receiver equipment and the possibility of viewing ten or more satellites at the same time, resulting in more accurate measurements in urban and mountainous environments.

The economic applications related to scientific, military, government, and particularly industrial and consumer-based applications continue to expand, and exponentially so. The current size of the satellite navigation, positioning and time industry in economic terms is over US$15 billion and is expected to grow to perhaps $30 billion in 2–3 years. By making other economic areas more efficient and by creating new ones, the economic impact of GNSS will be in the hundreds of billions of dollars (US).

Cross-References

- ▶ Current and Future GNSS and Their Augmentation Systems
- ▶ Global Navigation Satellite Systems: Orbital Parameters, Time and Space Reference Systems and Signal Structures
- ▶ International Committee on GNSS

Notes

1. Cesium clocks and global timekeeping. http://www.rfcafe.com/references/general/atomic-clocks.htm
2. Global Navigational Satellite System (GLONASS). Access Science. http://accessscience.com/content/Global-Navigation-Satellite-System-(GLONASS)/YB980695
3. E. Biba, Inside the GPS revolution: 10 applications that make the most of location. *Wired Magazine*. http://www.wired.com/gadgets/wireless/magazine/17-02/lp_10coolapps?currentPage=all. Also see GPS Applications http://www.gps.gov/applications/
4. International Committee on the GNSS. www.oosa.unvienna.org/oosa/SAP/gnss/icg.html

References

OECD, The Space Economy at a Glance 2011 (2012). http://www.oecd.org/dataoecd/63/5/48301203.pdf

J. Newman, What the next generation of GPS satellites will mean for you. PC World (March 6, 2012). http://www.pcworld.com/article/246128/what_the_nextgen_gps_satellite_upgrade_means_for_you.html

Report of the Third United Nations Conference on the Exploration and Peaceful Uses of Outer Space, Vienna, 19–30 July 1999 (United Nations Publication, Sales No. E.00.I.3), Chap. I, Resolution 1

Global Navigation Satellite Systems: Orbital Parameters, Time and Space Reference Systems and Signal Structures

23

Rogerio Enríquez-Caldera

Contents

Introduction 574
GNSS Orbits 575
MEO Orbits 576
Reference Systems 576
Ephemeris 578
Ground Tracks 578
Polar Plots 580
Point Positioning 580
Differential GNSS 582
Kinematic and Real Time Kinematic Systems 585
GNSS Augmentation Systems 585
International Atomic Time (TAI) 586
Signal Structure 588
Carrier Frequencies 589
Data Stream and Messages 591
Multiple Access and Pseudorandom Codes 593
Modulation Techniques 595
Pseudo-Range 596
Sources of Error and Error Budgets 598
Cross-References 600
References 600

Abstract

Global navigation satellite systems (GNSS) are part of the most complex modern space systems humankind has created, and therefore, their orbits, orbital

R. Enríquez-Caldera
Centre for Space Science and Technology Education for Latin America and the Caribbean, México Campus (CRECTEALC), National Institute of Astrophysics, Optics and Electronics (INAOE), Coordinación de Electrónica, Apdo. Postal 51 y 216, Tonantzintla, Puebla, 72000, Mexico
e-mail: rogerio@inaoep.mx

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_93, © Springer Science+Business Media New York 2013

parameters, and their two main terrestrial mappings are firstly described. Different frames of space-time reference systems are treated as part of such descriptions.

Communication systems engineering are important sections to allow for a GNSS precise fix positioning. All signal structures and data streams are treated for a clear understanding permitting the reader to see how ranging is obtained from space.

Theoretical and practical error budgets are considered to give the reader a perception of limitations during scientific and/or technical user campaigns or for simple common life enjoyment.

Keywords

MEO orbit • Reference systems • International atomic time (TAI) • Signal structure • Frequency issues • Relativistic time dilation • Modulation techniques • Pseudorandom codes • Carrier frequencies • C/A and P(Y) codes • Data stream • Pseudo-range • Accumulated delta range • Sources of error and error budgets • Differential GNSS • Kinematic and real time kinematic systems.

Introduction

Global navigation satellite systems are comprised of tens of satellites placed in Middle Earth Orbits (MEO), somewhere around 20,000 kms from the Earth. Scientific, technological, and technical aspects of modern human capabilities were weaved into GNSS as in no other systems and are finding applications in practical everyday life. The use of these space systems, originally developed for military purposes is increasingly contributing to the well-being of humankind. Space topics, such as aeronautics, propulsion, tracking, communications, handheld multiuser receptors, signal processing, electronic and inertial sensors, estimation theory, navigation, mapping, electromagnetic propagation, relativistic physics, reference systems, timing, and many other disciplines, constitute the elements of these GNSS, a wonderful human creation which is closest to a ubiquitous system.

Keeping in mind the above, the reader would, first of all, appreciate the complexity involved in sending a constellation of spacecraft to specific orbits, envisioning in perspective the trajectories followed by each and all satellites that comprise a global navigation satellite system and to time the arrival of their signals to a ground receiver.

In order to have a complete global concept of how GNSS systems work, one must understand different frames of reference around the globe by considering both, the dimensional space description and the abstract reference time.

Through basic communication engineering, we can understand how electromagnetic signals from space and Earth objects can be joined to give space-time ranging calculations, thus making possible many position- or time-dependent applications in an ample variety of fields including scientific, engineering, technical, social, and

even the very life-sustaining capabilities of a society. Thus, the basis for GNSS signal structure and modulation communications are presented as well as those techniques that permits a user to achieve a desired positioning precision.

Since reception of satellite signals can be obtained from either lumped or distributed receiver-transmitter systems, static or dynamic environments, clear, partial, or almost completely obstructed access for electromagnetic waves, a description of error sources and ways of dealing with measurements made in each type of situation is a must.

GNSS Orbits

To be able to cover the terrestrial globe at any given static or dynamic spot at any given time, GNSS are not placed in GEO orbits. For the same purpose, GNSS could not be comprised of a single spacecraft but rather a well-established set of spacecraft forming a proper and well-calculated system. The specific kinematic geometry for the set of satellites that form such system is called constellation. The very first GNSS was the well-known GPS and can be taken as a good example of a constellation placed on MEO orbits shown in Fig. 23.1.

Modern display and tracking of the actual real time orbits can be followed at NASA's science@NASA Satellite tracking (http://science.nasa.gov/realtime/jtrack/3d/JTrack3D.html/). Choosing any GPS satellite one can obtain a practical idea of what a MEO orbit means with respect to Earth.

Now, following the GPS example, it can be seen that for a constellation of 24 GNSS satellites at MEO orbits with an intended global coverage of at least four satellites above the celestial vault for a given point on Earth, there ought to be four satellites in a particular MEO orbit (defining an orbital plane), and there ought to be six orbital planes. Satellite separation in each plane is about 90° and, in the GPS case, planes differ from each other 60° and the origin plane is taken as being 55° respect to the equatorial plane.

Up to date, there are two well-established GNSS: GPS (Mohinder et al. 2001) owned by the United States of America and GLONASS (Global Navigation Satellite System: GLONASS 1998) owned by the Russian Federation. There are two other GNSS under development and planned to be active in the near future: GALILEO (Wilson 2006) owned by the European Union (EU) and COMPASS (Chen et al. 2009) owned by the People's Republic of China. The Quasi-Zenith Satellite System (QZSS) is a proposed three-satellite regional time transfer system and Satellite-Based Augmentation System (SBAS) for the Global Positioning System that would be receivable within Japan. The first satellite "Michibiki" was launched on 11 September 2010. Full operational status is expected by 2013. The GPS Aided Geo Augmented Navigation (GAGAN) is another planned regional SBAS by the Indian government. These systems will improve the accuracy of a GNSS receiver by providing reference signals (see ► Chap. 25, "Current and Future GNSS and Their Augmentation Systems").

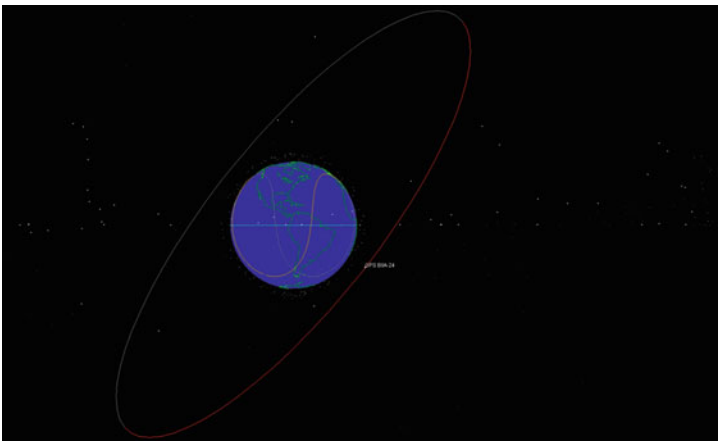


Fig. 23.1 Example of a GPS satellite MEO orbit (http://science.nasa.gov/realtime/jtrack/3d/JTrack3D.html/)

MEO Orbits

Table 23.1 shows all orbit data related to the GNSS (Shaw et al. 2008; Revnivykh 2008; Verhoef 2008; China Satellite Navigation Project Centre 2008).

Reference Systems

Despite the fact that one can visualize what 90° of separation between satellites in a particular plane and 60° among all planes means, such angles should be referred to a specific frame of reference. The most widespread 3D reference system is that constructed from three straight lines with right angles among them: the Cartesian Co-ordinate System (CCS) because its geometry and associated algebra are easy to understand.

From practical human-size measurements, our world can be thought as being Cartesian in a spatial square box of 100 m each side, meanwhile, for global Earth-size measurements, a geometrical abstraction reference named geoid has been defined and finally, for planetary-size measurements, a sphere centered in the Sun is the most common idea of a reference.

When quasi-spheres are involved in mobile outer space trajectories, two common frames of reference are used: one related to the center of the Earth (Conventional Terrestrial System CTS) and the other one related to a celestial or astronomical Reference system (ARS) the former being not an inertial frame of reference the latter one almost fulfills the inertial frame requirement.

Figure 23.2 shows the CCS, CTS, and ARS reference systems and their respective coordinate axis known as *horizontal datum and vertical datum* (Rogers 2003).

Table 23.1 MEO orbits for GNSS

GNSS/space segment	GPS	GLONASS	Galileo	COMPASS/BeiDou 2
Constellation (Space vehicles)	24 Operational/ 3 spares 31 GPS-M	24	27 Operational/3 spares	30/5
Orbit	MEO	MEO	MEO	MEO/GEO
Orbital period	11 h 56 min ~ 12 sidereal h	11 h 15 min	14 h 05 min	12 h 50 min
Altitude/orbital radius (Km)	~20,200/ 26,600	19,100/ 25,508	~23,616/ 29,600	21,500/27,840 35,785/42,164
Planes	6	3	3	3
Planes position (°from equator) latitude/separation	55/60	64.8/120	56/120	55/120
Satellites per plane/ degrees of separation	4/90	8/45	9/40 +1	10/36
Declination				58.75E,80E,110.5E,140E, 160E
Reference system/ time	WGS-84/GPST	PZ-90/ UTC(SU)	GTRF/GST	China geodetic system (CGS 2000) ~ ITRF/China UTC (BDT) ~ UTC

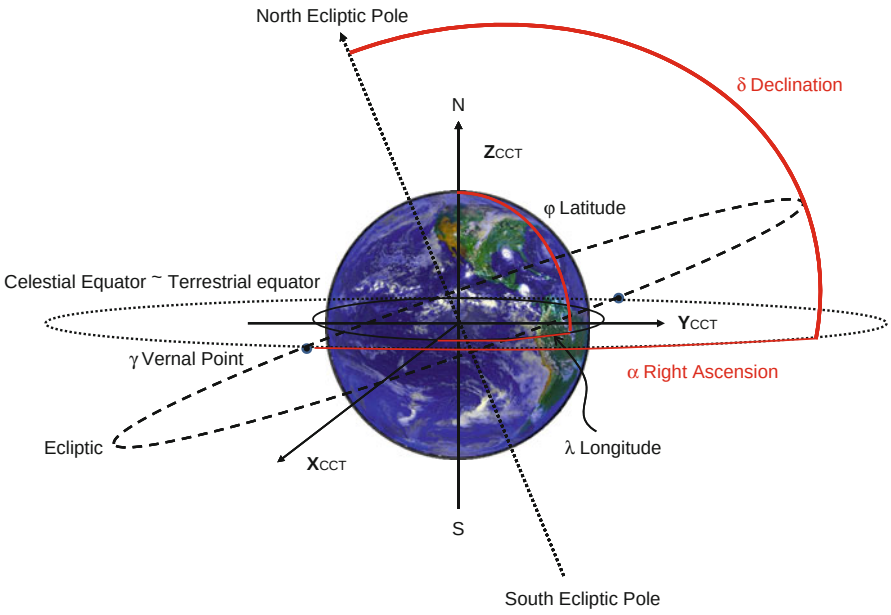


Fig. 23.2 Horizontal datum and vertical datum in the CCS (X, Y, Z), CTS (λ, φ), and ARS (α, δ) reference systems

Relationships among CCS, CTS, and ARS values are given in terms of coordinate transformations (Xu 2003). To view the effects of such transformations the reader could use one of the most practical and modern informatics tool: the Google Earth (http://www.google. com/earth/learn/beginner. html) where one can switch views from CTS to CCS and vice versa using the TOOLS command and changing datum views at the 3D view window tab. Thus, vertical datum (Orthometric Heights H) and horizontal datum (latitude φ and longitude λ) are shown either in degrees for the CTS or in meters for the CCS under the Universal Transversal Mercator (UTM) Projection (Carnes 2007) (In Fig. 23.2, H is changed from the geoid reference to the ellipsoid reference or Geometrical height h).

Ephemeris

From the previous concepts, it can be seen that satellite position coordinates, for non GEO orbital satellites, change with time. In Fig. 23.3 the 3D position vector r of a dynamical satellite can be most properly written as $r(t)$ which is known as the *satellite's ephemeris* (Beutler 2005).

Ephemeris are calculated using celestial mechanics, and can also be predicted using estimators from observed data. Satellite observations are a task for the tracking ground systems which are part of the so-called terrestrial segment of any satellite system. For the purpose of information completeness, satellite systems are composed by two segments, the space segment, composed by the satellite constellation, and the terrestrial segment, composed by the ground control and tracking subsystem. The terrestrial segment also includes a subsystem that is formed by all the GNSS users. Figure 23.4 shows all necessary GNSS segments.

From Fig. 23.3 ephemeris are given by: Right Ascension (α), Declination (δ), Argument of latitude (ω) and Range $\| r(t) \|$.

Ground Tracks

Figure 23.5 represents an ephemeris projected over the surface of the terrestrial globe for a single orbit and for all the time that takes that orbit to be completed. Those projections are known as ground tracks and in fact are closed paths that resemble sinusoidal trajectories.

Ground tracks together with geographical maps are useful to actually see ephemeris and ease satellite tracking or, in a more general case, to visualize the position of any space vehicle around Earth. Figure 23.5 shows a typical ground track map consisting of a flattened 3D map where a specific ground track has been drawn. There could be as many ground track maps as necessary to properly and clearly display as many satellites as are being tracked. Obviously, these maps are 2D but allow the visualization of the past, the present, and the future of satellite paths.

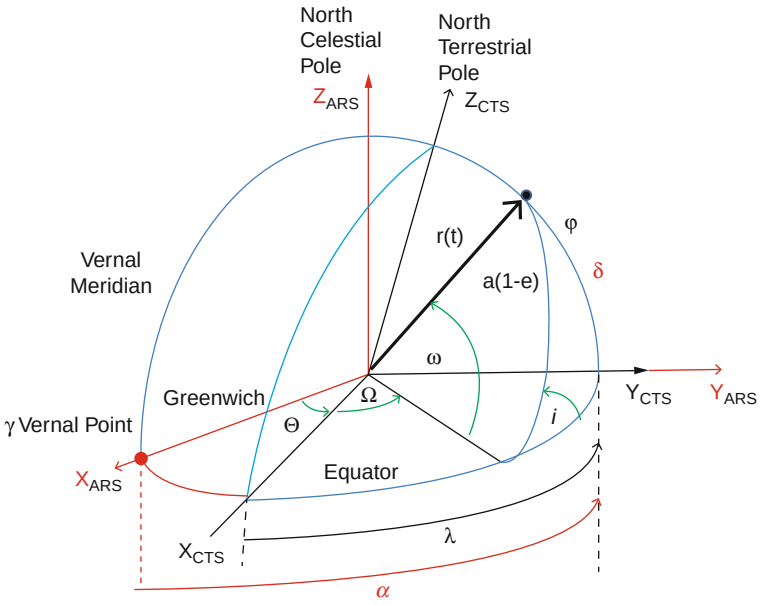


Fig. 23.3 Satellite ephemeris

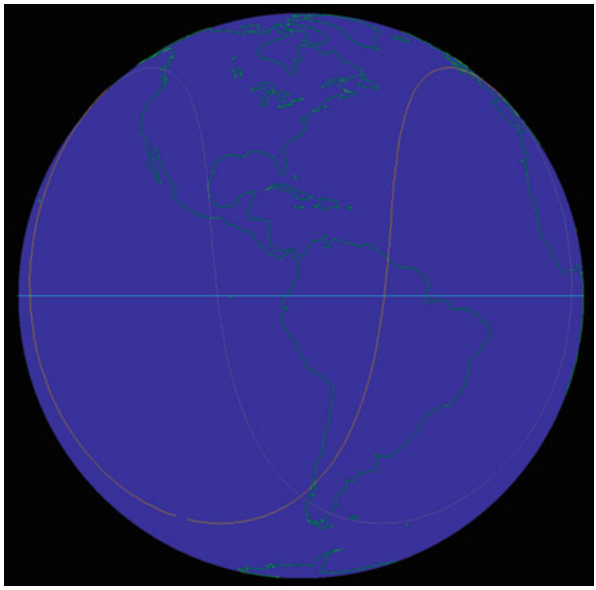


Fig. 23.4 Ground track projection (http://science.nasa.gov/realtime/jtrack/3d/JTrack3D.html/)

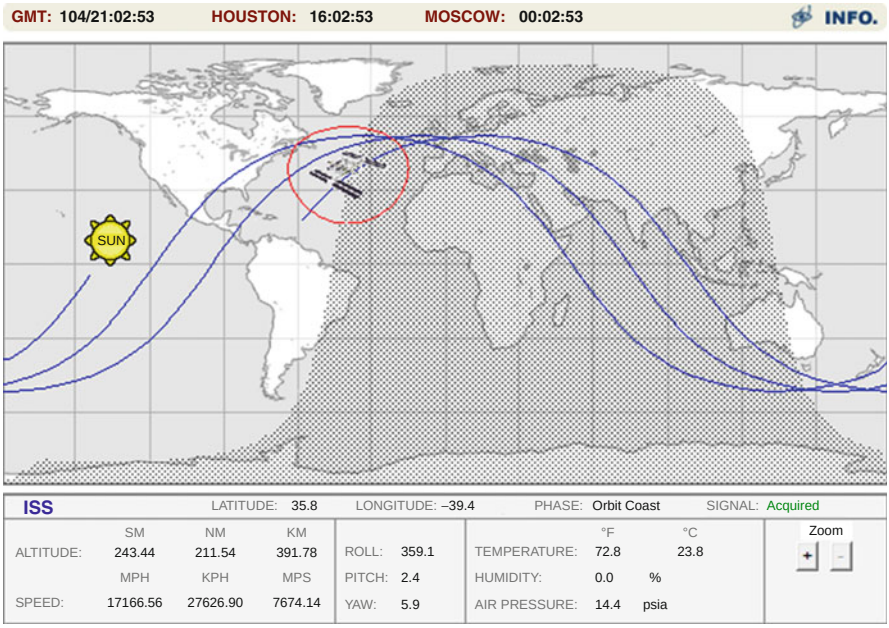


Fig. 23.5 Typical ground track map (http://spaceflight.nasa.gov/realdata/tracking/index.html)

Polar Plots

There is another 2D kind of satellite movement representation: polar plots. A polar plot is nothing else than a polar projection of the whole sky dome above a particular observer looking up the zenith and oriented toward North. Because this is a 2D projection, the vertical datum, i.e., the satellite's range is not considered. Thus, the horizontal datum of satellites above the observer is represented on one hand, by a compass giving the satellite's right ascension and on the other, by circles that represent satellite's declination angle starting from the zenith and ending at the observer's horizon. Figure 23.6 depicts an example of a polar plot with four satellites in view.

It should be noted that polar plots just show the position of satellites for a given time. Nevertheless, history of trajectories can be consulted as well as future positions using the memory and/or predicting tools proper of the modern informatics system.

Point Positioning

Point positioning is the process of finding the vector R of a specific user in terms of the satellite range, ρ, and the instantaneous satellite ephemeris $r(t)$ or simply r. Point positioning is shown in Fig. 23.7.

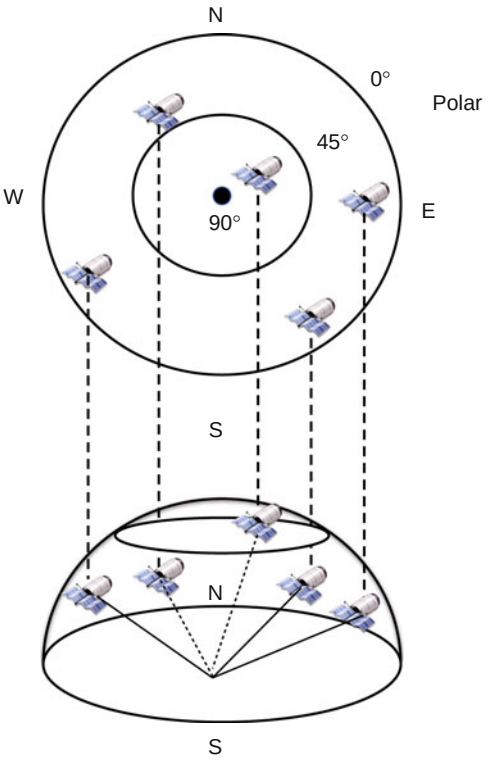


Fig. 23.6 Polar plot with four satellites above an observer

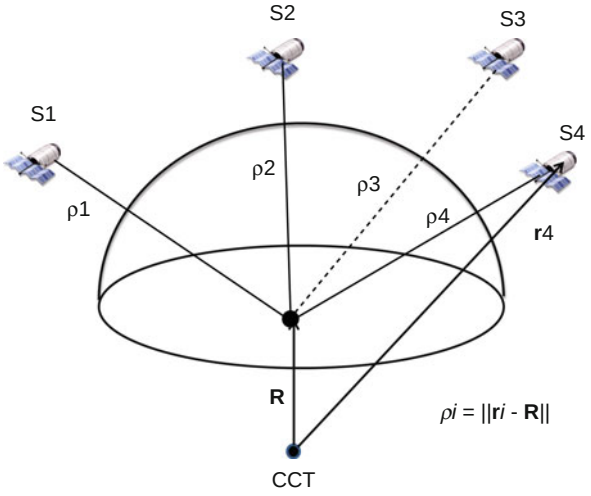


Fig. 23.7 Point positioning

Thus, the user vector R is given by the equation

$$R = r - \rho$$

Differential GNSS

Sometimes R is not the quantity we are interested in, but rather a position with respect to another point taken as a new reference. In such case we can talk about differential positioning. Figure 23.8 shows this concept.

The Differential GNSS (DGNSS) equation is easily obtained from Fig. 23.8 as:

$$\Delta R = R_2 - R_1 = \rho_1 - \rho_2$$

where $R_i = r - \rho_i$ and r is the ephemeris to the very same satellite used for positioning fix.

There is an advantage to using DGNSS and that is that the uncertainties in the determination of the time of arrival of signals from the satellites "in view" of the GNSS receiver are reduced, and therefore, DGNSS is used when high-precision positioning is required.

However, the differential concept extends to other ways of calculating positions on Earth. There are many ways of doing differential positioning (Wells 1987): differences between observations points, differences between satellites in view, differences between epochs of satellites, and some other linear combination of observations. It is even possible to define a network of local references. The purpose of such a network is to obtain data redundancy, thus allowing for diminishing positioning errors.

An advanced technique to combine observations is known as *bundle* and it has applications where a very high precision is required in practical measurements (Triggs et al. 2000).

Any least-squares adjust deals with two equally important models: stochastic and mathematical models (Fig. 23.9).

The n observations are the primary input to the adjustment and are denoted with the vector $\mathbf{L}_b$. The variances of these distributions comprise the stochastic model which gives information about (relative) precision of the observations, accomplished through the variance-covariance matrix of the observations $\mathbf{\Sigma L}_b$. Because in surveying the observations belong to a normal distribution, the vector of observations is a sample from the multivariate normal distribution. If $\mathbf{L}_P$ denotes the vector mean of the population, $\mathbf{L}_b$ can be expressed as:

$$\mathbf{L}_b \sim N(\mathbf{L}_p, \mathbf{\Sigma L}_b)$$

The adjustment requires either, the cofactor matrix $\mathbf{Q}_{\mathbf{L}_b}$ of the observations, which corresponds to the scaled variance-covariance matrix:

$$\mathbf{Q}_{\mathbf{L}_b} = \left(1/\sigma_0^2\right) * \mathbf{\Sigma L}_b$$

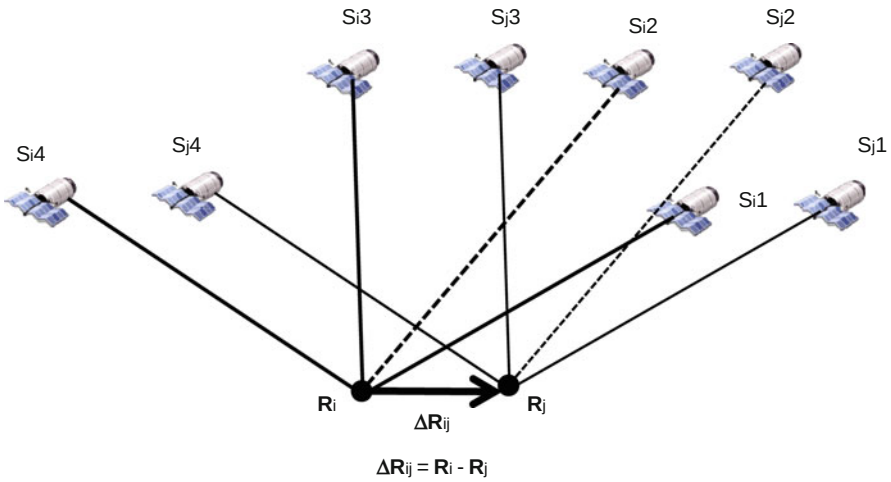


Fig. 23.8 Basic differential positioning

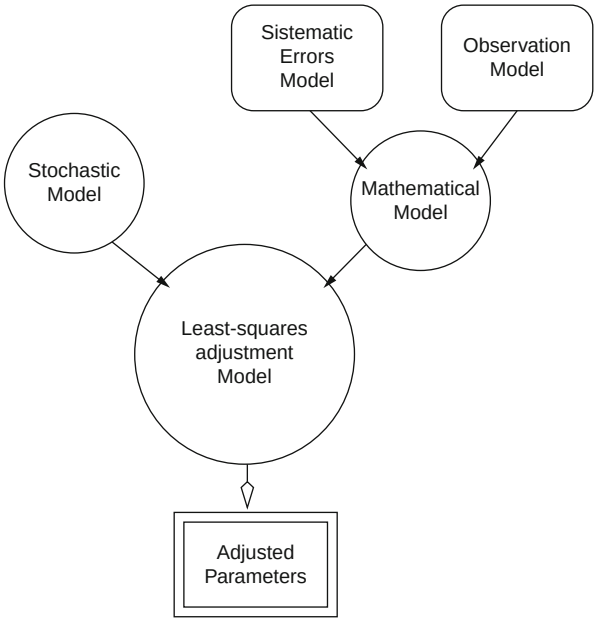


Fig. 23.9 Least-squares adjustment elements

or the weight matrix **P**:

$$\mathbf{P} = \mathbf{P}_{\mathbf{L}_b} = \mathrm{Q}_{\mathbf{L}_b}^{-1} = \sigma_0^2 \boldsymbol{\Sigma}_{\mathbf{L}_b}^{-1}$$

The scale factor σ_0^2 is called the a priori variance of unit weight.

The original observations are modified to yield the so-called model observation. In most surveying applications the mathematical model is *nonlinear*. If the observations are explicitly related to the parameters, such as in:

$$\mathrm{F}(\mathbf{X}_a) = \mathbf{L}_a$$

where $\mathbf{X}_a$ are the u adjusted parameters (unknowns) and $\mathbf{L}_a$ is the vector of n adjusted observations. This is the *observation equation model,* and it has particular advantage that each observation generates one equation,allowing its implementation seamlessly on a computational algorithm.

If $\mathbf{X}_0$ is a vector of known approximated values of the parameters, then the parameter corrections $\mathbf{X}$ are:

$$\mathbf{X} = \mathbf{X}_a - \mathbf{X}_0.$$

If $\mathbf{L}_b$ denotes the vector of observations, then the residuals are defined by:

$$\mathbf{V} = \mathbf{L}_a - \mathrm{L}_b$$

and using both previous equations, the mathematical model can be written as

$$F(\mathbf{X_0} + \mathbf{X}) = \mathbf{L}_b + \mathbf{V}.$$

This nonlinear mathematical model can be linealized around a known point of expansion ($\mathbf{X_0}$) giving:

$$\mathbf{L}_b + \mathbf{V} = F(\mathbf{X}) + \frac{\partial[F(\mathbf{X})]}{\partial \mathbf{X}} dx \bigg|_{\mathbf{X}=\mathbf{X_0}}$$

If $F(\mathbf{X})|_{\mathbf{X}=\mathbf{X_0}} = \mathbf{L_0}$, that is, the value of the observations from the approximated parameters $\mathbf{X_0}$ (the point of expansion), and every partial derivative $\frac{\partial[F(\mathbf{X})]}{\partial \mathbf{X}} dx \Big|_{\mathbf{X}=\mathbf{X_0}}$ is expressed as the product of the design matrix $\mathbf{A}$ by the parameter matrix $\mathbf{X}$, then:

$$\mathbf{L}_b + \mathbf{V} = \mathbf{L_0} + \mathbf{AX}$$

which finally gives

$$\mathbf{V} = (\mathbf{L_0} - \mathbf{L}_b) + \mathbf{AX} = \mathbf{L} + \mathbf{AX}$$

where the difference $\mathbf{L}$ is often called the misclosure.

The previous linealized equations are called observation equations and can be written as

$${}_n\mathbf{V}_1 = {}_n\mathbf{A}_u \, {}_u\mathbf{X}_1 + {}_n\mathbf{L}_1 = \mathbf{0}$$

The least-square estimate $\hat{\mathbf{X}}\mathbf{E}$ is based on the minimization of the function $\mathbf{V}^{\mathrm{T}}\mathbf{P}\mathbf{V}$. A solution is obtained by introducing a vector of *Lagrange multipliers*, λ, and minimizing the function:

$$\Phi(\mathbf{V}, \lambda, \mathbf{X}) = \mathbf{V}^{\mathrm{T}}\mathbf{P}\mathbf{V} - 2\lambda^{\mathrm{T}}(-\mathbf{V} + \mathbf{A}\mathbf{X} + \mathbf{L})$$

The solution starts by noting that P is a squared matrix and can be inverted. Thus, the expression for the residuals is:

$$-\mathbf{V} = \mathbf{P}^{-1}\lambda$$

and the solution for the Lagrange multiplier is:

$$\lambda = -\mathbf{P}(\mathbf{A}\,\hat{\mathbf{X}}\mathbf{E} + \mathbf{L})$$

from where the estimate vector $\hat{\mathbf{X}}\mathbf{E}$ follows

$$\hat{\mathbf{X}}\mathbf{E} = -(\mathbf{A}^{T}\mathbf{P}\mathbf{A})^{-1}\mathbf{A}^{T}\mathbf{P}\,\mathbf{L}$$

Kinematic and Real Time Kinematic Systems

Kinematics positioning is the terminology utilized when the user vector R is changing dynamically (as in a moving vehicle), and therefore, the user position vector becomes a function of time, *R(t)*. From Fig. 23.7 the equation for $R(t)$ is given by:

$$R(t) = r(t) - \rho(t)$$

In the previous equation, when *R(t)* changes slowly with respect to all elements then, real time kinematics positioning takes place. Practical uses of kinematics positioning are found in GNSS receivers, which display velocity and acceleration of the platform where the receiver is placed.

Real time relative kinematics positioning can also be used when a reference receiver transmits the observed ranges for the very same satellites that are used by the receiver at the mobile platform. The transmission is done by a radio link (e.g., using Radio Technical Commission for Maritime Services, known as RTCM signals). If the message sent by the reference station is similar to those transmitted by the GNSS satellites, the reference station is called a pseudo-satellite or pseudolite.

GNSS Augmentation Systems

DGNSS differential positioning has evolved from local reference stations – which broadcast error correction – to the most actual and modern satellite systems to support GNSS. Such DGNSS satellite systems are named augmentation systems

or satellite-based augmentation system (SBAS). The former was used mainly for Maritimes users (RTCM-104 1994) while the latest took the Radio Technical Commission for Maritime Services (RTCM) broadcasting correction techniques to implement pseudo-satellite systems (Wang 2002) which are mainly used in airports.

Nations have undertaken to develop and construct augmentation systems for their own specific territories. Such are the cases of those of the North America's Wide Area Augmentation System (WAAS), for Europe the European Geostationary Navigation Overlay Service (EGNOS), the Indian GPS Aided Geo Augmented Navigation (GAGAN), and the Japanese Multi-functional Satellite Augmentation System (MSAS). Using various kinds of differential positioning, commercial systems such as John Deere's StarFire^{TM} for precision agriculture offer position accuracy levels ranging from +/− 13 to 1 in. Other services may use cellular phone or satellite telephony as in the case OmniSTAR which can provide up short-term positioning accuracies of 1–2 in. and long-term repeatability of 10 cm (Hofmann-Wellenhof et al. 2008).

International Atomic Time (TAI)

It is common to introduce time as the parameter to describe both the ephemeris and equations of movement of a user in a moving platform, but that also introduces a common mistake: to think time as being a unique quantity for all observers. However, after Albert Einstein's demonstration that time is a relative measurement, there is a need to define a reference time system. First there ought to be a time unit, under *Système international d'unités* (SI) concepts, and such a unit should be the same for any observer in an inertial frame of reference. Nowadays, such standard time unit is defined and measured by modern atomic clocks. The typical atomic clock realization is one that uses a Cesium atom standard.

A "cesium atomic clock" is a device that uses as a reference the exact frequency of the microwave spectral line emitted by atoms of the metallic element cesium. A cesium clock operates by exposing cesium atoms to microwaves until they vibrate at one of their resonant frequencies and then counting the corresponding cycles as a measure of time. This frequency provides the fundamental unit of time, which may thus be measured by cesium clocks. Today, cesium clocks measure frequency with an accuracy of from 2 to 3 parts in 10^{14}, i.e., 0.00000000000002 Hz. This corresponds to a time measurement accuracy of 2 ns per day or 1 s in 1,400,000 years. It is the most accurate realization of a standard unit that mankind has yet achieved.

The SI second was defined in terms of the cesium atom in 1967, and in 1971 it was renamed International Atomic Time (TAI, from the French name Temps Atomique International). TAI is a high-precision atomic coordinate time standard based on the notional passage of proper time on Earth's geoid. It is the basis for Coordinated Universal Time (UTC), which is used for civil timekeeping all over the Earth's surface, and for Terrestrial Time, which is used for astronomical calculations. This time references are explained below.

TAI as a time scale is a weighted average of the time kept by over 200 atomic clocks in about 70 national laboratories worldwide. In the 1970s, it became clear that the clocks participating in TAI were ticking at different rates due to gravitational time dilation, and the combined TAI scale therefore corresponded to an average time of various clocks at different altitudes. Starting from 1 January 1977, corrections were applied to the output of all participating clocks, so that TAI would correspond to proper time at mean sea level (the geoid) (see ▶ Chap. 26, "Introduction and History of Space Remote Sensing"). As of 1 January 2011, TAI was exactly 34 s ahead of UTC (this is the case since 1 January 2009): an initial difference of 10 s at the start of 1972, plus 24 leap seconds in UTC since 1972 (McCarthy and Seidelmann 2009).

Historically, time was measured using Earth's rotation with respect to the Sun. When the Sun is observed from a particular point on the surface of the geoid we get the *true solar time*; however, because days do not last the same all year round, it was more adequate to talk about *mean solar time*. For most civil uses, the solar mean time measured with respect to the meridian at Greenwich is known as Greenwich Mean Time (GMT) and its importance resides in that it helped to define internationally the Universal Time (UT).

It should be clear that measuring time with respect to the Sun is different from the time measured with respect to a more inertial reference frame as that defined by distance stars and, such time is called sidereal time (ST). ST is used by astronomers to locate celestial objects at astronomical ephemeris almanacs.

UT and ST, despite of having different origin, have very slight differences between their respective unit time, and therefore, they both properly define the Terrestrial Time (TT).

Modern refined measurements established that the duration of days are affected by as many factors as those that change the moment of inertia of the entire Earth (dynamics of water, land, and wind), making the Earth's rotation to slow down; thus, UT is corrected by leap seconds and such time defines the Universal Time Coordinate (UTC). Such proper name was chosen as an international compromise to agree on a specific name to enclose all types of contemporary reference times coexisting when all sorts of corrections have been considered trying to adjust the TT.

Furthermore and in agreement with Einstein's theory of relativity, when the reference system is not an inertial one, then clocks measure time at different rate; thus, when a system is accelerated, the time unit shrinks or stretches, and, as a result, reference clocks tick differently when they move under acceleration. This effect is noticeable for GNSS because in general, all users' platforms are accelerated at different rates when they are at the surface of the Earth compared to when they are above it and the effect is accentuated when very different heights are involved.

As an example of such relativity effect, and because of both the high-precision time measurements involved and that the GNSS satellites and user's platforms are located at different heights moving at a particular relative velocity, there is the need to consider a correction factor of about 4.45×10^{-10} Hz to their respective clocks (Audion 2001). GLONASS clock is at 5 MHz at ground but in sky is -2.18×10^{-3}.

Table 23.2 Time reference frames

	Duration of day		
Reference	Solar	Sidereal	ITU[a] adjustments
Local meridian	Aparent solar time	UT0	
Greenwich meridian	Mean solar time (GMT)	Ephemeris time (ET)	Leap years
International celestial reference frame @ zero meridian	Universal time (UT)	Terrestrial time (TT)	
Geoid *temps atomique international* (**TAI**)	UT1 civil uses	**Universal time coordinate (UTC)** Maritime, aerial, computers, etc.	Leap seconds
Corrections			
Phenomena	Tides UT1R Seasonal UT2	Terrestrial gravitation *Temps-coordonnée géocentrique* TCG	
Phenomena	Both UT2R	Solar system gravitation *Temps-coordonnée barycentrique* (TCB)	

[a]International Telecommunications Union

Sometimes, for scientific purposes it could be convenient to avoid relativity effects due to gravitational field differences, thus measuring time at the center of the Earth which leads to a corrected time named *Temps Coordonnée Géocentrique* (TCG) or Geocentric Coordinate Time (Guinot 1986). Furthermore, when measuring time in other planets, or other space artifacts, then TCG has to be exchanged by another relativistic corrected time, the Barycentric Coordinate Time (TCB). The instant that the gravitational correction started to be applied serves as the epoch for Barycentric Coordinate Time, Geocentric Coordinate Time, and Terrestrial Time. All three of these timescales were defined to read 1 January 1977 00:00:32.184, exactly at that instant. (The offset is to provide continuity with the older Ephemeris Time.)

Table 23.2 summarizes the diversity of TT.

Nonetheless, the basis for all those time references is the TAI, and therefore, there are equations that relate all those different times.

Signal Structure

GNSS satellites broadcast information that is used by GNSS receptors to calculate positions on Earth. Since the information is sent using radio transmission, it is important to consider some basic telecommunications techniques.

First of all, to be able to transmit data information in which long distances are involved, data are modulated in a carrier wave which, as in the GNSS case, is one of the electromagnetic type because the transmission is wireless in open space.

Second, to send digital information in a continuous wave carrier, the modulation technique depends on whether the information changes the amplitude, the

frequency or the phase of the carrier wave (Spilker 1977). These different modulation techniques respectively are: Amplitude Shift Keying (ASK), Frequency Shift Keying (FSK), or Phase Shift Keying (PSK). A practical and simple implementation of a digital modulation is the BPSK when one binary symbol is transmitted using one bit at a time. Other implementations are: (1) QPSK, when two bits are used to transmit the information, and therefore, 4 binary symbols are transmitted simultaneously; (2) QAM, a combination of QPSK and ASK when four bits transmit 16 binary symbols, etc.

Third, to access all different satellites in any specific GNSS constellation, there are techniques that use time, frequency, or codes assigned to each satellite which allows distinguishing among all multiple satellites that compose the constellation. Correspondingly, these multiple access techniques are (Lahi and Ding 2009): Time Division Multiple Access (TDMA), Frequency Division Multiple Access (FDMA), and Code Division Multiple Access (CDMA). Other distinguishing possibilities are: (1) the polarization of the traveling carrier wave; (2) Orthogonal Frequency Division Multiple Access (OFDM), which stresses the orthogonality frequency feature among waves; and (3) Wave Division Multiple Access WDMA, which exploits wavelength as a distinction among carriers.

Table 23.3 summarizes the signal structure related to actual GNSS.

Finally, to recuperate any message that is sent by satellite, there is a format under which a particular full message is to be sent. A practical example can be taken from GPS in which such format (Borre et al. 2007) is formed in a top-down fashion as: a master frame, frames, subframes, and digital words. Such format is shown in Fig. 23.10.

Carrier Frequencies

Since the carrier wave is an essential part in GNSS, let us consider some useful carrier electromagnetic wave characteristics. With respect to selecting a specific satellite operating communications frequency, when using global communication environments it is essential to respect international telecommunication regulations as well as physical laws for electromagnetic waves. For example, GNSS use the L and C bands (Wither 2000) of the international radio-electrical spectrum since those bands contain the frequencies for mobile communications. And with respect to physical environments, the carrier frequency should be selected taking into account that such frequency is going to travel large distances across the Ionosphere, atmosphere, the biosphere, and within cities' concrete canyons, etc. (Bensky 2008).

From the electrical point of view when selecting the operating frequency the design engineer works backward from the given technical specifications, such as speed of information transmission, selected multiple access and modulation techniques, transmission power, and efficient use of the satellite's energy, all this of course within the allocated transmission frequencies for GNSS. It is worth to note that all necessary frequencies employed by satellites are

Table 23.3 GNSS signal structure

GNSS				**GPS**			**GLONASS**			**GALILEO**			**BEIDOU**	
Multiple Access Method				CDMA			FDMA			CDMA			CDMA	
International L band 1 - 2 GHz Subdivisions E, G & B			**Carrier Frequency** [GHz]	Ranging Code	Length [chips]	Code-Data Modulation /speed [MHz]	Ranging Code	Length [chips]	Code-Data Modulation /speed [MHz]	Service	Length 10230 [chips]	Code-Data Modulation /speed [MHz] 1.023MHz	Service	Modulation
L1	G1		1602.00				CA/P		BPSK/0.511					
	E1	B1-2	1589.74										OS	QPSK/MBOC
L1c	E1 L1 E2		1575.42	CA/M SPS/PPS	1023	BPSK /1.023/5.115				OS/SoL/ PRS/CS	4092/8184	BOCc(1,1) data BOC(15,2.5) pilot 1.023Mcps		
	E2	B1-1	1561.10										OS	QPSK/MBOC
L2c			1381.05	CA/M		BPSK-BOC /1.023	CDMA modernization							
	E6		1278.75							CS/PRS	✓	BPSK pilot Data 5.115Mcps BOC(10,5)		
	E3	B3	1268.52										OS	MOBC
L2	G2		1246.00				p		BPSK/5.11					
L2			1227.60	C/P(Y)/M	10230	BPSK/0.5115 /10.23/5.115								
L3	E5 b		1207.14							OS/CS/S oL		BPSK Data Pilot		
	E5		1196.91							OS/CS/S oL	✓	QPSK/BPSK BOC(15,10)d ataI cos pilotQ sin		
L5	E5 a		1176.45	CS/M	10230	BPSK-BOC /10.23				OS/CS/S oL	✓	bpsk 10.23		
	E5 a+	B2	1027.14							OS/SoL /SAR	✓	QPSK pilotQ DataI	OS	QPSK
POWER [dBW]				-158 to -150			-167 to -161			-155			NA	

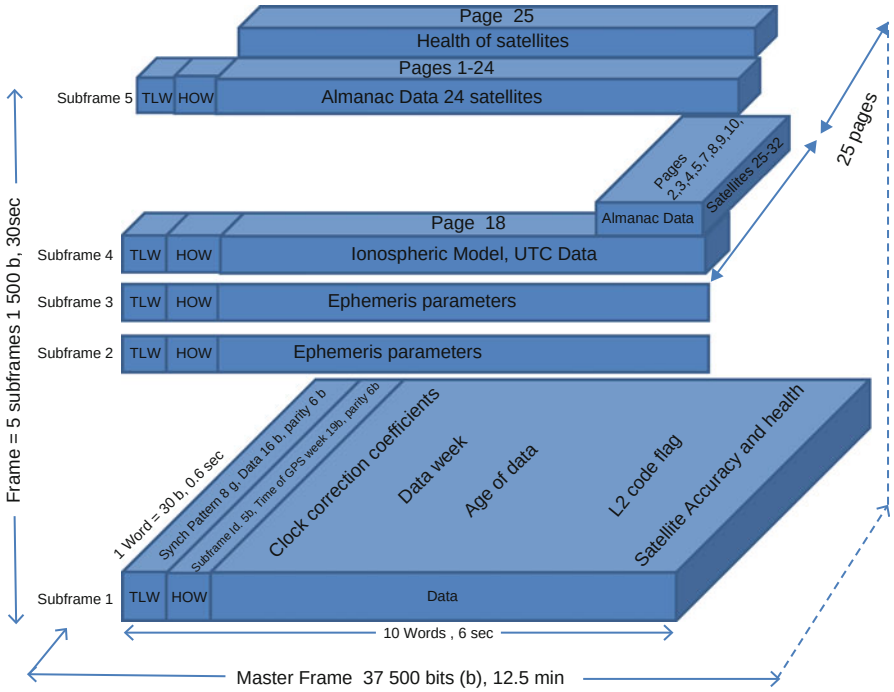


Fig. 23.10 GPS message format

derived from a unique atomic clock which runs in a very-well-determined frequency in agreement with all previous described technological communication considerations.

Since high-precision positioning is pursued, there are other frequency design considerations such as relativistic effects, ionosphere effects, multipath trajectories, and even encryption of the information for security reasons. On one hand, two frequencies are normally transmitted by the GNSS to allow corrections for ionospheric effects and, on the other, as the satellite atomic clock runs slower with respect to the master control clock on Earth, a correction is needed to compensate for the different gravitational potential between the satellite and the platform where the user is as well as their relative velocities.

Figure 23.11 shows the two sinusoidal waveform carriers in the L band in both time and frequency domains.

Data Stream and Messages

Once the carrier is selected, as described before, then the data stream and the message format should to be defined. Data stream refers to both transmission

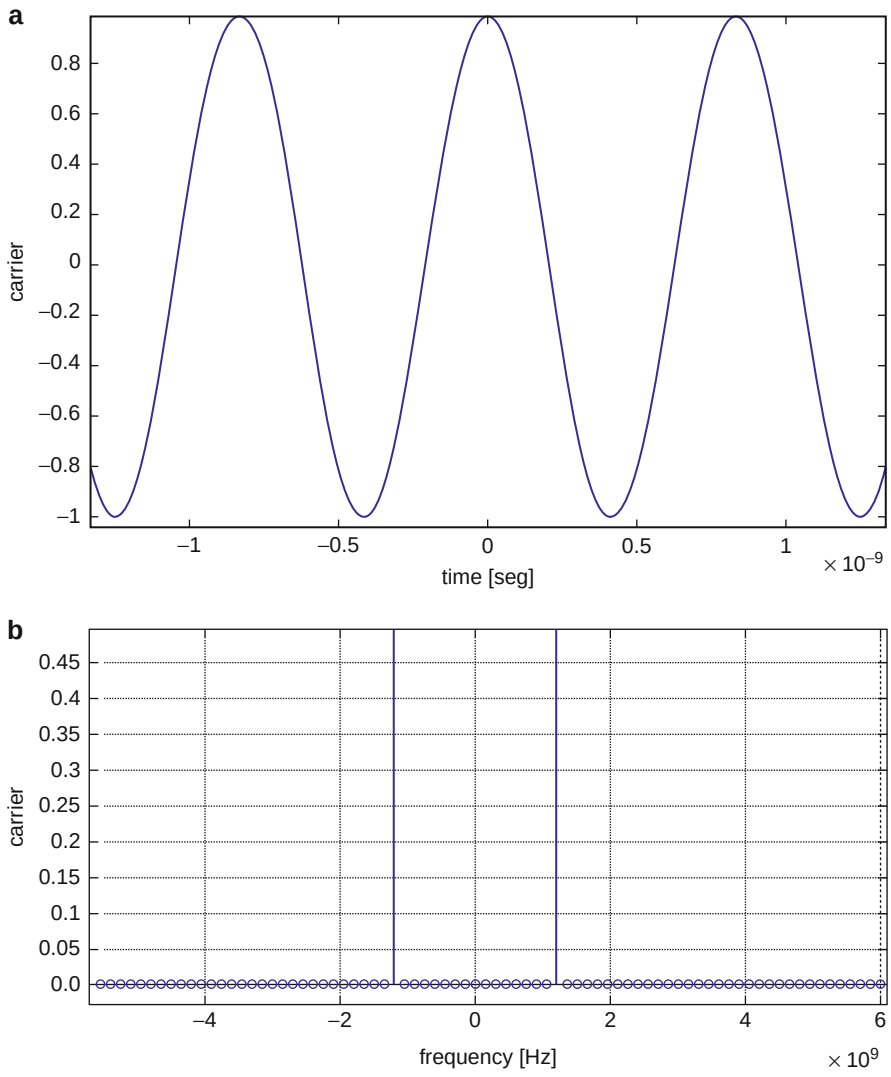


Fig. 23.11 Carriers in L band: (**a**) time domain and (**b**) frequency domain

speed and the way data is organized, while the message refers to all data needed for a GNSS receiver to calculate the position of a user.

The speed of data transmission affects directly the time at which anyone receptor performs a position fix for the first time. Remembering that at least four satellites are needed for GNSS positioning, the specific multiple access technique selected affects directly the transmission speed of data which forces the designer to make a compromise between the specific radio communication techniques and the amount of information needed to calculate the user's position.

In Table 23.4 we can find data transmission speeds for the main GNSS.

At the average data speed of 50 bps, it is common, for a receiver with the current processing technology, to take 12 min to calculate the user's position for the very first time.

With respect to all existing different GNSS messages, each GNSS has organized its message in a specific structure which basically contains information that, beside making position fix, timing, and prediction possible, also allows the user to know the satellite constellation ephemeris status and diagnosis, quality and availability of data for a given specific receiver location and data necessary to initiate acquisition of GNSS signals and for correction of errors due to clock and ionosphere.

For example the GPS message follows the format described by Fig. 23.10

Multiple Access and Pseudorandom Codes

It was mentioned that to distinguish and access each and all satellites in a specific GNSS constellation, a multiple access communication technique is needed. GPS uses CDMA to access all different satellites in its constellation and GLONASS uses FDMA. Thus, GLONASS only needs to use a single modulation to transmit all useful information but, if GLONASS wants to encrypt its message, an encryption code should be defined.

Encryption codes or CDMA codes are generally binary and generated using different amounts of bits which defines automatically the length of the code. Bits used to generate codes in fact do not transmit information and that is why those bits are rather called chips. Either, by using a code to create a CDMA or, by using a code for encryption, both produce the same effect on a binary modulated signal that is, both spread the total power radiated by the satellite (Dixon 1976). Despite of this spreading of power across the spectrum density of the signal still, Spread Spectrum systems (SS) keep the spectrum constrained to the length and speed at which the code is transmitted. CDMA codes or encryption codes receive the generic name of direct sequences (DS).

For an observer without the DS, the spread GNSS signal will look just like a sinusoidal wave embedded in noise and such apparent noise would have specific statistics. Since the DS code is digitally generated with a given amount of chips, the code will repeat itself, and therefore, the apparent noise receives the name of pseudorandom noise (PRN). Pursuing a low auto correlation code sequence is the purpose of the Gold codes random processes (Proakis 2008).

Different codes can be used for a given GNSS and named after the purpose for which they were defined. For example, for the GPS, a code for CDMA is called C/A code or Civil Applications code while, a proper code for encryption is called Y (formerly P) code. Y code was originally conceived to give the Precise Positioning Service (PPS) which allowed positioning with the full precision available from GPS. While C/A codes are well known and open for all public, Y codes are not.

Table 23.4 GNSS messages

GNSS				**GPS**		**GLONASS**		**GALILEO**		**BEIDOU**	
Multiple Access Method				CDMA		FDMA		CDMA		CDMA	
International L band 1 - 2 GHz Subdivisions E, G & B			**Carrier Frequency** [GHz]	Service	Data speed [sps]	Service	Data speed [sps]	Service	Data speed [sps]	Service	Modulation
L1	G1		1602.00			CA/P	50				
	E1	B1-2	1589.74							OS	QPSK/MBOC
L1c	E1 L1 E2		1575.42	CA/M SPS/PPS	50			OS/SoL/PRS/ CS	250		
	E2	B1-1	1561.10							OS	QPSK/MBOC
L2c			1381.05	CA/M	100	CDMA modernization					
	E6		1278.75					CS/PRS	1000		
	E3	B3	1268.52							OS	MOBC
L2	G2		1246.00			p	50				
L2			1227.60	C/P(Y)/M	50						
L3	E5 b		1207.14					OS/CS/SoL	250		
	E5		1196.91					OS/CS/SoL	250/50		
L5	E5 a		1176.45	CS/M	100			OS/CS/SoL	50		
	E5 a+	B2	1027.14					OS/SoL /SAR	250/50	OS	QPSK

It is interesting to note that, on one hand, a 1,024 chip length code transmitted at 1.024 Mbps per second will take 1 ms to be transmitted and thus one can expect to acquire the specific satellite transmitting such code in about that time, while a code that is 16,777,216 chips long or more will take at least 1 week to be detected. The latter is the case for military or restricted codes, furthermore Y codes are changed every week making it almost impossible to decode.

There is another advantage of using DS coding in that their use also makes it possible to calculate the range between two communications points (Zigangirov 2004). To do so, by knowing the encrypted or the CDMA codes, all that GNSS receivers have to do is to correlate a locally generated code with the arriving spread signal and the resulting time for correlation will give the pseudo-range information (see Pseudo-Range below). The correlation action basically unspreads the signal as well and, therefore, reconstructs the original information signal.

GALILEO has defined other spread spectrum technique which is intended to reduce interference between the BPSK modulated based CDMA signals. This new technique receives the name of Binary Offset Code (BOC) and, because it uses a single carrier but with a difference in phase of 90°, it produces two subcarriers the BOCsine and BOCcosine which simultaneously makes it possible to permit two different types of services. All codes and their respective services for GNSS can also be found in Table 23.3.

Modulation Techniques

With respect to modulation, BPSK is the most common realization in GNSS; however, Galileo uses BOC modulation which is a modified version technique. Figure 23.12 depicts both modulations. Specific modulation techniques for each GNSS is also given in Table 23.3.

True information bits are sent in the carrier wave as well of the corresponding chips for the DS code which are mixed under the XOR (bitwise eXclusive OR) binary operation with the purpose of recuperating easily the message at the receiver end. Thus, the signal received at the users end is:

$$r(t) = d(t)\sqrt{2P}\cos(2\pi ft + \vartheta_0)c(t-\tau) + v(t)$$

where $d(t)$ is the modulating data, P is the carrier power, f is the instantaneous frequency at the receiver antenna including the Doppler shift, $c(t)$ is the DS ranging code, τ is the time for correlation, $v(t)$ is the noise added to the original signal, and ϑ_0 is the initial carrier phase.

Figure 23.13 shows a GNSS signal modulated under the information and DS code in the time domain and its typical power.

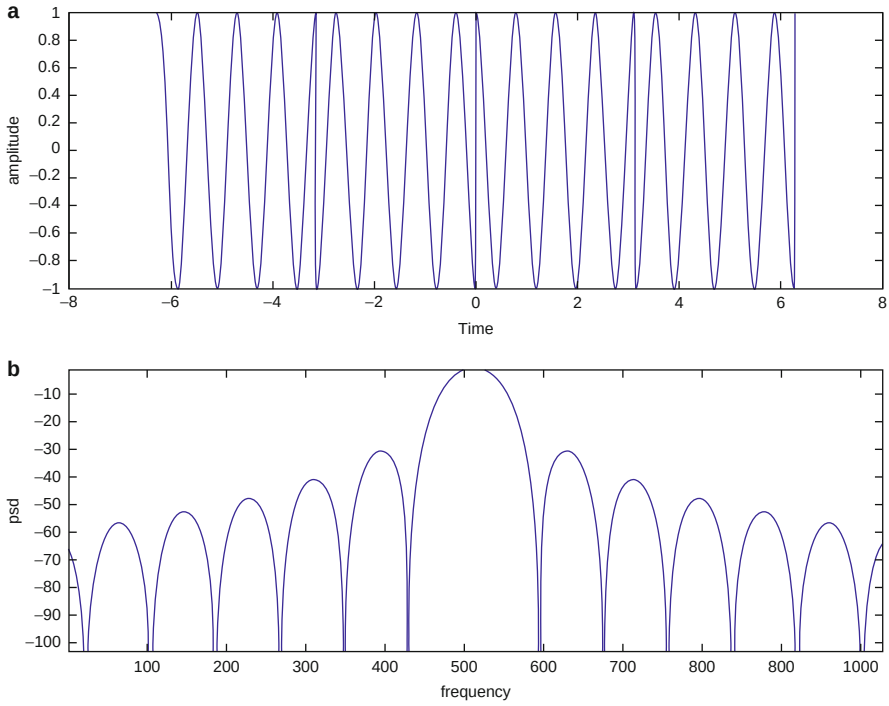


Fig. 23.12 BPSK and the BOC modulation: (**a**) time domain; (**b**) power spectrum density

Pseudo-Range

In the origins of GPS, to have access to the PPS, it was necessary to have special receivers and the corresponding military permit. Such a restriction challenged some researchers who wanted to benefit science by using the most advanced and high-precision positioning GNSS system available in contemporary times. Their challenging task was to calculate precise ranges to GNSS satellites without knowing any of the GNSS message at all.

Remember that because GNSS receivers perform a correlation to spread out GNSS signals, a by-product of such operation exists: the correlation time. In fact correlation time is exactly the time that it takes the signal to travel from the satellite to the receiver. Correlation time multiplied by the velocity of the carrier wave gives the range to the emitting satellite. However, since satellite and receiver operate in different time references, such difference introduces a bias into the measurements. When using this biased correlation time to calculate the range then biased range is named pseudo-range (Samama 2008).

Pseudo-range provides the possibility to calculate the position of the receiver without the navigation message because by knowing the correlation time and either, the period of the carrier or the period of the PRN code, it is easy to calculate

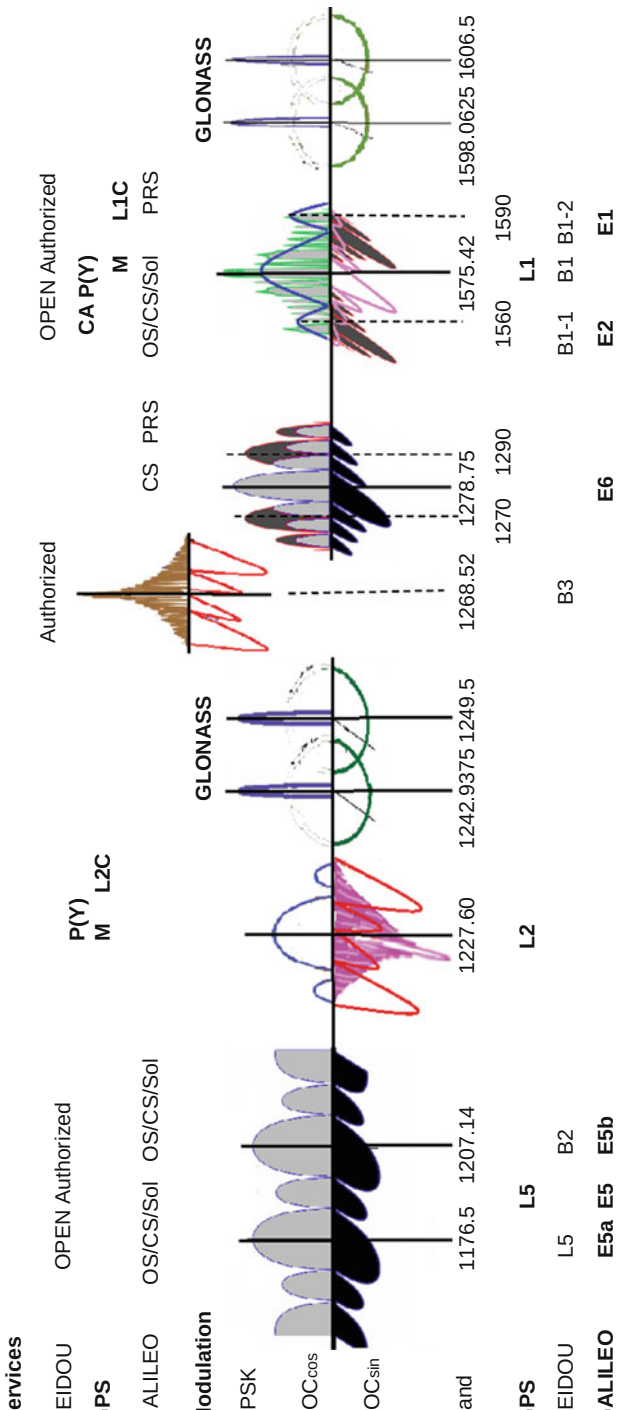


Fig. 23.13 GNSS signals

the number of cycles that would fit in the pseudo-range. In either case, carrier or code tracking are the corresponding observed variables commonly known as observations.

On one hand, observations may have cycle slips during the correlation process producing an error in their respective pseudo-range creating an ambiguity term (Strang and Borre 1997). Ambiguity adds to the total error budget in GNSS. On the other hand, carrier phase measurements are affected by the Doppler shift effect which can be estimated by calculating how the pseudo-range carrier observation varies with time, this rate of change gives the so-called accumulated delta range (Draganov 2006). Range measurement is improved when calculated from this observation.

Ranges can be estimated using stationary observations by the Weiner estimator but Kalman estimation is more adequate when the delta ranges observations are used (Mendel 1995).

Sources of Error and Error Budgets

We have seen that errors are introduced in the observations. There are two kinds of errors, the systematic and the properly stochastic errors. Systematic errors just produce bias; however, stochastic errors make the positioning a process that can only give estimates of the exact user position.

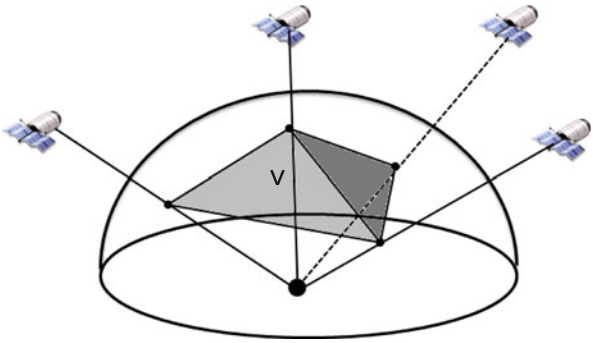
Bias errors come from different sources: from the space segment (Clock bias, ephemeris errors), from the propagation channel (atmospheric refraction, both by the Ionosphere and Troposphere, multipath due to reflection from fixed obstacles also called imaging), or from the user segment (clock bias and phase centre variation) (Leick 2004).

Systematic or bias errors can be eliminated or greatly reduced by using one or a combination of the differential positioning techniques (Muñoz et al. 2009). When observations are made simultaneously, electronic signals can be considered to have the same epoch, and thus, clock bias can be assumed to be uncorrelated. By doing differential positioning among two receivers and one satellite it is possible to cancel these specific space segment biases. Meanwhile, doing the counterpart differential positioning, that is, among two satellites and one receiver, cancels the specific user segment bias. A combination of both differential modes cancels most bias errors.

Bias error can be modeled and introduced into the observations equations. As an example, when several observations are carried out for the same point and made at different times, which implies that the corresponding signals have different epochs, then errors can be assumed correlated and easily modeled (Seeber 2003). Because advanced research in GNSS allowed clock, ephemeris, and atmosphere bias to be modeled, another way to reduce bias effects is the use of augmentation systems and broadcasting correction information.

It is very important to notice that since ranging is practically a geometry calculation, the actual and specific geometric arrangement between the satellite

Fig. 23.14 DOP



segment and the user segment impacts directly in the calculations. The error introduced by this geometry is given by the dilution of precision (DOP) (Bar-Shalom et al. 2001). The meaning of DOP can be easily understood based on Fig. 23.14.

Despite the fact that GNSS signals may be modulated with the specific purpose of reducing interference or/and jamming, there still a great dependence on the total power delivered onto the user receiver. Nowadays due to the presence of many electromagnetic fields in practical real life, undesired noise is present during observations affecting the signal to noise ratio, the carrier to noise power density ratio and may favors jamming or interference.

Thus, true stochastic errors may include: residual biases, randomly multipath imaging due to moving obstacles or to the kinematics of the receiver platform, cycle slips due to the stochasticity of the receiver's way of detection, and random observation errors.

Remember, however, that accumulated delta range is only affected by bias errors, and therefore, it is the technique that gives the most precise positioning calculation.

At the end, sources of error may vary depending on the specific observation and the technique used to reduce noise effects. When using a given modulation technique with the purpose of reducing interference and jamming, there are two advantages: errors become uncorrelated and independent of the observation point. Thus, both advantages simplify the calculation of the total error because now, the total squared error is just the sum of the squared errors. Nevertheless, such calculations ought to be done for the specific user segment and receiver and the differential technique used.

Conclusion

GNSS have been presented as space based systems that cover most of the space-time around the Earth. Such systems have also extended capabilities through Augmentation Systems to send electronic signals that allow position fixing for a single user or a network of users with an extraordinary precision.

Errors were described so one can easily calculate what will be the total error involved in a particular observation or surveying campaign and have a figure of positioning quality.

Cross-References

- ▶ Current and Future GNSS and their Augmentation Systems
- ▶ International Committee on GNSS
- ▶ Introduction to Satellite Navigation Systems

References

C. Audion, *The Measurement of Time: Time, Frequency and The atomic Clock* (Cambridge University Press, Cambridge, 2001)

Y. Bar-Shalom, X. Rong-Li, T. Kirubarajan, *Estimation with Applications to Tracking and Navigation* (Wiley, New York, 2001)

A. Bensky, *Wireless Positioning Technologies and Applications* (Artech House, Boston, 2008)

G. Beutler, *Methods of Celestial Mechanics*, vol. I (Springer, Berlin, 2005)

K. Borre et al., *A Software Defined GPS and Galileo Receiver* (Birkhauser, Boston, 2007)

J. Carnes, *UTM Using your GPS with the Universal Tranverse Mercator Map Coordinate System* (Maptools, Woodside, 2007)

H.C. Chen et al., The performance comparison bewteen GPS and BeiDou-2/COMPASS: a perspective from Asia. J. Chin. Inst. Eng. **32**(5), 679–689 (2009)

China Satellite Navigation Project Centre, COMPASS/Beidou navigation satellite system development. 3rd meeting of the international committee on GNSS, Pasadena, 8–12 Dec 2008

R.C. Dixon, *Spread Spectrum Systems* (Wiley, New York, 1976)

A. Draganov, *GPS Accumulated Delta Range Processing for Navigation Applications*. United States Patent Application Publication. US 2006/0195262 A1, 2006

Global Navigation Satellite System: GLONASS, *Interface Control Document* (Coordination Scientific Information Centre, Moscow, 1998)

B. Guinot, Is the international atomic time TAI a coordinate time or a proper time? Celest. Mech. **28**, 155–161 (1986)

B. Hofmann-Wellenhof, H. Lichtenegger, E. Wasle, *GNSS Global Navigation Satellite Systems* (Springer, Wien/New York, 2008)

B.P. Lahi, Z. Ding, *Modern Digital and Analog Communication Systems*. Oxford Series in Electrical and Computer Engineering (Oxford University Press, Oxford, 2009)

A. Leick, *GPS Satellite Surveying*, 3rd edn. (Wiley, Hoboken, 2004)

D.D. McCarthy, P.K. Seidelmann, *Time: From Earth Rotation to Atomic Physics* (Wiley, Weinheim, 2009)

J. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications, and Control* (Prentice-Hall, Englewood Cliffs, 1995)

G. Mohinder, W. Lawrence, A. Angus, *Global Positioning Systems, Inertial Navigation and Integration* (Wiley, New York, 2001)
D. Muñoz et al., *Position Location Techniques and Applications* (Elsevier, Amsterdam, 2009)
J.G. Proakis, *Digital Communications*, 4th edn. (McGraw-Hill, New York, 2008)
S. Revnivykh, GLONASS sattus and progress. 3rd meeting of the international committee on GNSS (ICG)/UNOOSA, Pasadena, 8–12 Dec 2008
R. Rogers, *Applied Mathematics in Integrated Navigation Systems* (AIAA, Reston, 2003)
RTCM-104, *Recommended Standards for Differential NAVSTAR-GPS services. V 2.1* (Radio Technical Commission for Maritime Services, Washington, DC, 1994)
N. Samama, *Global Positioning* (Wiley, Hoboken, 2008)
G. Seeber, *Satellite Geodesy* (Humbert, Germany, 2003)
M. Shaw et al., U.S. space-based PNT, policy and program review. 3rd meeting of the international committee on GNSS (ICG)/UNOOSA, Pasadena, 8–12 Dec 2008
J.J. Spilker, *Digital Communications by Satellite* (Prentice-Hall, Englewood Cliffs, 1977)
G. Strang, K. Borre, *Linear Algebra, Geodesy, and GPS* (Wellesley-Cambridge, Wellesley, 1997)
B. Triggs et al., Boundle adjustment, a modern synthesis. Calculations, in *Vision Algorithms: Theory and Practice*. Lecture Notes in Computer Science (LNCS), vol. 1883 (Springer, Berlin, 2000), pp. 298–372
P. Verhoef, European GNSS programmes Galileo and EGNOS. 3rd meeting of the international committee on GNSS (ICG)/UNOOSA, Pasadena, 8–12 Dec 2008
J. Wang, Pseudolite applications in positioning and navigation: progress and problems. J. Glob. Position. Syst. **1**(1), 48–56 (2002)
D. Wells, *Guide to GPS Positioning* (Canadian GPS Associates, Fredericton/New Brunswick, 1987)
A. Wilson, *The First Galileo Satellites. Galileo in-orbit Validation Element: GIOVE* (ESA publications Division, Noordwijk, 2006)
D.J. Wither, *Radio Spectrum Management*. IEE telecommunications Series, vol. 45 (The Institution of electrical engineers, London, 2000)
G. Xu, *GPS Theory, Algorithms and Applications* (Springer, Berlin, 2003)
K.S.H. Zigangirov, *Theory of Code Division Multiple Access Communication* (Wiely, Hoboken, 2004)

International Committee on GNSS

24

Sergio Camacho-Lara

Contents

Introduction 604
Establishment of the ICG 605
Membership of the ICG 607
Objectives of the ICG 608
Work of the ICG 609
Working Groups of the ICG 609
ICG Working Group A: Compatibility and Interoperability 609
ICG Working Group B: Enhancement of Performance of GNSS Services 610
ICG Working Group C: Information Dissemination 611
Working Group D: Reference Frames, Timing, and Applications 611
The Future of the GNSS International Coordination Process 612
Providers Forum 613
Objectives of the Providers Forum 614
Cross-References 615
Notes 615
References 615

Abstract

Global navigation satellite systems (GNSS), with their extremely high accuracy, global coverage, all-weather operation, and usefulness at high velocities, are a dual-use technology that are becoming a new global utility that increasingly improve people's daily lives. GNSS applications are growing and their quality is improving in such areas as aviation, maritime and land transportation, mapping and surveying, agriculture, power and telecommunications networks, disaster warning and emergency response, and a host of commercial and social applications.

S. Camacho-Lara
Centro Regional de Enseñanza de Ciencia y Tecnología del Espacio para América Latina y el Caribe (CRECTEALC), Luis Enrique Erro 1, Tonantzintla, Puebla, Mexico
e-mail: sergio.camacho@inaoep.mx

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_12, © Springer Science+Business Media New York 2013

At the turn of the millennium, it became apparent that the two global navigation satellite systems that had existed, the Global Positioning Systems (GPS) of the United States and the Global Navigation Satellite System (GLONASS) of the Russian Federation, would soon be joined by the Galileo system of Europe and the Compass/Beidou of China, as well as by the regional Quasi-Zenith Satellite System (QZZS) of Japan and the Indian Regional Navigation Satellite System (IRNSS) of India. The emergence of new GNSS and regional augmentations focused attention on the need for the coordination of program plans among current and future operators in order to enhance the utility of GNSS services. It also made clear that the providers of GNSS services should pursue greater compatibility and interoperability among all current and future systems in terms of spectrum, signal structures, time, and geodetic reference standards to the maximum extent possible.

Although coordination between the providers of the GNSS was already taking place on a bilateral basis, the desirability of having a forum in which all GNSS providers participated became an attractive idea. Such a forum would allow discussion and coordination on issues of common interest such as protection of the radionavigation spectrum from disruption and interference, global compatibility, and interoperability of space-based position, navigation, and timing services (PNT) that could be used separately or together without interfering with each other. After 1999, and following several years of discussing terms of reference, objectives, and work plan, the International Committee on GNSS (ICG) became such a forum.

Keywords

Argos • Chinese navigation satellite system (Compass/BeiDou) • Compatibility and interoperability of GNSS • European geostationary navigational overlay system (EGNOS) • European Satellite Navigation System (Galileo) • Global navigation satellite system (GNSS) • Global positioning satellite (GPS) • Global Navigation Satellite System (GLONASS) • Indian global positioning system-aided GEO-augmented navigation system (GAGAN) • Indian regional navigation satellite system (IRNSS) • International Committee on GNSS (ICG) • Japanese multi-functional satellite augmentation system (MSAS) • Japanese regional navigation satellite system; Quasi-Zenith satellite system (QZSS) • Position, navigation, and timing services (PNT) • UNISPACE III • United Nations Committee on the Peaceful Uses of Outer Space (COPUOS) • Wide-Area Augmentation System (WAAS)

Introduction

Satellite navigation builds upon terrestrial-based radio navigation that has been used by aviation and shipping over the past 100 years. For a number of years, navigation satellite systems were limited to the Joint United States-France Argos Satellite System, the U.S. Navstar Global

Positioning Satellite System, and the Union of Soviet Socialist Republics/Russian Federation Glonass systems. Today the number, capability, and complexity of the global navigation satellite systems have doubled and continue to expand.

Navigation satellites broadcast signals that are used by a receiver to determine exactly the receiver's position, velocity, and precise time worldwide. User receivers of satellite navigation signals measure the distance of the receiver equipment to the satellite using a technique called "passive ranging." In this technique, the distance to each satellite is derived from the measurement of the time the navigation signal needs to travel from the satellite to the receiver. The three-dimensional position of the receiver can be calculated if signals from at least three satellites are available. The signal from a fourth satellite is used to avoid the need for a precise atomic clock at the receiver.

Standard GNSS signal processing provides around 100-m accuracy at the location of the receiver with four satellites in view, while precision signal processing provides around 20–10-m accuracy. Reference stations make differential GNSS (DGNSS) services possible providing higher time and position accuracies. If, in addition to the signals from the satellites, a user receiver also receives the signal of a ground-based reference station, the accuracy at the location of the user receiver is around 1 m. If in addition, a space-based augmentation system is used as a reference station, the position accuracy increases to the order of centimeters or less.

For the everyday user of the GNSS signal, without interoperability of the systems, the accuracies that can be obtained are limited by the number of satellites of one GNSS that the receiver equipment can view. For mountainous terrain and for urban settings, the number of satellites that can be viewed is likely to be limited as there might not be a direct line of view of the GNSS receiver to one or more of the satellites. This is often the case in canyons or among tall buildings. When full interoperability is achieved between the four global navigation satellite systems and the two regional systems, receivers in mountainous or urban areas will be able to view a large number of satellites at the same time leading to very high accuracies in position and timing measurements. This will be a spectacular result of the work being carried out by the ICG.

Establishment of the ICG

Following the Third United Nations Conference on the Exploration and Peaceful Uses of Outer Space (UNISPACE III), held in 1999, in its resolution 54/68, the United Nations General Assembly endorsed the "Vienna Declaration: Space Millennium for Human Development" (Report of the Third United Nations Conference 1999). The Vienna Declaration called for action, among other matters, to improve the efficiency and security of transport, search and rescue, geodesy and other activities by promoting the enhancement of, universal access to and compatibility of, space-based navigation and positioning systems. In response to that call, in 2001

the United Nations Committee on the Peaceful Uses of Outer Space (COPUOS) established the Action Team on Global Navigation Satellite Systems (GNSS) to carry out those actions under the chairmanship of Italy and the United States.

The Action Team on GNSS, consisting of 38 member States and 15 intergovernmental and nongovernmental organizations, recommended, among other things, that an International Committee on GNSS (ICG) should be established to promote the use of GNSS infrastructure on a global basis and to facilitate exchange of information. The United Nations Committee on the Peaceful Uses of Outer Space included this recommendation in the Plan of Action proposed in its report to the General Assembly on the review of the implementation of the recommendations of UNISPACE III.[1] In 2004, in its resolution 59/2, the General Assembly of the United Nations endorsed the Plan of Action. In the same resolution, the General Assembly invited GNSS and augmentation system providers to consider establishing an ICG in order to maximize the benefits of the use and applications of GNSS to support sustainable development.

At the "United Nations International Meeting for the Establishment of the International Committee on Global Navigation Satellite Systems (ICG)" held on December 1–2, 2005, in Vienna, Austria, the ICG was established on a voluntary basis as an informal body for the purpose of promoting cooperation, as appropriate, on matters of mutual interest related to civil satellite-based positioning, navigation, timing, and value-added services, as well as compatibility and interoperability among the GNSS systems, while increasing their use to support sustainable development, particularly in the developing countries. This fact was noted with appreciation by the General Assembly in its resolution 61/111, of December 14, 2006.

In establishing the ICG, in 2006, the representatives of GNSS core system providers, GNSS augmentation providers (see ▶ Chap. 25, "Current and Future GNSS and Their Augmentation Systems" for a description of the satellite systems), and the international organizations primarily associated with the use of GNSS identified the overlap of GNSS mission objectives and the interdisciplinary nature of applications of GNSS services. Aware that the complexity and cost of user equipment should be reduced whenever possible, the founders of the ICG considered the need to protect the investment of the current user base of GNSS services through the continuation of existing services, particularly thorough greater compatibility and interoperability among all current and future GNSS systems in terms of spectrum, signal structures, time, and geodetic reference standards.

With the above understandings, the representatives of GNSS core system providers, GNSS augmentation providers, and the international organizations primarily associated with the use of GNSS developed terms of reference for the International Committee on GNSS in which they agreed on the objectives of the ICG, its participants (members, associate members and observers), procedures of work, structure, and organization. The terms of reference that were adopted at its meeting in 2008 also provided for their revision on the basis of proposals made by members or associate members and adopted by consensus of the members.



Fig. 24.1 The United Nations Office for Outer Space Affairs, Executive Secretariat of the International Committee on GNSS, is located at the Vienna International Centre, Vienna, Austria (Courtesy of the United Nations Office at Vienna)

The structure of the ICG consists of a chairperson, a plenary session of the Committee, an executive secretariat, and working groups. The chair will rotate on an annual basis among the Members and Associate Members. The Office for Outer Space Affairs, part of United Nations Office at Vienna, is the Executive Secretariat for the ICG and its Provider´s Forum (Fig. 24.1).

Membership of the ICG

The participants in the ICG are the following governments, organizations, and associations:

(a) Members:

Current and future core system providers, including China (Compass/BeiDou Navigation Satellite System (CNSS)), the European Union (European Satellite Navigation System (Galileo)), the Russian Federation (Global Navigation Satellite System (GLONASS)), and the United States of America (Global Positioning System (GPS)); States Members of the United Nations with an active program in implementing or promoting a wide range of GNSS services and applications (Italy, Malaysia, United Arab Emirates); current and future space-based regional or

augmentation system providers including, for example, the European Space Agency (European Geostationary Navigation Overlay Service (EGNOS)), India (GPS and Geostationary Augmented Navigation System (GAGAN) and Indian Regional Navigation Satellite System (INRSS)), Japan (Multi-functional Transport Satellite (MTSAT) Satellite-based Augmentation System (MSAS) and Quasi-Zenith Satellite System (QZSS)), Nigeria (Nigerian Communication Satellite Space Based Augmentation System (NigComsat-1 SBAS)), the Russian Federation (System of Differential Corrections and Monitoring (SDCM)), and the United States (Wide-Area Augmentation System (WAAS))

(b) Associate members:

International and regional organizations and associations dealing with GNSS services and applications, including the Office for Outer Space Affairs of the United Nations Secretariat, the Civil GPS Service Interface Committee (CGSIC), the International Association of Geodesy (IAG), the International Association of Geodesy Reference Frame Sub-Commission for Europe (EUREF), the International Cartographic Association (ICA), the International GNSS Service (IGS, formerly International GPS Service), the International Society for Photogrammetry and Remote Sensing (ISPRS), the International Earth Rotation and Reference Systems Service (IERS), the Fédération internationale des géomètres (FIG), the European PositionDetermination System (EUPOS), and the International Astronautical Federation

(c) Observers:

The Committee on Space Research (COSPAR), the Bureau international des poids et mesures (BIPM), the International Association of Institutes of Navigation (IAIN), the Union radio-scientifique internationale (URSI), the International Telecommunication Union (ITU), and the Interagency Operations Advisory Group (IOAG)

Objectives of the ICG

The objectives of the ICG are to:

(a) Benefit users of GNSS services through consultations among members of the ICG
(b) Encourage coordination among providers of GNSS core systems and augmentations in order to ensure greater compatibility and interoperability
(c) Encourage and promote the introduction and utilization of satellite positioning, navigation, and timing services, particularly in the developing countries through assistance with the integration of GNSS services into their infrastructures
(d) Assist both the members of the ICG and the international user community by, inter alia, serving as the focal point for international information exchange related to GNSS activities, respecting the roles and functions of GNSS service providers and intergovernmental bodies such as the International Telecommunication Union (ITU), the International Civil Aviation Organization (ICAO), and the International Maritime Organization (IMO)

(e) Better address future user needs in the GNSS development plans and applications
(f) Report periodically on its activities to the Committee on the Peaceful Uses of Outer Space

The participants in the ICG agreed that these objectives will be accomplished by an indicative workplan that would be reviewed at meetings of the ICG for accomplishments and updating as might be desirable or necessary.

Work of the ICG

The ICG meets least once every year in plenary session. Meetings of the ICG are organized by the designated host, assisted by the United Nations Office for Outer Space Affairs which acts as the Executive Secretariat of the ICG.

The ICG may establish, as mutually agreed and on an ad hoc basis, working groups to investigate specific areas of interest, cooperation, and coordination. The chairpersons of such working groups report at each plenary session on accomplishments and future plans.

All recommendations of the ICG or its working groups are decided on the basis of consensus of its members, do not create legal obligations, and will be acted upon at the discretion of each member, associate member, or observer.

The ICG's indicative workplan contains the following broad scope elements:
(a) Compatibility and interoperability
(b) Enhancement of performance of GNSS services
(c) Information dissemination
(d) Reference frames, timing, and applications
(e) Coordination

Working Groups of the ICG

In order to carry out the work associated with items (a) to (d) of the above listed elements throughout the period between meetings of the ICG, the Committee decided to establish Working Groups A–D, assigning to them specific tasks. The Working Groups report and make recommendations to the ICG on a yearly basis. Recommendations accepted by consensus in the ICG can lead to additional items on the workplan of the Working Groups. The initial workplans of the Working Groups and the highlights and modifications to their workplans as agreed at the sixth meeting of the ICG, held in Tokyo, Japan, in 2011, are presented below.

ICG Working Group A: Compatibility and Interoperability

As compatibility and interoperability are highly dependent on the establishment of standards for service provision and user equipment, the ICG decided to address the

topic of the adoption of common guidelines. However, the ICG would not itself set guidelines; instead, it will identify applications where no guidelines currently exist and recommend possible organizations that could appropriately set new guidelines. As required, consultation with existing standard-setting bodies, such as the International Civil Aviation Organization (ICAO), the International Maritime Organization (IMO), the International Telecommunication Union (ITU), and the International Organization for Standardization (ISO), is carried out.

Among the first actions assigned to Working Group A was the establishment of a Providers Forum to enhance compatibility and interoperability among current and future global and regional space-based systems.

In 2011, the Working Group A on compatibility and interoperability addressed all four areas of its current work plan. Interference detection and mitigation, and open service provision and performance monitoring by multi-GNSS networks were the major areas of focus, leading to three of the working group's four recommendations.[2] The session on multi-GNSS monitoring was held jointly with Working Groups B and D, as was the session on interoperability. This resulted in constructive dialogue with these working groups and an agreed plan of practical steps including establishment of a subgroup to collectively investigate international GNSS monitoring and assessment.

ICG Working Group B: Enhancement of Performance of GNSS Services

As a unique combination of GNSS service providers and major user groups, the ICG is exceptionally placed to promote and coordinate activities aimed at enhancing GNSS performance, recommending system enhancements, and meeting future user needs.

Among the first actions assigned to Working Group B were the development of a reference document on models and algorithms for ionospheric and tropospheric corrections, the examination of the problem of multi-path and related mitigation actions affecting both GNSS systems and user receivers, especially for mobile receivers and the extension of GNSS service to indoor applications.

In 2011, the Working Group B discussed, among other aspects, the dissemination of disaster information. The Working Group concluded that satellite navigation systems may provide essential contributions but the service concept still needs further elaboration. Due to the importance of this issue a new work item was introduced in the work plan of the Group on the basis of seven recommendations that were approved by the ICG.[3] In addition, at the Sixth Meeting of the ICG, the existing actions in the work plan were confirmed and good progress was shown in various areas of the Working Group's workplan, including indoor positioning, signal authentication, precise positioning, transportation, maritime and space applications. Since more and more application-related issues are introduced and discussed within Working Group B, it was agreed to form a dedicated subgroup on applications.

ICG Working Group C: Information Dissemination

A great many people, in their work environments and everyday life, already rely on GNSS products. The awareness of the improvements that are resulting both from better technology in receiver equipment and in the signal that is being provided, coupled with greater knowledge on the use the GNSS signal, will improve the benefits that can be derived in a large range of every day work-life areas. To support greater awareness and knowledge of GNSS and of the use of their signals, the ICG is promoting the establishment of user information centers by GNSS providers.

Among the first actions assigned to Working Group C were the establishment of a GNSS web information portal drawing on contributions from Members, Associate Members, and Observers of the ICG,[4] to consider the use of the Regional Centres for Space Science and Technology Education, affiliated to the United Nations located in Africa (Nigeria and Morocco), Asia and the Pacific (India) and Latin America and the Caribbean (Brazil and Mexico), to promote GNSS use and applications as well as the development of a GNSS curricula that could be introduced at these Centers and other institutions of higher education. At the Sixth Meeting of the ICG, held in September 2011 in Tokyo, Japan, the Provider's Forum invited its members to make proposals for an updated design of the interim ICG Information Portal.

In 2011, the Working Group C on information dissemination and capacity-building addressed further aspects of its workplan, including training for capacity-building in developing countries, distance learning programs, web-based courses and tutorials, interactive programs for middle/high schools, multimedia softwares and demonstration data sets to enrich the training and research, programs promoting the use of GNSS technologies as tools for scientific applications; the International Space Weather Initiative; and regional workshops on applications of GNSS. A new item on education and training programs on GNSS was added to the workplan.[5]

Working Group D: Reference Frames, Timing, and Applications

The ICG is establishing links with national and regional authorities and relevant international organizations, particularly in developing countries.On the basis of these links, the ICG organizes and sponsors regional workshops and other types of activities in order to fulfill its objectives.

Among the first actions assigned to Working Group D was the definition of minimum operational performance standards for GNSS performance monitoring networks, the establishment of working groups focused on: (1) Site Quality, Integrity and Interference Monitoring (SQII); (2) developing a strategy for support by the International Committee of regional reference systems (e.g., the African Geodetic Reference Framework (AFREF), the European Position Determination System (EUPOS), the IAG Reference Frame Sub-Commission for Europe (EUREF), and the Geocentric Reference System for the Americas (SIRGAS));

and (3) developing a strategy for support by the ICG of mechanisms to detect and mitigate sources of electromagnetic interference, taking existing regulatory mechanisms into consideration.

In 2011, The Working Group D completed the development of templates describing the geodetic and timing references for the navigation satellite systems currently represented in the ICG. The Working Group proposed that the templates be published on the ICG Information Portal. ICG also welcomed progress by the Bureau international des poids et mesures (BIPM) toward production of the "Rapid Coordinated Universal Time (UTC)" as a more immediately accessible time reference that could be used to better harmonize the UTC broadcast by each GNSS. The Working Group recommended that interested System Providers supply data from their respective Monitor Stations for inclusion in regular processing with the IGS network of reference stations. Such inclusion is aimed at improving the alignment of the various GNSS reference frames with each other and with the International Terrestrial Reference Frame. Working Group D also recommended that the ICG support and endorse the IGS Multi-GNSS Global Experiment (IGS M-GEX) and actively encourage participation and/or contributions from, among others, GNSS providers, international organizations related to GNSS, entities involved in timing, navigation, aviation, transportation, GIS, and relevant fields, including national mapping agencies, space agencies, universities and research institutions, as well as industry receiver manufacturers and service providers.[6]

The Future of the GNSS International Coordination Process

Because of its membership composition, the ICG brings together the providers of the global and regional navigation satellite systems, the large professional associations of users and the international organizations that have a mandate to regulate the use of the GNSS spectrum. These are the key actors that need to discuss issues of compatibility and interoperability among the systems and the protection of the GNSS frequency spectrum in order to provide better and more cost-effective services to all.

In the future, the ICG will consider, make recommendations, and agree on actions to promote appropriate coordination across GNSS programs. Furthermore, the ICG encourages its Members, Associate Members, and Observers to maintain communication, as appropriate, with other groups and organizations involved in GNSS activities and applications through the relevant channels within their respective Governments and organizations.

The ICG could also support the establishment of national and/or regional planning groups for GNSS that would address regulations associated with the use of GNSS services and suggest organizational models to use at the national level for co-coordinating and governing GNSS use.

In addition to the coordination provided by the ICG plenary, high-level coordination is carried out by the Provider's Forum, established at the recommendation of the ICG's Working Group A.

Providers Forum

The Providers Forum was established at the second meeting of the ICG in 2007, in Bangalore, India, with the aim to promote greater compatibility and interoperability among current and future providers of the Global Navigation Satellite Systems (GNSS). The current members of the Providers Forum, including China, India, Japan, the European Union, the Russian Federation, and the United States, addressed key issues such as ensuring protection of GNSS spectrum and matters related to orbital debris/orbit de-confliction.

Global and regional system providers, members of the Providers Forum, at the third meeting of the ICG held in 2008, in Pasadena, United States, agreed that at a minimum, all global navigation satellite systems (GNSS) signals and services must be compatible. To the maximum extent possible, open signals and services should also be interoperable, in order to maximize benefit to all GNSS users. For many applications, common carrier frequencies are essential to interoperability and commonality of other signal characteristics is desirable. In some cases, carrier frequency diversity may be preferable to improve performance. The Providers Forum will continue to investigate the benefits of carrier frequency commonality and diversity, as well as of compatibility and interoperability, as these latter terms and desired outcomes are defined below:

Interoperability refers to the ability of global and regional navigation satellite systems and augmentations and the services they provide to be used together to provide better capabilities at the user level than would be achieved by relying solely on the open signals of one system.

1. Interoperability allows navigation with signals from different systems with minimal additional receiver cost or complexity.
2. Multiple constellations broadcasting interoperable open signals will result in improved observed geometry, increasing end-user accuracy everywhere, and improving service availability in environments where satellite visibility is often obscured.
3. Geodetic reference frames' realization and system time standards should adhere to existing international standards to the maximum extent practical.
4. Any additional solutions to improve interoperability should be encouraged.

Compatibility refers to the ability of global and regional navigation satellite systems and augmentations to be used separately or together without causing unacceptable interference and/or other harm to an individual system and/or service.

1. The International Telecommunication Union provides a framework for discussions on radiofrequency compatibility. Radiofrequency compatibility should involve thorough consideration of detailed technical factors, including effects on receiver noise floor and cross-correlation between interfering and desired signals.
2. Compatibility should also respect spectral separation between each system's authorized service signals and other systems' signals. Recognizing that some

signal overlap may be unavoidable, discussions among providers concerned will establish the framework for determining a mutually acceptable solution.
3. Any additional solutions to improve compatibility should be encouraged.

Objectives of the Providers Forum

To achieve the desired compatibility and interoperability of the GNSS, the providers of the systems agreed on their own Terms of Reference,[7] which have been updated to adapt to evolving circumstances, including a statement of objectives of the Providers Forum, its membership, and work procedures, structure, and organization. The objectives of the Providers Forum are to:

(a) Promote compatibility and interoperability among current and future global and regional space-based systems by exchanging detailed information about planned or operating systems and the policies and procedures that govern their service provision, consistent with the template for information sharing among providers that was circulated prior to the first meeting
(b) Act as a mechanism to continue discussions on important issues addressed by the ICG that require focused inputs from system providers

The Providers Forum is not a policy-making body, but provides a means to promote discussion among system providers based on agreed guidelines for provision of open services, including transparency, cooperation, performance monitoring, and spectrum protection, and agreed principles for ensuring compatibility and interoperability among systems.

Conclusion

With new systems and increasing of new frequency bands and applications of new services, more and more space systems, there are increasing demands on use of the frequency spectrum around the bands utilized by the GNSS. One should expect this trend to continue. An uncoordinated approach in which the use frequency bands are awarded to industry or government institutions based only on considerations of the use of one or two space systems or on the benefits to a specific application could result in an adverse impact to the signal reception of all GNSS. Since GNSS in its civil application is a service-oriented system, which is aimed at serving all humankind, this issue warrants significant attention across the world.

It is precisely these types of issues that are being dealt with by the International Committee on GNSS. There are many GNSS events around the world on a yearly basis, each contributing to one or more aspects that promote and strengthen the use of GNSS in a myriad of applications. However, the meetings of the ICG and its Provider Forum and the intersessional work carried out by their Working Groups are the unique mechanisms where the operators of GPS, GLONASS, Galileo, Compass/BeiDou, IRNSS, and QZSS can identify issues that could adversely affect the use of the GNSS.

The issues addressed at the ICG include whether frequency coordination related to satellite navigation satellites should be conducted only on a bilateral basis, as has been the case, or whether frequency coordination should also be a multilateral process. Another key area dealt with by the ICG are the practical and technical issues of how to separate civil- and defense-related services in the context of signaling, frequency assignments, and transmission requirements. With a membership that includes all GNSS operators and the large user associations, the ICG is the best, and perhaps the only, place to discuss and resolve complex situations that could prevent the achievement of compatibility and interoperability among the GNSS to the detriment of current and future users of GNSS.

Cross-References

- ▶ Current and Future GNSS and Their Augmentation Systems
- ▶ Global Navigation Satellite Systems: Orbital Parameters, Time and Space Reference Systems and Signal Structures
- ▶ Introduction to Satellite Navigation Systems

Notes

1. Report of the Action Team on Global Navigation Satellite Systems (GNSS): Follow-up to the Third United Nations Conference on the Exploration and Peaceful Uses of Outer Space (UNISPACE III) (United Nations publication, Sales No. E.05.I.3).
2. http://www.unoosa.org/pdf/icg/2011/icg-6/wgA/ICG_WGA_2011.pdf
3. http://www.unoosa.org/pdf/icg/2011/icg-6/wgB/ICG_WGB_2011.pdf
4. A GNSS web portal with information on ICG matters has been established by the United Nations Office for Outer Space Affairs at http://unoosa.org/oosa/en/SAP/gnss/icg.html.
5. http://www.unoosa.org/pdf/icg/2011/icg-6/wgC/ICG_WGC_2011.pdf
6. http://www.unoosa.org/pdf/icg/2011/icg-6/WgD/ICG_WGD_2011.pdf
7. http://unoosa.org/oosa/en/SAP/gnss/icg/providersforum.html

References

Report of the Third United Nations Conference on the Exploration and Peaceful Uses of Outer Space, Vienna, 19-30 July 1999 (United Nations publication, Sales No. E.00.I.3), chap. I, resolution 1

Current and Future GNSS and Their Augmentation Systems

25

Sergio Camacho-Lara

Contents

Introduction 619
The Global Positioning System (GPS) of the United States 620
Global Positioning System Satellite Constellation 620
Current and Future Satellite Generations 621
Current and Planned GPS Signals 622
Signal-in-Space Health 625
Services Provided and Provision Policies 625
United States Space-Based Positioning, Navigation, and Timing Policy 626
Wide-Area Augmentation System 626
International Cooperation to Ensure Compatibility and Pursue Interoperability 627
The Global Navigation Satellite System (GLONASS) of the Russian Federation 627
The GLONASS – System Description 628
Current and Future Satellite Generations 628
GLONASS First Generation Satellites 629
GLONASS Second Generation Satellites 629
GLONASS Third Generation Satellites 630
Current and Planned GLONASS Signals 631
Performance Standards Versus Actual Performance 632
Services Provided and Provision Policies 633
System for Differential Correction and Monitoring 634
International Cooperation to Ensure Compatibility and Pursue Interoperability 635
The Compass/Beidou Global Navigation Satellite System of China 635
Space Segment System Description 635
Compass Second Generation Satellites 636
Current and Planned Beidou-2 Signals 637
Performance Standards Versus Actual Performance 638
Services Provided and Provision Policies 639
Perspective on Compatibility and Interoperability 639

S. Camacho-Lara
Centro Regional de Enseñanza de Ciencia y Tecnología del Espacio para América Latina y el Caribe (CRECTEALC), Luis Enrique Erro 1, Tonantzintla, Puebla, Mexico
e-mail: sergio.camacho@inaoep.mx

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_25, © Springer Science+Business Media New York 2013

The European Satellite Navigation System (Galileo) 640
Space Segment 640
Current and Future Satellite Generations 641
Current and Planned Galileo Signals 644
Performance Standards Versus Actual Performance 645
Services Provided and Provision Policies 645
European Geostationary Navigation Overlay Service (EGNOS) 646
Perspective on Compatibility and Interoperability 646
The Multifunctional Transport Satellite (MTSAT) Satellite-Based Augmentation System (MSAS) and the Quasi-Zenith Satellite System (QZSS) of Japan 647
Description of MSAS 647
Description of QZSS 649
The Indian Regional Navigation Satellite System (IRNSS) and the Global Positioning System-Aided GEO-Augmented Navigation (GAGAN) System 651
System Description 652
Current and Planned Signals 652
GPS-Aided GEO-Augmented Navigation System (GAGAN) 653
Services Provided and Provision Policies 653
Cross-References 654
Notes 654
References 654

Abstract

Global Navigation Satellite System (GNSS) is the standard generic term for satellite navigation systems that provide autonomous geospatial positioning with global coverage. GNSS allows small electronic receivers to determine their location (longitude, latitude, and altitude) to within a few meters using time signals transmitted along a line-of-sight by radio from satellites. Receivers on the ground, air, or water calculate the precise time as well as position, which can be used as a reference for scientific experiments and numerous everyday applications.

As of 2012, the Navstar Global Positioning System (GPS) of the United States and the Global Navigation Satellite System (GLONASS) of the Russian Federation are the only fully operational global GNSS. The European Union's Galileo positioning system is a GNSS in the initial deployment phase, scheduled to be operational in 2014. The People's Republic of China has decided to expand its regional BeiDou/Compass navigation system into a complete global navigation system by 2015 although, with 13 satellites in orbit, it already has limited global coverage. The global coverage for each system is generally achieved by a constellation of 24–30 Medium Earth Orbit (MEO) satellites distributed between several orbital planes. The actual systems vary, but use orbit inclinations greater than 50° and orbital periods of roughly 12 h (height 20,000 km / 12,500 miles). These global systems are being joined by the regional Quasi-Zenith Satellite System (QZZS) of Japan and the Indian Regional Navigation Satellite System (IRNSS) of India. These regional systems utilize satellites at smaller inclinations in elliptical orbits with apogees around 24,000 and 39,000 km or in inclined geostationary orbits at around 36,000 km. As accuracy in position, time, or speed measurements increases with the number of satellites that can be observed by a receiver, the signals received

from the global from GNSS satellites are complemented by signals provided by satellite-based augmentations systems (SBAs). Such is the motivation for the Wide-Area Augmentation System of the United States, the System for Differential Correction and Monitoring (SDCM) of the Russian Federation, the European Geostationary Navigation Overlay Service (EGNOS), the GPS and Geo-Augmented Navigation system (GAGAN) of India, and the Multifunctional Transport Satellite (MTSAT) Satellite-based Augmentation System (MSAS) of Japan. Altogether, by 2020 there will be around 120 navigation and positioning satellites in orbit at any given moment. It is possible that a user could receive signals from as many at ten satellites, leading to accuracies only available at the research level today. This chapter presents the characteristics of all current and future generations of navigation and positioning satellites.

Keywords

GNSS • GPS • GLONASS • Galileo • Compass/BeiDou • IRNSS • QZSS • WAAS • SDCM • EGNOS • GAGAN • MSAS • Positioning • Navigation and Timing (PNT) system • Location Based Service • Satellite-Based Augmentation System (SBA) • Compatibility and Interoperability • GNSS user services and policies • Standard Positioning Service (SPS)

Introduction

The Global Navigation Satellite Systems (GNSS) have grown into a global utility whose multiuse services are integral to national security of the providers of the systems as well to global security, economic growth, transportation safety, search and rescue activities, and scientific research. As such, the GNSS capabilities are an essential element of the worldwide economic and social infrastructure.

The existing and the soon to be built GNSS will transmit both open service signals and restricted signals (encrypted) for military or restricted commercial services. With four global position, navigation, and timing (PNT) systems that will be fully operational by 2020 and at least five operational satellite-based augmentation systems, the issues of compatibility and interoperability among the systems become of paramount importance both for the providers and the users of the systems. This means that the signals of the various systems should not interfere with each other and that, to the extent possible, a receiver could use the signal of more than one GNSS.

Compatibility refers to the ability of multiple satellite navigation systems to be used separately or together, without interfering with the navigation performance of any of the various systems. Interoperability refers to the ability of the open services of multiple satellite navigation systems to be used together to provide better capabilities at the user level than would be achieved by relying solely on one service, without significantly increasing the complexity and cost of the receivers.

This chapter presents descriptions of the satellite constellations of the currently operating GNSS and of those under construction as well as of their augmentation systems. The current and future signal frequency bands of the open and restricted

service and other characteristics of the GNSS and their augmentation systems are provided. The services that each GNSS will provide and their policy for users are also indicated. The reader should note that through their participation in the "International Committee on GNSS" (see ▶ Chap. 24), all providers have agreed to the definition of compatibility and interoperability as stated in the previous paragraph and have stated their commitment to attempt to achieve compatibility and interoperability with other systems through bilateral and multilateral coordination. This is an excellent goal but as can be seen in the description of the current frequency bands of the BeiDou/Compass and Galileo systems, it is difficult to achieve.

The Global Positioning System (GPS) of the United States

The Navstar Global Positioning System, hereafter referred to as GPS, is a space-based radio navigation system owned by the United States Government (USG) and operated by the United States Air Force. GPS has provided positioning, navigation, and timing services to military and civilian users on a continuous worldwide basis since first launch in 1978. An unlimited number of users with a civil or military GPS receiver can determine accurate time and location, in any weather, day or night, anywhere in the world.

In an effort to ensure beneficial services are available to the greatest number of users without degrading security interests, two GPS services are provided. The Precise Positioning Service (PPS) is available primarily to the military of the United States and its allies for users properly equipped with PPS receivers. The Standard Positioning Service (SPS), as initially described in the *SPS Signal Specification* (see below), was originally designed to provide civil users with a less accurate positioning capability than PPS, through a feature known as Selective Availability (SA). In view of technological advancements, the USG has committed to maintain the discontinuance of the SA feature to degrade globally the SPS. The US President announced in 2007 that Selective Availability will not be built into modernized GPS III satellites. The United States National Space Policy released in June 2010 reaffirms the long-standing and stable US policy on space-based navigation.

Global Positioning System Satellite Constellation

The Global Positioning System (GPS) Standard Positioning Service (SPS) consists of space-based positioning, navigation, and timing (PNT) signals delivered free of direct user fees for peaceful civil, commercial, and scientific uses worldwide.

The GPS baseline constellation consists of 24 slots in six orbital planes, with four slots per plane. Three of the slots are expandable and can hold no more than two satellites. Satellites that are not occupying a defined slot in the GPS constellation occupy other locations in the six orbital planes. This 24-slot arrangement ensures there are at least four satellites in view from virtually any point on the planet. The Air Force normally flies more than 24 GPS satellites to maintain coverage whenever the baseline satellites are serviced or decommissioned. The extra satellites may

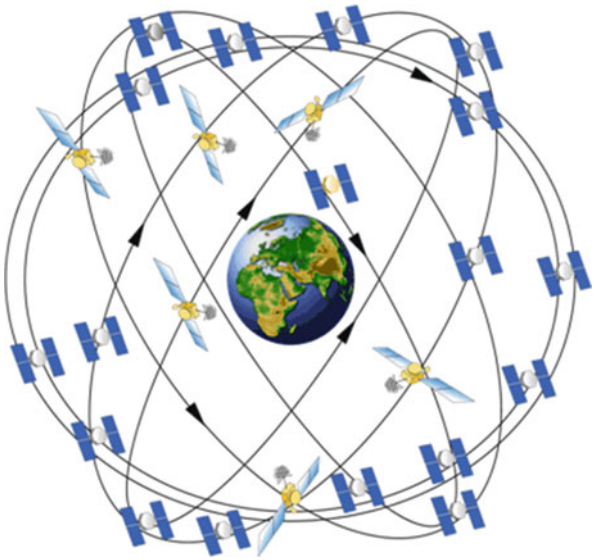


Fig. 25.1 Schematic of the GPS Constellation (Courtesy National Coordination Office for Space-based PNT)

increase GPS performance but are not considered part of the core constellation. Figure 25.1 shows a schematic of the GPS constellation of satellites.

The constellation is being reconfigured to an expanded 24 baseline operational satellites and spare, also operational, to provide better coverage and availability around the world. As of 29 November 2010, GPS had 31 operational satellites in orbit to ensure a baseline constellation of 24 satellites plus three spares. Four additional satellites are in residual status and could be reactivated if one of the currently operational satellites experienced a sudden breakdown. Eight of the Block IIR-M satellites and one new Block IIF satellite are broadcasting a second civil signal called L2C. As of January 2012, two IIF satellite are also broadcasting a new civil signal at L5, which eventually will be used for safety-of-life applications (see below).

Current and Future Satellite Generations

The GPS constellation is a mix of new and legacy satellites. The following text describes the various generations, or blocks, of GPS satellites that are currently flying. It also describes the satellites under development as part of the GPS modernization program.

Block IIA is an upgraded version of the GPS Block II satellites launched in 1989–1990. The "II" refers to the second generation of GPS satellites, although Block II was actually the first series of operational GPS satellites. The "A" stands for *advanced* (Fig. 25.2).

The IIR series were produced to replace the II/IIA series as the II/IIA satellites gradually degraded or exceeded their intended design life. The "R" in Block IIR stands for *replenishment*.



Fig. 25.2 A Block IIA satellite (Courtesy National Coordination Office for Space-based PNT)

The IIR(M) series of satellites are an upgraded version of the IIR series, completing the backbone of today's GPS constellation. The "M" in IIR(M) stands for *modernized*, referring to the new civil and military GPS signals added with this generation of spacecraft. As of February 2012, there were 12 IIR satellites in the GPS constellation, forming the backbone of today's GPS along with seven healthy IIR(M) satellites in the constellation.

The Block IIF series expand on the capabilities of the IIR(M) series with the addition of a third civil signal in a frequency protected for safety-of-life transportation. The "F" in IIF stands for *follow-on*. Compared to previous generations, GPS IIF satellites have a longer life expectancy and a higher accuracy requirement. Each spacecraft uses a mix of rubidium and cesium atomic clocks to keep time within 8 billionths of a second per day. As mentioned above, as of January 2012 there are two operational IIF satellites in the GPS constellation (Fig. 25.3).

Currently under development, the GPS III series is the newest block of GPS satellites. GPS III will provide more powerful signals in addition to enhanced signal reliability, accuracy, and integrity – all of which will support precision, navigation, and timing services. Based on the current contracts and funding, four GPS III satellites will be produced with options to purchase an additional eight satellites. Future versions will feature increased capabilities to meet demands of military and civilian users alike.

Current and Planned GPS Signals

The modernization program of the US Global Positioning System (GPS) continued with the first new Block IIF satellite launched in 2010 and now operating normally,



Fig. 25.3 A Block IIF satellite (Courtesy National Coordination Office for Space-based PNT)

broadcasting a new civil signal, L5, in addition to other civil signals: L2C and L1 C/A. These signals, which are transmitted at specific frequencies, are described below.

GPS L1

The L1 frequency, transmitted by all GPS satellites, contains a course/acquisition (C/A) code ranging signal with a navigation data message that is available for peaceful civilian, commercial, and scientific use, and a precision P(Y) code ranging signal with a navigation data message available to users with valid cryptographic keys. GPS satellites also transmit a second P(Y) code ranging signal with a navigation data message on the L2 frequency.

GPS L2C

The second civil signal, known as "L2C," has been designed specifically to meet commercial needs. When combined with L1 C/A in a dual-frequency receiver, the L2C signal enables ionospheric correction, improving accuracy. For professional users with existing dual-frequency operations, L2C signals deliver faster signal acquisition, enhanced reliability, and greater operating range for differential applications. The L2C modulation also results in a signal that is easier to receive under trees and even indoors. This also supports the further miniaturization of low-power GPS chipsets for mobile applications.

The first GPS IIR-M satellite featuring L2C capabilities was launched in 2005. Every GPS satellite fielded since then has included an L2C transmitter. As at January 2010, there were seven GPS satellites broadcasting L2C signals. In June 2011, the Air Force successfully completed a GPS constellation expansion known as the "Expandable 24" configuration. Three of the 24 slots were expanded, and six satellites were repositioned, so that three of the extra satellites became part of the constellation baseline. As a result, GPS now effectively operates as a 27-slot

constellation with improved coverage in most parts of the world. Interface specification information for the L2C signal can be found on the website of the Los Angeles Air Force Base.[1]

GPS L5

The third civil signal, known as "L5," is broadcast in a radio band reserved exclusively for aviation safety services and radio navigation satellite services. With a protected spectrum, higher power, greater bandwidth, and other features, the L5 signal is designed to support safety-of-life transportation and other high-performance applications. Future aircraft will use L5 signals in combination with L1 C/A (also in a protected band) to improve accuracy via ionospheric correction and robustness via signal redundancy. The use of L5 signals will increase capacity, fuel efficiency, and safety in airspace, railroads, waterways, and highways. When used in combination with L1 C/A and L2C, L5 will provide a very robust service that may enable submeter accuracy without augmentations and very long-range operations with augmentations. The operational L5 signal is available with the follow-on series of GPS satellites, Block IIF, beginning in 2010. Interface specification information on the L5 signal can also be found on the website of the Los Angeles Air Force Base.[2]

GPS L1C

The fourth civil signal, known as "L1C," has been designed to enhance interoperability between GPS and international satellite navigation systems. The United States and the European Union originally developed L1C as a common civil signal for GPS and the European Satellite Navigation System (Galileo). It features a multiplexed binary offset carrier (MBOC) waveform designed to improve mobile reception in cities and other challenging environments. Other satellite navigation systems, such as Japan's Quasi-Zenith Satellite System (QZSS) and China's Compass/BeiDou system, also plan to broadcast signals similar to L1C, to enhance interoperability with GPS. The United States will launch its first L1C signal with GPS III in 2014. L1C will broadcast at the same frequency as the original L1 C/A signal, which will be retained for backward compatibility. Interface specification information for the L1C signal can be found on the website of the Los Angeles Air Force Base.[3]

Although GPS will provide operationally three new modernized civil signals in the future: L2C, L5, and L1C, the performance specifications in the Signal Performance Standards (SPS) apply only to users of the L1 (1575.42 MHz) Coarse/Acquisition (C/A) signal, as this is the only civil GPS signal that has reached full operational capability at this time.

The modernized GPS constellation has been performing at very high accuracy levels – the 1 year performance as of July 2010 provided a user range error of one-half meter, the best ever. The reliability of the constellation has been enhanced by solid performance from the Block IIR and IIR-M satellites which have solar array and power capacity that far exceeds the specified mean mission duration and there have been no clock failures in these satellites to date. The first Block IIF satellite was declared operational in August 2010 and the second one launched in January 2012. Ten more Block IIFs are in the pipeline with satellites three through five already in production.

Signal-in-Space Health

For accurate and trustworthy measurements of position, velocity, and timing, it is important for the user to know that the satellite sending the signal is functioning properly. The SPS Signal-in-Space (SPS SIS) health is the status given by the real-time health-related information broadcast by each satellite as an integral part of the SPS SIS. The SPS SIS health is also sometimes referred to as "satellite health" or "space vehicle health" or "SV health." For this standard, there are three possible SPS SIS health conditions: "healthy," "marginal," "unhealthy." The mapping of the real-time health-related information broadcast by the satellite to these three conditions is given as follows.

The SPS SIS accuracy is described in two statistical ways; one way is as the 95th percentile (95%) SPS SIS user range error (URE) at a specified age of data (AOD), the other is as the 95% SPS SIS URE over all AODs. With either statistical expression, the SPS SIS accuracy is also known as the SPS SIS pseudo-range accuracy. Other accuracy-related SPS SIS performance parameters include the SPS SIS pseudo-range rate (velocity) accuracy defined as the 95% SPS SIS pseudo-range rate error over all AODs and the SPS SIS pseudo-range acceleration (rate rate) accuracy defined as the 95% SPS SIS pseudo-range acceleration error over all AODs.

The SPS SIS integrity is defined to be the trust which can be placed in the correctness of the information provided by the SPS SIS. SPS SIS integrity includes the ability of the SPS SIS to provide timely alerts to receivers when the SPS SIS should not be used for positioning or timing. The SPS SIS should not be used when it is providing misleading signal-in-space information (MSI), where the threshold for "misleading" is a not-to-exceed (NTE) tolerance on the SIS URE. For this *SPS PS*, the four components of integrity are the probability of a major service failure, the time to alert, the SIS URE NTE tolerance, and the alert (either one or the other of two types of alerts).

It is recognized that GPS receivers cannot always monitor the broadcast NAV message data since interruptions may be caused by temporary signal blockages, abnormal receiving antenna orientation, radio frequency interference (particularly jamming), and intermittent environmental effects. Although the GPS receiver is responsible for taking appropriate action when it cannot monitor, process, or apply the current real-time health-related information in the NAV message data, it is possible for the Control Segment to aid some GPS receivers by giving them some advance warning of impending SPS SIS health changes. This action will only be beneficial for SPS SIS integrity if the SPS SIS health changes from healthy to marginal or from healthy to unhealthy.

Services Provided and Provision Policies

GPS provides two levels of service: a standard positioning service, which uses the C/A code on the L1 frequency, and a precise positioning service, which uses the C/A code on the L1 frequency and the P(Y) code on both the L1 and L2 frequencies. Authorized access to the precise positioning service is restricted to the United States Armed

Forces, federal agencies, and selected allied armed forces and governments. The standard positioning service is available to all users worldwide on a continuous basis and without any direct user charge. The specific capabilities provided by the GPS open service are published in the GPS Standard Positioning Service Performance Standards. The United States Department of Defense, as the operator of GPS, will continue enabling codeless/ semi-codeless GPS access until 31 December 2020, by which time the L2C and L5 signals will be available on at least 24 modernized GPS satellites.

United States Space-Based Positioning, Navigation, and Timing Policy

The current United States space-based positioning, navigation, and timing policy, signed by the President in July of 2010, states that the United States must maintain its leadership in the service, provision, and use of global navigation satellite systems (GNSS). To this end, the United States shall take the following steps:

- Provide continuous worldwide access, for peaceful civil uses, to the Global Positioning System (GPS) and its government-provided augmentations, free of direct user charges.
- Engage with foreign GNSS providers to encourage compatibility and interoperability, promote transparency in civil service provision, and enable market access for US industry.
- Operate and maintain the GPS constellation to satisfy civil and national security needs, consistent with published performance standards and interface specifications. Foreign positioning, navigation, and timing (PNT) services may be used to augment and strengthen the resiliency of GPS.
- Invest in domestic capabilities and support international activities to detect, mitigate, and increase resiliency to harmful interference to GPS, and identify and implement, as necessary and appropriate, redundant and back-up systems or approaches for critical infrastructure, key resources, and mission-essential functions.

The policy promotes the global use of GPS technology through the following key provisions:

- No direct user fees for civil GPS services
- Open and free access to the information necessary to develop and build equipment
- Performance improvements for United States space-based positioning, navigation, and timing services
- Promotion of GPS standards
- International compatibility and interoperability for the benefit of end users
- Protection of the radio navigation spectrum from disruption and interference

Wide-Area Augmentation System

A satellite-based augmentation system (SBAS) is a system that supports wide-area or regional augmentation through the use of additional satellite-broadcast

messages. The Wide-Area Augmentation System (WAAS) of the United States is an SBAS developed by the Federal Aviation Administration to augment the Global Positioning System (GPS), with the goal of improving its accuracy, integrity, and availability. Essentially, WAAS is intended to enable aircraft to rely on GPS for all phases of flight, including precision approaches to any airport within its coverage area. Such systems include multiple ground stations, located at accurately surveyed points. The ground stations take position measurements of one or more of the GNSS satellites, the satellite signals, or other environmental factors which may affect the signal received by the users. Using these measurements, information messages are created and sent to one or more satellites for broadcast to the end users.

WAAS currently relies on the service of two leased geostationary satellites positioned at 107° W latitude and 133° W longitude. On 3 April 2010, the telemetry tracking and control system on the Intelsat satellite (positioned at 133° W longitude) failed. Mitigation efforts are under way to ensure that dual coverage requirements are met over the long term. The objective of this System is to provide a user receiver with at least two geostationary satellites in view during localizer performance vertical operations.

There are 38 wide-area reference stations throughout North America (in Canada, Mexico, and the United States, including Alaska and Hawaii) and Puerto Rico. The Federal Aviation Administration of the United States plans to upgrade the wide-area reference stations with receivers capable of processing the new GPS L5 signal.

International Cooperation to Ensure Compatibility and Pursue Interoperability

In addition to participating in the International Committee on GNSS (ICG), the United States pursues its international GNSS coordination objectives in many other ways. These include working through and with the Asia-Pacific Economic Cooperation forum, as well as standard-setting bodies such as relevant United Nations specialized agencies, the International Telecommunication Union (ITU), the International Civil Aviation Organization (ICAO), and the International Maritime Organization. Finally, the United States also pursues bilateral cooperation with other system providers.

The Global Navigation Satellite System (GLONASS) of the Russian Federation

The Global Navigation Satellite System, hereafter referred to as GLONASS is a space-based radio navigation system owned by the government of the Russian Federation and operated for the Russian government by the Russian Aerospace Defense Forces. It both complements and provides an alternative to the United States' Global Positioning System (GPS) and is currently the only alternative navigational system in operation with global coverage and similar precision. It is possible to purchase GNSS receivers that can acquire both the GPS and the GLONASS frequencies.

The Soviet Union military identified, in the late 1960s a need for a Satellite Radio Navigation System (SRNS) for use in precision guidance of a new generation of ballistic missiles. The existing Tsiklon satellite navigation system required several minutes of observation by the receiving station to fix a position making them unusable for navigation positioning purposes. In 1968 to 1969, research institutes of the Ministry of Defense, Academy of Sciences, and Soviet Navy joined together to establish a single solution for air, land, sea, and space forces. This resulted in a 1970 requirements document that established the requirements for such a space-based system. After further basic research, in 1976 a decree was issued by the Soviet Union establishing the Global Navigation Satellite System (GLONASS).

The GLONASS – System Description

The nominal baseline constellation of GLONASS comprises 24 GLONASS-M satellites that are uniformly deployed in three roughly circular orbital planes at an inclination of 64.8° to the equator. The altitude of the orbit is 19,100 km. The orbit period of each satellite is 11 h, 15 min, 45 s. The orbital planes are separated by 120° right ascension of the ascending node. Eight satellites are equally spaced in each plane with a 45° argument of latitude. Moreover, the orbital planes have an argument of latitude displacement of 15° relative to each other. This constellation configuration provides for continuous, global coverage of the Earth's surface and near-Earth space and for minimizing the effect of disturbances on deformation of the orbital constellation.

Current and Future Satellite Generations

Development of GLONASS began in the Soviet Union in 1976. Beginning on 12 October 1982, numerous rocket launches added satellites to the system until the constellation was completed in 1995. However, due to the time that it took to orbit all the satellites of the system, the earlier satellites became nonfunctional and the system quickly became incomplete.

In the 2000s the restoration of the system was made a top government priority and funding was substantially increased. This resulted in GLONASS becoming the most expensive program of the Russian Federal Space Agency (ROSCOSMOS). By 2010, GLONASS had achieved 100% coverage of Russia's territory and in October 2011, the full orbital constellation of 24 satellites was restored, enabling full global coverage. Information on the Official GLONASS website[4] indicates that as of May 2012, the Russian Federation has 31 GLONASS satellites in orbit, with 24 operating to provide global coverage, four spares, two under maintenance, and one undergoing flight tests. The complete GLONASS constellation needs 24 functional and 2–3 reserve satellites to operate with global coverage.

The main contractor of the GLONASS program is Joint Stock Company Reshetnev Information Satellite Systems (formerly called NPO-PM). The company, located in Zheleznogorsk, is the designer of all GLONASS satellites, in

cooperation with the Institute for Space Device Engineering and the Russian Institute of Radio Navigation and Time. Serial production of the satellites is accomplished by the company PC Polyot in Omsk.

Over the three decades of development, the satellite designs have gone through numerous improvements, and can be divided into three generations: the original GLONASS (since 1982), GLONASS-M (since 2003), and GLONASS-K (since 2011).

GLONASS First Generation Satellites

The true first generation of GLONASS (also called Uragan) satellites were all three-axis stabilized vehicles, generally weighing 1,250 kg and were equipped with a modest propulsion system to permit relocation within the constellation. Over time they were upgraded to Block IIa, IIb, and IIv spacecraft, with each block containing evolutionary improvements.

Six Block IIa satellites were launched in 1985–1986 with improved time and frequency standards over the prototypes, and increased frequency stability. These spacecraft demonstrated a 16-month average operational lifetime. Block IIb spacecraft, with a 2-year design lifetimes, appeared in 1987, of which a total of 12 were launched, but half were lost in launch vehicle accidents. The six spacecraft that made it to orbit worked well, operating for an average of nearly 22 months.

Block IIv was the most prolific of the first generation. Used exclusively from 1988 to 2000, and continued to be included in launches through 2005, a total of 25 satellites were launched. The design life was 3 years; however, several spacecraft exceeded this.

Block II satellites were typically launched three at a time from the Baikonur Cosmodrome using Proton-K Blok-DM-2 or Proton-K Briz-M boosters. The only exception was when, on two launches, an Etalon geodetic reflector satellite was substituted for a GLONASS satellite.

GLONASS Second Generation Satellites

The second generation of satellites, known as Glonass-M, was developed beginning in 1990 and first launched in 2003. These satellites possess a substantially increased lifetime of 7 years and weigh slightly more at 1,480 kg. They are approximately 2.4 m in diameter and 3.7 m high, with a solar array span of 7.2 m for an electrical power generation capability of 1,600 W at launch. The aft payload structure houses 12 primary antennas for L-band transmissions. Laser corner-cube reflectors are also carried on board to aid in precise orbit determination and geodetic research. Onboard cesium clocks provide the local clock source. Figure 25.4 below shows a GLONASS-M satellite.

A total of 14 second generation satellites were launched through the end of 2007. As with the previous generation, the second generation spacecraft were launched in triplets using Proton-K Blok-DM-2 or Proton-K Briz-M boosters.



Fig. 25.4 A GLONASS-M satellite (Courtesy of Roscosmos and Information Satellite Systems Reshetnev Company)

GLONASS Third Generation Satellites

GLONASS-K is a substantial improvement over the previous generation. It is the first unpressurised GLONASS satellite with a much reduced mass (750 kg versus 1,480 kg of GLONASS-M). It has an operational lifetime of 10 years, compared to the 7-year lifetime of the second generation GLONASS-M. It will transmit more navigation signals to improve the system's accuracy, including new CDMA signals in the L3 and L5 bands which will use modulation similar to modernized GPS, Galileo, and Compass/Beidou satellites. The new satellite's advanced equipment – made solely from Russian components – will allow the doubling of GLONASS' accuracy. As with the previous satellites, these are three-axis stabilized, nadir pointing satellites with dual solar arrays for power sources. The first GLONASS-K satellite was successfully launched on 26 February 2011. Figure 25.5 shows a GLONASS-K satellite.

Due to their weight reduction, GLONASS-K spacecraft can be launched in pairs from the Plesetsk Cosmodrome launch site using the substantially lower cost Soyuz-21b boosters or in six-at-once from the Baikonur Cosmodrome using Proton-K Briz-M launch vehicles.

The present constellation officially has two reserve satellites, GLONASS 714 and 726 but neither of these satellites will ever be brought back to active service.

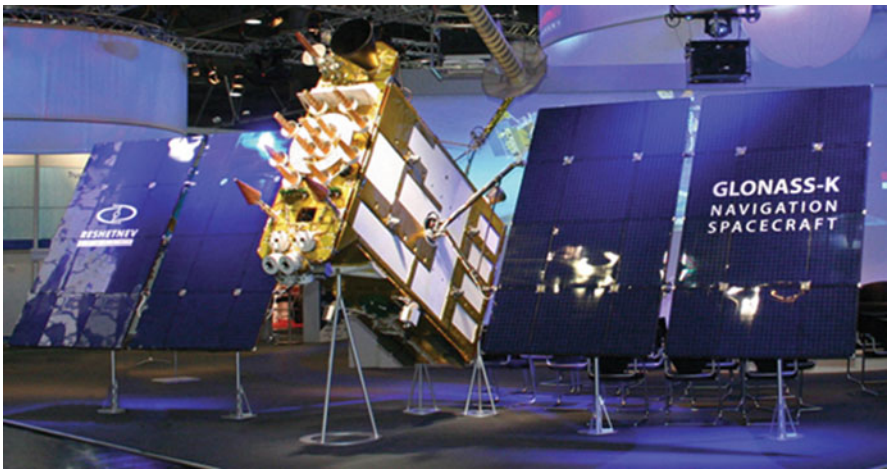


Fig. 25.5 A GLONASS-K satellite (Courtesy of Roscosmos and Information Satellite Systems Reshetnev Company)

714 was a flight-test satellite, apparently, and 726 had a failure of its navigation payload. Rather than being possible replacement satellites, these vehicles are being used to train the ground team to operate spare satellites in a full or nearly full constellation. GLONASS 727, in orbital slot 3, which was taken out of service on 8 September 2010 has also had a failure of its navigation payload and will not be returning to service. It is anticipated that 11 more GLONASS-M satellites will be launched by the end of 2012.[5]

Current and Planned GLONASS Signals

Each GLONASS satellite transmits two types of navigation signals in two sub-bands of L-band: a standard accuracy signal and a high accuracy signal. L1 carrier frequencies are in the band 1,598.06–1,604.40 MHz, and L2 carrier frequencies are in the band 1,242.94–1,248.63 MHz with right-hand circular polarization. GLONASS uses the frequency division multiple access (FDMA) in both L1 and L2 sub-bands.

ROSCOSMOS has announced that there will be two versions of the new GLONASS-K satellites: GLONASS-K1 and GLONASS-K2. As indicated previously, the first GLONASS-K1 satellite was launched in February of 2011. Only one more GLONASS-K1 satellite will be built and launched after that. The K1 satellites will test an open service code division multiple access (CDMA) signal on the GLONASS L3 frequency in the 1205 MHz band (R.B. Langley, GLONASS Update Delves into Constellation Details, http://www.gpsworld.com/gnss-system/glonass/news/glonass-update-delves-constellation-details-10499).

With a completely new design, launching of GLONASS-K2 satellites will start in 2013. GLONASS-K2 satellites will have a 10-year design life. Besides

transmitting CDMA signals on L3, GLONASS-K2 will also transmit CDMA signals on L1 and L2. The GLONASS-K satellites will transmit the legacy FDMA satellites in addition to the CDMA signals.

A modernized GLONASS-K satellite, GLONASS-KM, tentatively to be launched after 2015, is now under study. In addition to transmitting legacy FDMA signals on L1 and L2 and CDMA signals on L1, L2, and L3, CDMA signals may also be transmitted on the GPS L5 frequency at 1176.45 MHz. This transmission coding will facilitate interoperability with the other GNSS. Also being studied is an alternative to the present three-plane, equally spaced satellite constellation. A different constellation design would be possible using CDMA signals. Such a move would require that the legacy FDMA signals be switched off. However, any such move would require at least 10-years' advance notice (R.B. Langley, GLONASS Update Delves into Constellation Details, http://www.gpsworld.com/gnss-system/glonass/news/glonass-update-delves-constellation-details-10499). In 2020, Russia plans to have 30 satellites in orbit, including six in reserve. To support the orbital grouping, Russia plans to launch 22 new-generation GLONASS-KM satellites in the period 2012–2020 to replace the outdated ones.

The design characteristics of the various generations of GLONASS satellites are shown in Fig. 25.6 and the modernization of the GLONASS signals are shown in Fig. 25.7 below.

In Figs. 25.6 and 25.7, the types of signals transmitted are identified with respect to access techniques as: OF = open-access FDMA, SF = special (military) FDMA, OC = open-access CDMA, OCM = open-access CDMA modernized.

Performance Standards Versus Actual Performance

The document that defines requirements related to the interface between the space segment and the navigation user segment is the Interface Control Document (version 5.1, 2008). An update to this document is being prepared and should be available in the 2012–2013 time frame. At present, the main performance characteristics for GLONASS civil service are defined by the GLONASS Standard Positioning Service Performance Requirements.[6] According to this document, for "GLONASS-M" satellite constellation:

- The signal-in-space user range error value over any 24-h interval for all healthy satellites should be less than or equal to 6.2 m, with a 0.95 probability when using open service signals containing ephemeris and clock data transmitted by the operational constellation.
- The position dilution of precision availability (the percentage of time over any 24-h interval that position dilution of precision availability is less than or equal to 6 for the constellation of operational satellites) should be equal to or better than 98% for the full 24-satellite constellation.
- The corresponding real-time and absolute mode positioning accuracy in the state reference frame using signal-in-space only (neglecting user clock bias and errors due to propagation environment and receiver) and assuming that

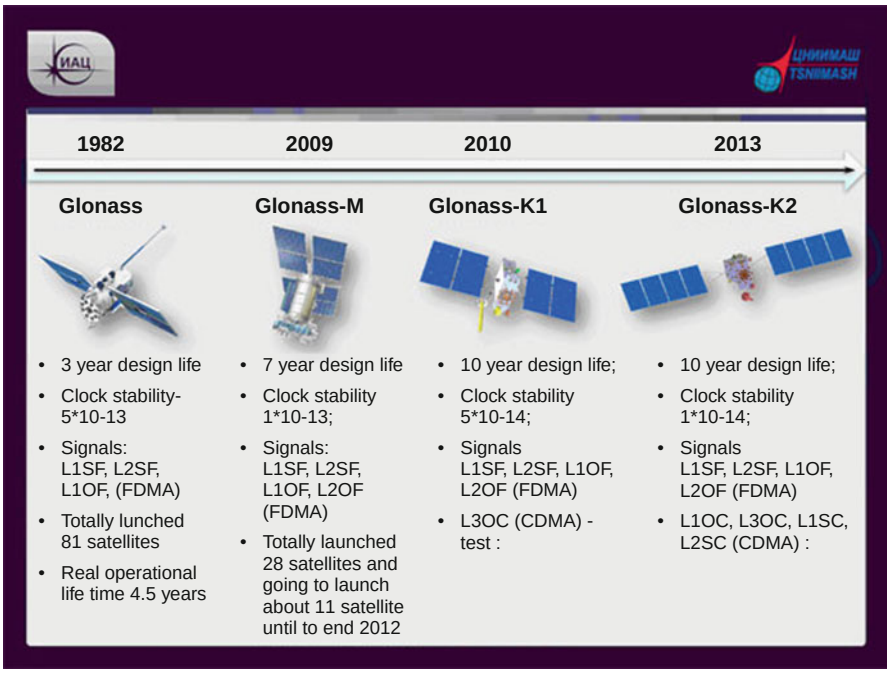


Fig. 25.6 Design characteristics of the GLONASS generations of satellites (Courtesy ROSCOSMOS)

position dilution of precision availability is equal to 2, should be 12.4 m over any 24-h interval for any point within the service volume with 0.95 probability.

Additional information on GLONASS performance monitoring is contained in the website of the Information and Analysis Center[7] (IAC) of ROSCOSMOS.

Services Provided and Provision Policies

GLONASS is a part of the critical state space-based position, navigation, and timing (PNT) infrastructure providing for national security and economic development. Thus, creating, developing, and sustaining the PNT infrastructure is a State responsibility.

GLONASS satellites broadcast two types of navigation signals in L1 and L2 frequency bands: the standard positioning signal and the high accuracy positioning signal. The Russian Federation maintains a policy of no direct user fees for the standard positioning GLONASS signal. The high accuracy positioning signal is modulated by a special code and is reserved for military and selected users and applications. The operator also provides open, free access to GLONASS

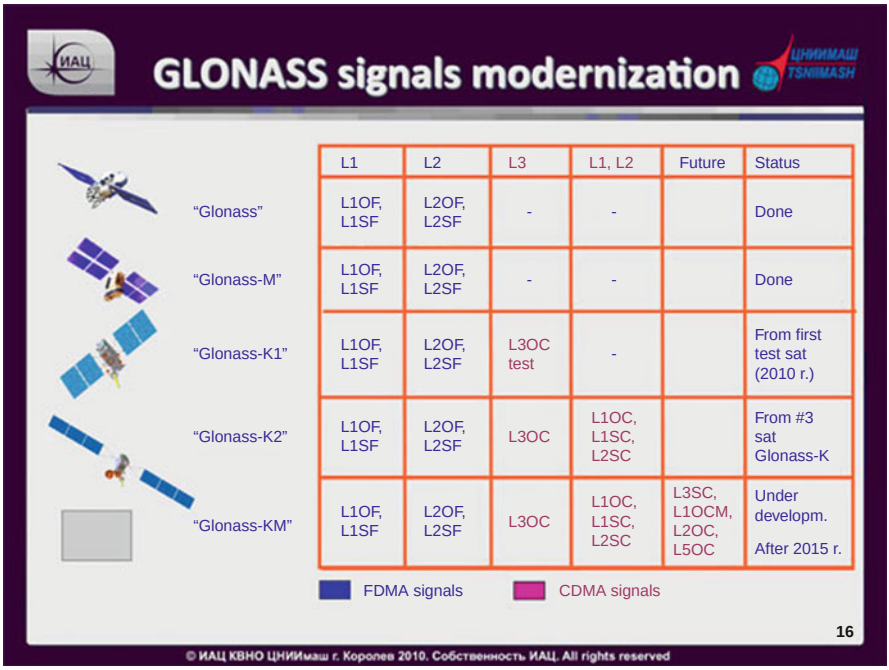


Fig. 25.7 Modernization of the GLONASS (Courtesy ROSCOSMOS)

information necessary to develop and build user equipment. The Russian Federation participates in international cooperation with the providers of other GNSS to ensure compatibility and pursue interoperability.

System for Differential Correction and Monitoring

On 11 December 2011, Roscosmos launched the Luch-5A geostationary relay satellite. Roscosmos reported that the satellite's antennas and solar panels deployed successfully and that the satellite would be positioned in a geostationary orbit at 16° west longitude.

Luch 5A carries a transponder to enable the System for Differential Correction and Monitoring (SDCM). SDCM is a satellite-based augmentation system (SBAS) for GLONASS and GPS, compatible with the US Wide-Area Augmentation System and other SBASs. Luch-5A is the first in a series of new data relay satellites designed to rebuild the Luch Multifunctional Space Relay System that was used to transmit to Earth live TV images, communications, and telemetry from Mir, the Soviet/Russian space station; the Russian segment of the International Space Station; and other low-Earth-orbiting (LEO) spacecraft.

The SBAS transponder will transmit correction and integrity data for GLONASS and GPS on the GPS L1 frequency with a C/A pseudorandom noise code to be assigned

by the GPS Directorate. The data will be provided by the SDCM, which uses a ground network of monitoring stations on Russian territory as well as some overseas stations.

As the SDCM primary service area is Russian territory, the main lobe of the SBAS antenna beam will be directed to the north with an angle of 7° relative to the equator. The transmitted power will be 60 W and will give a signal power level at the Earth's surface roughly equal to that of GLONASS and GPS signals, about –158 dBW.

The current international SBAS data format has a limited capability for broadcasting corrections for both GLONASS and GPS satellites combined. There is space for only 51 satellites, insufficient for the current number of satellites on orbit. Studies are being carried out in an attempt to resolve this problem. One option is to use a dynamic satellite mask, where an SDCM satellite would only broadcast corrections and integrity data for those GLONASS and GPS satellites in view of users in the territory of the Russian Federation.

Luch-5A is the first of three MSRS/SDCM satellites. Luch-5B will be launched in 2012 into a slot at 95° East longitude and Luch-4, in 2014, into a slot at 167° East longitude.

International Cooperation to Ensure Compatibility and Pursue Interoperability

The Russian Federation recognizes that spectral separation of authorized service signals and other systems' signals is not, in practice, always feasible and that such overlap exists now and might continue to do so in the future and believes that providers of the GNSS should try to resolve those issues through consultations and negotiations. In that spirit, the Russian Federation has coordinated with the United States, on GLONASS-GPS compatibility and with the European Union/European Space Agency (ESA) on GLONASS-Galileo compatibility and interoperability and participates in the ICG and its Providers Forum.

The Compass/Beidou Global Navigation Satellite System of China

The BeiDou Navigation System or BeiDou (Compass) Navigation Satellite System is a project by the People's Republic of China to develop an independent satellite navigation system. The name may refer to either one or both generations of the Chinese satellite navigation system.

Space Segment System Description

The first BeiDou system, officially called BeiDou Satellite Navigation Experimental System, and known as BeiDou-1, consists of three satellites and has limited coverage and applications. This system was a demonstration Phase 1 and has been

offering navigation services mainly for customers in China and from neighboring regions since 2000.

Phase II is BeiDou Navigation Satellite (regional) System which is aimed at providing service for areas in China and its surrounding areas from 2012 but which in fact can provide services to many areas around the world. Phase III will see the BeiDou Navigation Satellite System established completely and will provide a fully global service by 2020.

The second generation of the system, known as Compass or BeiDou-2, approved by the Government in 2004, will be a global satellite navigation system consisting of 35 satellites, and is still under construction. The Compass or BeiDou-2 Navigation Satellite System will consists of five geostationary satellites and 30 non-geostationary satellites. BeiDou-2 became operational with coverage of China in December 2011 with 10 satellites in use. BeiDou-2 will offer services to users in Asia-Pacific region from 2012 and to the global community by 2020.

On 2 November 2006, China announced that from 2008 BeiDou would offer an open service with an accuracy of 10 m, timing of 0.2 ns, and a speed of 0.2 m/s (Marks 2006).

On 25 February 2012, China successfully launched a BeiDou-2 satellite from the Xichang Satellite Launch Center in the southwestern Sichuan province. The satellite was carried by a Long March-3C carrier rocket into a geosynchronous orbit and became the fifth geostationary and eleventh overall spacecraft in the current Beidou-2 constellation. The geostationary (GEO) satellites are located at 58.75° E, 80° E, 110.5° E, 140° E, and 160° E in orbits inclined at 55°.

The system is initially capable of providing high accuracy continuous, real-time passive 3-D geospatial positioning and speed measurement services for users in China and its neighboring regions, covering an area of about 120° longitude in the Northern Hemisphere.

On 30 April 2012, a Chinese Long March 3B rocket successfully lifted off to carry the Compass-M3 and Compass-M4 satellites that became China's 12th and 13th global navigation satellites in orbit. As a result, China is well on its way to assembling its own GNSS. The non-geostationary satellites are being placed in three 55° inclined orbits at an altitude of 21,500 km, in Medium Earth orbit (MEO), and in inclined geosynchronous orbit.

Compass Second Generation Satellites

The Compass-M satellites were developed from the DFH-3B satellite platform and are deployed in near circular orbits with a 55° inclination, at altitudes ranging between 21,500 and 24,100 km. The first Compass-M satellite was launched in April 2007. China has developed two models for Compass-M satellites. The two satellites that have been launched into MEO are based on the DFH-3 bus and are equipped with an apogee propulsion system for final orbit insertion. The second model is not equipped with an apogee propulsion system and its platform is completely different from the DFH-3B bus. Still



Fig. 25.8 A Compass-M series satellite (Courtesy NASA Spaceflight.com)

under development, the latter model will be flown in the final construction phase of the BeiDou-2 (Compass) constellation. Figure 25.8 shows a satellite of the Compass-M series.

Current and Planned Beidou-2 Signals

The frequency bands of the Compass/BeiDou Navigation Satellite System, as indicated in 2010 by the China National Administration of GNSS and Applications, included:
B1: 1559.052–1591.788 MHz
B2: 1166.22–1217.37 MHz
B3: 1250.618–1286.423 MHz

The frequencies bands for Compass, identified as B1, B2, and B3 above, are in fact allocated in four bands: E1, E2, E5B, and E6 and overlap with Galileo signals (see figure below). The fact of overlapping could be convenient from the point of view of the receiver design, but on the other hand raises the issues of intersystem interference, especially within E1 and E2 bands, which are allocated for Galileo's publicly regulated service. Compass proposes to use frequencies planned for Galileo's Public Regulated Service (PRS) – and for the GPS military code – meaning that in an emergency, Europe could not jam the Chinese signal without also jamming its own encrypted, security-related signals as well. The same problem holds for the US military.[8]

However, under the International Telecommunication Union (ITU) policies, the first nation to start broadcasting in a specific frequency will have priority to that frequency, and any subsequent users will be required to obtain permission prior to using that frequency, and otherwise ensure that their broadcasts do not interfere with the original nation's broadcasts. It now appears that Chinese Compass

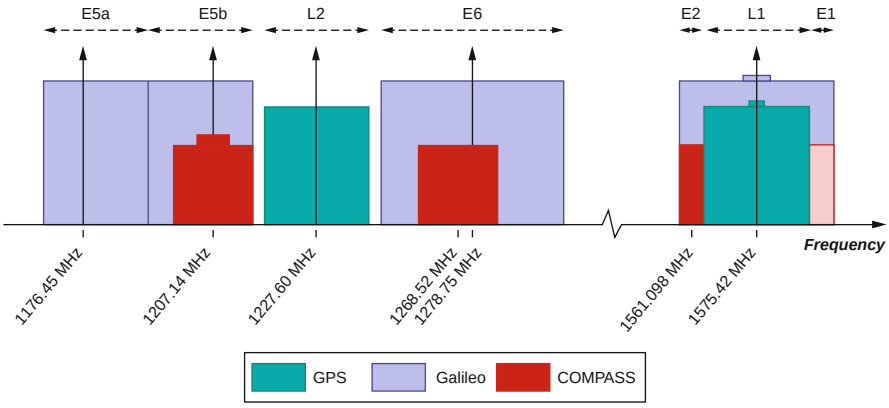


Fig. 25.9 Proposed frequency transmission bands for Compass, Galileo, and GPS (Credit: Wikipedia)

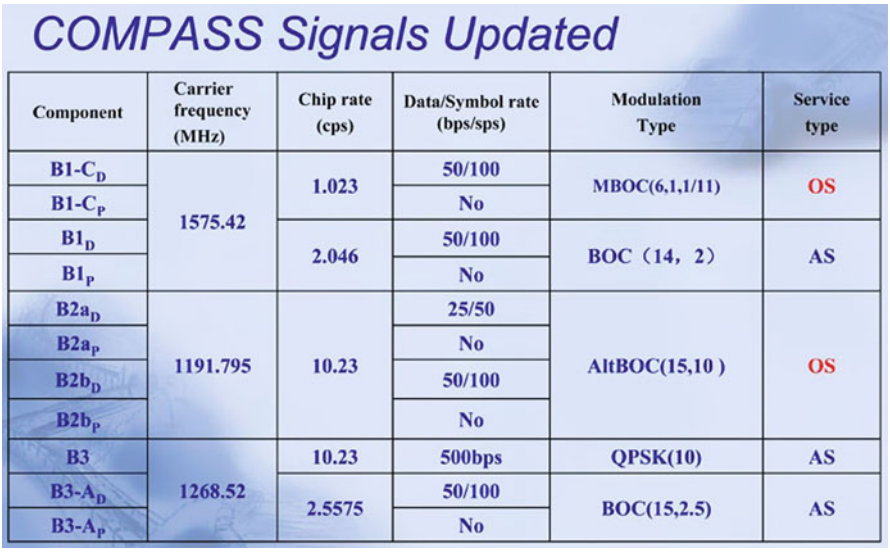
satellites will start transmitting in the E1, E2, E5B, and E6 bands before Europe's Galileo satellites and thus have primary rights to these frequency ranges (Levin 2009). This is an issue that is being discussed bilaterally and in multilateral fora such as the ITU and the ICG (Fig. 25.9).

In July 2009, China presented one possible solution to the impasse on frequency overlay between that country's Compass system and the European Galileo program. At a meeting of Working Group A, on issues related to Compatibility and Interoperability, of the International Committee on GNSS held in Vienna, China showed plans to move the Compass signal modulation to binary offset carrier (BOC), with an alt BOC (15,10) open service (OS) signal in the aeronautical radio navigation band at E5b (centered at 1191.8 MHz) (see Fig. 25.10).[9]

Performance Standards Versus Actual Performance

The performance standards indicated for Compass/Beidou-2 are: coverage area, global positioning accuracy 10 m (95%), velocity accuracy 0.2 m/s, and timing accuracy 20 ns. With the 13 GNSS satellites that are in place in the constellation, Compass can already offer limited global coverage and has been offering the open service signal to users in China and neighboring areas.

Compass features a multiplexed binary offset carrier (MBOC) waveform designed to improve mobile reception in cities and other challenging environments and to increase interoperability with GPS, Galileo, and other satellite navigation systems, such as Japan's Quasi-Zenith Satellite System (QZSS). The technical characteristics of the signals for users and manufacturers of receiver equipment are available at the BeiDou website in the "BeiDou Navigation Satellite System Signal-In-Space Interface Control Document (Test Version)."[10]



COMPASS Signals Updated

Component	Carrier frequency (MHz)	Chip rate (cps)	Data/Symbol rate (bps/sps)	Modulation Type	Service type
B1-C_D	1575.42	1.023	50/100	MBOC(6,1,1/11)	OS
B1-C_P			No		
$B1_D$		2.046	50/100	BOC（14，2）	AS
$B1_P$			No		
$B2a_D$	1191.795	10.23	25/50	AltBOC(15,10)	OS
$B2a_P$			No		
$B2b_D$			50/100		
$B2b_P$			No		
B3	1268.52	10.23	500bps	QPSK(10)	AS
B3-A_D		2.5575	50/100	BOC(15,2.5)	AS
B3-A_P			No		

Fig. 25.10 Technical information on the updated Compass signals (Courtesy: China National Administration of GNSS and Applications)

Services Provided and Provision Policies

The Compass/BeiDou Navigation Satellite System can provide two types of service at the global level: open service and authorized service. Through its open service, it provides cost free positioning, velocity, and timing services. Through its authorized service, it provides higher accuracy positioning, velocity, and timing services, as well as system integrity information, for authorized users. The Compass/BeiDou Navigation Satellite System can provide two kinds of authorized services, including a wide-area differential service (with a positioning accuracy of 1 m) and a short-message communication service in China and nearby areas.

Perspective on Compatibility and Interoperability

The Compass/BeiDou Navigation Satellite System will achieve frequency compatibility with other satellite navigation systems under the ITU framework through bilateral coordination and taking advantage of multilateral discussions such as those carried out in Working Group A of the International Committee on GNSS. Presently, the COMPASS/BeiDou Navigation Satellite System has held coordination meetings with the operators of the GPS, Galileo, GLONASS, and QZSS systems.

The Compass/BeiDou Navigation Satellite System will achieve interoperability with other satellite navigation systems by coordinating through bilateral or

multilateral platforms, including within the framework of the International Committee on GNSS (ICG). On the bilateral level, the System has held coordination meetings with the operators of the GPS and Galileo systems concerning compatibility and interoperability.

The European Satellite Navigation System (Galileo)

The European global navigation satellite system Galileo is a space-based radio navigation system owned by the European Union (EU) and, once functional, will be operated by the dedicated facilities of the European Space Agency (ESA) and several European national space agencies. Galileo is Europe's GNSS, providing a highly accurate, guaranteed global positioning service under civilian control. Galileo provides enhanced distress localization and call features for the provision of a Search and Rescue service interoperable with the COSPAS-SARSAT system.

As far back as the 1990s, the European Union saw the need for Europe to have its own global satellite navigation system. The conclusion to build one was taken in similar spirit to decisions in the 1970s to embark on other well-known European endeavors, such as the Ariane launcher and the Airbus. The European Commission and European Space Agency joined forces to build Galileo as an independent European system under civilian control.

European independence is the chief reason for taking this major step. However, other subsidiary reasons included that by being interoperable with GPS and GLONASS, Galileo will allow positions to be determined accurately for most places on Earth, even in high-rise cities where buildings obscure signals from satellites low on the horizon. This is because the overall number of satellites available from which to take a position is more than doubled. Another important reason building Galileo was that by placing satellites in orbits at a greater inclination to the equatorial plane than GPS, Galileo will achieve better coverage at the high European latitudes.

Space Segment

The fully deployed Galileo system will consist of 30 satellites (27 operational + 3 spares), positioned in three circular Medium Earth Orbit (MEO) planes at 23,222 km altitude above the Earth, and at an inclination of the orbital planes of 56° with reference to the equatorial plane. The satellites will have an approximate revolution period of 14 h. There will be nine operational satellites per orbital plane, occupying evenly distributed orbital slots. The three additional spare satellites (one per orbital plane) complement the nominal constellation configuration. Once this is achieved, the Galileo navigation signals will provide a good coverage even at latitudes up to 75° north and 75° south.[11] Figure 25.11 shows a schematic of the Galileo constellation.

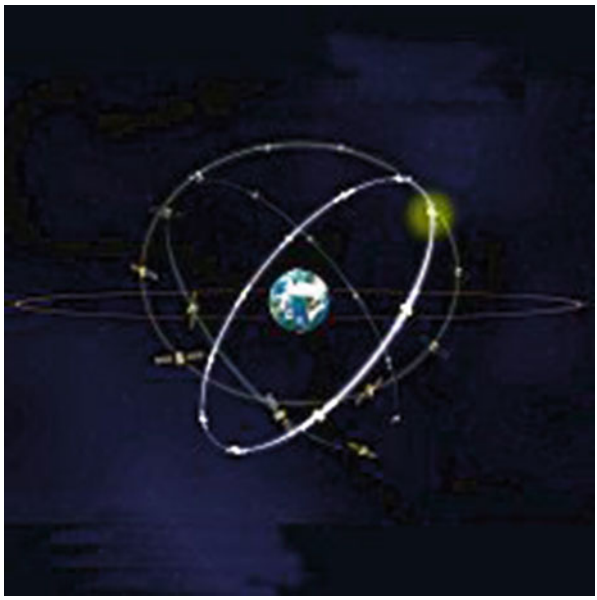


Fig. 25.11 The Galileo satellite constellation showing its high latitude coverage (Courtesy European Space Agency)

By offering dual frequencies as standard, Galileo will deliver real-time positioning accuracy down to the meter range. It will guarantee availability of the service under all but the most extreme circumstances and will inform users within seconds of any satellite failure, making it suitable for safety-critical applications such as guiding cars, running trains, and landing aircraft.

Current and Future Satellite Generations

As was the case with other GNSS, the Galileo satellite constellation will be built through several generations of satellites that will have evolving capabilities and carry out needed functions in a step-by-step manner. The two main phases of the Galileo program are described below.

In-Orbit Validation (IOV) Phase

During this phase, the system is assessed through tests, the operation of two experimental satellites and a reduced constellation of four operational satellites and their ground infrastructure. The two experimental satellites were launched in, respectively, December 2005 and April 2008. Their purpose was to characterize the Medium Earth Orbit (MEO) environment (radiations, magnetic field, etc.) and to test in such environment the performance of critical payload technology (atomic clocks and radiation hardened digital technology). They also provide an early experimental

signal-in-space making it possible to secure the frequency spectrum required for Galileo in accordance with World Radio Conference RNSS allocations.

The first two Galileo satellites were launched into the first orbital plane on 21 October 2011. The intended launch date for the next pair of In-Orbit Validation (IOV) Galileo satellites is scheduled for September 2012. The two IOV space vehicles (SVs) will launch together on a Soyuz rocket from Europe's spaceport in Kourou, French Guiana. During this phase, four operational satellites will complement the two experimental satellites (GIOVE-A and GIOVE-B) already in orbit. In the current design, the satellites each weigh approximately 700 kg and have a design lifetime of 12 years.

As with any satellite, separation marks the start of the critical Launch and Early Orbit Phase (LEOP), when the spacecraft must carry out a series of automated and, later, commanded actions, including deployment of the solar arrays to obtain power, the switching on of the satellite's systems and the setting of initial configurations. This short but crucial LEOP period was overseen by a tightly integrated team of mission operations specialists from ESA's European Space Operations Center (ESOC) and the Center National d'Etudes Spatiales (CNES). The LEOP phase was formally completed on 3 November 2011and control of the satellites was passed to the Galileo Control Center (GCC) at the Forschungszentrum der Bundesrepublik Deutschland für Luft-und Raumfahrt, the German Aerospace Center (DLR) in Oberpfaffenhofen, Germany.

When the navigation payload is switched on, it will mark the start of Galileo's in-orbit test campaign. This rigorous check of the navigation signals is being conducted from ESA's ground station in Redu, Belgium.

The Full Operational Capability (FOC) Phase

This phase consists of the deployment of the remaining ground and space infrastructure, including an intermediate Initial Operational Capability phase with 18 satellites in operation (the four IOV satellites plus 14 others). By 2015, 18 satellites should be in place, followed by the rest by 2020. The full system will consist of 30 satellites, control centers in Europe and a network of sensor stations and uplink stations installed around the globe.

Under an agreement with the European Union, ESA acts as the technical lead for Europe's GNSS program and is responsible for acquiring satellites for the program. In May 2012, ESA announced that the first Full Operational Capability (FOC) Galileo satellite had been assembled at OHB System AG in Bremen, Germany, and was being tested. This is the first of 22 Full Operational Capability (FOC) satellites to be built and is scheduled to be ready for delivery to ESA by the end of 2013.

First launch of Galileo FOC spacecraft from Europe's spaceport in Kourou, French Guiana, is currently expected to take place in the second quarter of 2013: two space vehicles (SVs) on board a Russian Soyuz rocket. The four IOV satellites are in almost all respects the same as the 14 FOC satellites being built by OHB System. Figure 25.12 shows the first operational Galileo satellite, FOC-FM1 being tested by OHB System AG.

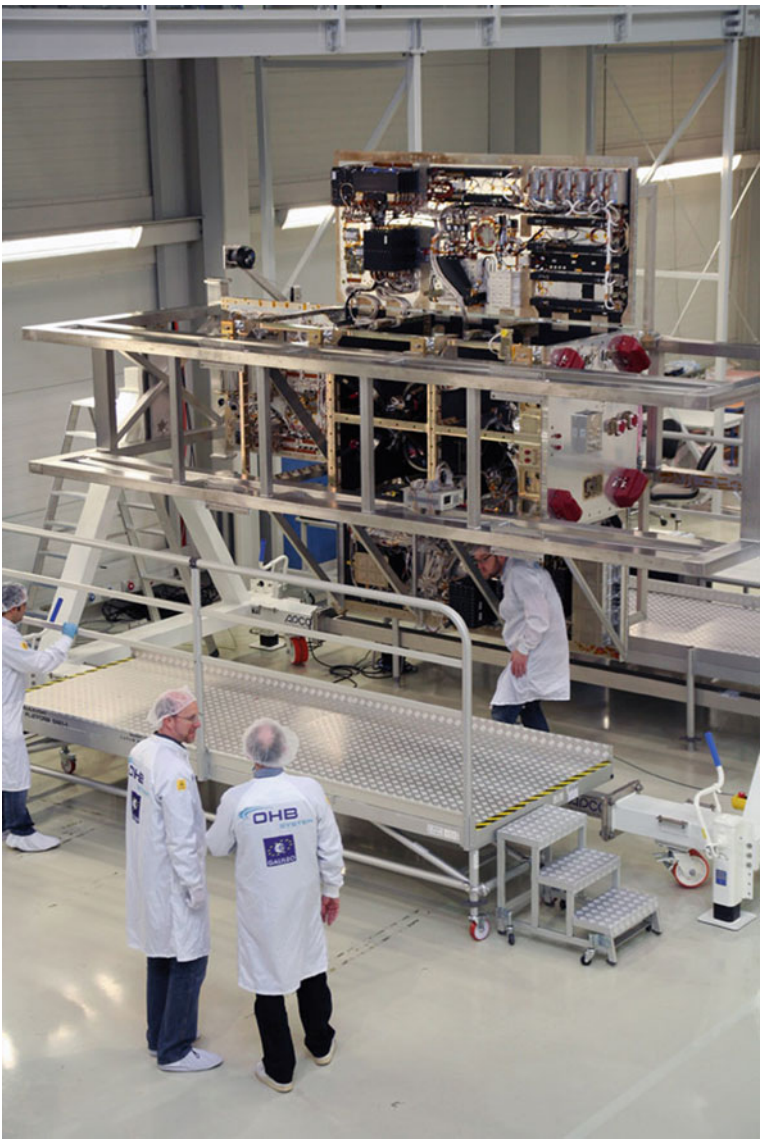


Fig. 25.12 The first operational Galileo satellite, FOC-FM1 being tested by OHB (Courtesy OHM System AG)

As prime contractor for the FOC satellites, OHB is responsible for developing the satellite platform and integrating the satellite with its payload – the part of the satellite that provides Galileo's precision positioning measurements and services to users worldwide. The SV1 payload was developed at Surrey Satellite Technology Ltd in Guildford, United Kingdom. In April 2012, SSTL delivered the first of 14 full

operational capability (FOC) payloads for Galileo to prime contractor OHB System AG and both companies have set up series production facilities to prepare the remaining satellites. In addition to these first 14, the OHB-SSTL consortium won a contract from the European Space Agency (ESA) in February 2012 to build a further eight Galileo FOC satellites.

As payload prime contractor, SSTL is tasked with the development, assembly, integration, and test (AIT) of the navigation payloads at their technical facility in Guildford, Surrey. This first payload has been shipped to prime contractor OHB in Bremen, Germany, for mechanical integration of the payload with the satellite platform and the beginning of the overall space vehicle AIT. SSTL's payload design is based on European-sourced atomic clocks, navigation signal generators, high-power traveling wave tube amplifiers, and antennas.

The definition phase and the development of the In-Orbit Validation phase of the Galileo program are carried out by the European Space Agency and co-funded by ESA and the EU. The FOC phase is funded by the EU and managed by the European Commission. The Commission and ESA have signed a delegation agreement by which ESA acts as design and procurement agent on behalf of the Commission.[12]

Current and Planned Galileo Signals

Galileo will transmit radio navigation signals in four different operating frequency bands: E1 (1,559 ∼ 1,594 MHz), E6 (1,260 ∼ 1,300 MHz), E5a (1,164 ∼ 1,188 MHz), E5b (1,195 ∼ 1,219 MHz). These signals, which are transmitted at specific frequencies, are described below.

Galileo E1

The Galileo E1 band is centered at 1,575.42 MHz. It comprises two signals that can be used alone or in combination with signals in other frequency bands, depending on the performance demanded by the application. The signals are provided for the open service and the public regulated service, both of which include a navigation message. Moreover, an integrity message for the safety-of-life service is included in the open service signal. The E1 carrier is modulated with a CBOC (6,1,1/11) (following the MBOC spectrum) code for the open source and a BOC cos (15,2,5) code for the public regulated service.

Galileo E6

The Galileo E6 signal is transmitted on a center frequency of 1,278.75 MHz and comprises commercial service and public regulated service signals, which are modulated with a binary phase shift keying (BPSK)(5) and BOCcos(10,5) code, respectively. Both signals include a navigation message and encrypted ranging codes.

Galileo E5

The wideband Galileo E5 signal is centered on a frequency of 1,191.795 MHz. This signal is generated with an AltBOC modulation of sideband subcarrier rate of

15.345 MHz. This scheme provides two side lobes. The lower side lobe of E5 is called the Galileo E5a signal, which is centered on a frequency of 1,176.45 MHz and provides a second signal (dual-frequency reception) for the open service and safety-of-life services, both of which include navigation data messages. The upper side lobe of E5 is called the Galileo E5b signal, which is centered on a frequency of 1,207.14 MHz and provides a safety-of-life service, including a navigation message with an integrity information message.

Search and Rescue Signal

The search-and-rescue downlink signal is transmitted by the Galileo satellites in the frequency range of between 1,544 and 1,545 MHz.

Performance Standards Versus Actual Performance

The Galileo open service aims at making positioning, navigation, and timing services widely available, free of charge. The target Galileo open service positioning, navigation, and timing accuracy performances are specified as the 95th percentile of the positioning, navigation, and timing error distribution for different user types and take into account any type of error, including those not under the responsibility of the Galileo system. Hence, the target positioning, navigation, and timing performance specifications are subject to several assumptions on the user terminal and local environment: clear sky visibility, absence of radio frequency interference, reduced multipath environment, mild local ionospheric conditions, absence of scintillations, and fault-free user receiver.

The nominal performance specification for single frequency open service user (E1) and dual-frequency open service user (E1-E5b) are, respectively, 15 m and 4 m in horizontal accuracy and 35 m and 8 m, with a timing accuracy of 30 ns with respect to UTC.

Detailed information on the interface between the Galileo Space Segment and the Galileo User Segment and on performance standards is being added to the Galileo Open Service Signal-In-Space Interface Control Document[13] (OS SIS ICD) as Galileo is being developed.

Services Provided and Provision Policies

The Galileo mission and services have been elaborated during the initial definition phase in consultation with user communities and the European Member States.

The services that are planned to be provided by Galileo are the following:

- *Open Service*: Basic signal provided free of charge.
- *Safety-of-Life Service*: Enhanced signal including an integrity function that will warn the user within a few seconds in case of a malfunction. This service will be offered to the safety-critical transport community, e.g., aviation. It will be certified according to the applicable standards, e.g., those of the International Civil Aviation Organization (ICAO).
- *Commercial Service*: Combination of two encrypted signals for higher data throughput rate and higher accuracy authenticated data.

- *Public Regulated Service*: Two encrypted signals with controlled access for specific users like governmental bodies.
- *Search and Rescue Service*: Galileo will contribute to the international COSPAS-SARSAT cooperative system for humanitarian search and rescue activities. Each satellite will be equipped with a transponder transferring the distress signal from the user to the Rescue Coordination Center and informing him/her that the situation has been detected for distress beacons fitted with Galileo open service receivers.

European Geostationary Navigation Overlay Service (EGNOS)

The European Geostationary Navigation Overlay Service (EGNOS) provides an augmentation signal to the GPS standard positioning service. The EGNOS signal is transmitted on the same signal frequency band and modulation as the GPS L1 (1575.42MHz) C/A civilian signal function. While the GPS consists of positioning and timing signals generated from spacecraft orbiting the Earth, thus providing a global service, EGNOS provides correction and integrity information intended to improve positioning navigation services over Europe.

The EGNOS space segment consists of three navigation transponders on board three geostationary satellites and broadcasting corrections and integrity information for GPS satellites in the L1 frequency band (1575.42 MHz). EGNOS uses the following three geostationary satellites: INMARSAT AOR-E, positioned at 15.5 W; INMARSAT IOR-W (F5), positioned at 25.0 E; and ARTEMIS, positioned at 21.5 E.

The EGNOS ground segment is composed of a network of ranging integrity monitoring stations, four mission control centers, six navigation land Earth stations, and the EGNOS wide-area network, which provides the communication network for all the components of the ground segment. Two additional facilities, the performance assessment and system checkout facility and the application specific qualification facility, are also part of the ground segment to support system operations and service provision.

The main objective of the EGNOS open service is to improve the achievable positioning accuracy by correcting several sources of errors affecting GPS signals. The accuracy achievable with the EGNOS open service is specified as the 95th percentile of the error distribution and is 3 m in horizontal accuracy and 4 in vertical accuracy. However, the typical measured positioning accuracy in the middle of the EGNOS open service area is significantly better than that specification (around 1 m (95%) in vertical accuracy).

Perspective on Compatibility and Interoperability

Galileo coordinates with other space-based positioning, navigation, and timing systems to ensure compatibility. Achieving compatibility is essential when coordinating and it involves both radio frequency compatibility and national security compatibility.

Galileo also coordinates at the bilateral and multilateral levels, including the ICG, to achieve interoperability among the GNSS and believes that in order to achieve interoperability:

- Common center frequency, common modulation, and common maximum power levels, based on the same link budget assumptions, are necessary.
- Highest minimum power level is desirable.
- The availability of information on open signals characteristics (such as a public signal-in-space interface control document) is necessary.
- Geodetic reference frames and system time references steered to international standards are necessary.
- Performance standards and system architecture descriptions must be published.

As indicated previously, the issue of overlapping of the planned frequency bands for transmission of some of its signals with the planned Compass frequency bands is still to be resolved. The agreements reached, or lack thereof, will be notified publicly in due course and reflected in their respective "Signal-In-Space Interface Control Documents."

The Multifunctional Transport Satellite (MTSAT) Satellite-Based Augmentation System (MSAS) and the Quasi-Zenith Satellite System (QZSS) of Japan

Description of MSAS

The Multifunctional Transport Satellite (MTSAT) Satellite-Based Augmentation System (MSAS) provides GPS augmentation information for civil aircraft onboard satellite navigation systems under the Fukuoka Flight Information Region; it is one of the satellite-based augmentation systems that complies with ICAO standards and recommended practices.

MSAS provides navigation services for all aircraft within Japanese airspace via two geostationary satellites: MTSAT-1R, located at 140.0E, and MTSAT-2, located at 145.0E.

The initial satellite in the series, MTSAT-1, was built by Space Systems/Loral (SS/L) as a Multifunctional Transport Satellite. This satellite was intended to be also an advanced geostationary satellite for air traffic control and for weather observation. MTSAT-1 was manufactured under a contract with the Japanese Ministry of Transport Civil Aviation Bureau & Meteorological Agency. This multifunctional satellite was to provide communications and navigational services for aircraft, and provide weather data to users throughout the entire Asia-Pacific region. The integration and ground testing of the satellite was conducted at SS/L's facilities in Palo Alto, California. The spacecraft was delivered in March, 1999, and launched on an H-2S rocket in 2000, but due to a rocket failure never reached orbit.[14]

MTSAT-1R was built by SS/L as a replacement for MTSAT-1 and was successfully launched into geostationary orbit by the launcher H-2A-2022 on 26 February 2005. It was positioned at an orbital slot of approximately 140° East and provides



Fig. 25.13 An artist representation of an MTSAT satellite (Courtesy: JAXA)

high quality digital voice and data communications in the L, Ku, and Ka bands. After launch, MTSAT-1R was renamed Himawari-6.

The MTSAT-2 is a multifunctional satellite with a dual purpose. It is an integral part of a next-generation global-scale air traffic safety system comprised of communications, navigation, tracking, and air traffic control for improving traffic congestion conditions and safety in the Asia-Pacific region. In addition, the MTSAT-2 is designed to obtain, collect, and deliver meteorological images and/or data.[15] Figure 25.13 shows a satellite of the MTSAT series.

MTSAT-2 was built by Mitsubishi Electric as prime contractor and by the Boeing Company as a subcontractor. MTSAT-2 was launched on 26 February 2006 by the Japanese H-2A-2024 launch vehicle and placed in geostationary orbit at 145.0 E. MTSAT-2 was renamed Himawari 7 after launch.

The master control stations generate augmentation information based on the GPS and MTSAT signals received at the ground-monitoring stations and at the monitoring and ranging stations. The ground-monitoring stations monitor GPS satellite signals and transfer the information to the monitoring and ranging stations which monitor the MTSAT orbits.

Current and Planned Signals

MSAS navigation signals transmit from the L1 C/A GPS satellites at a center frequency of 1,575.42 MHz. The signal is modulated by a BPSK (Bi-Phase Shift Keying) technique with pseudorandom noise (PRN) spreading codes having a clock rate of 1.023 MHz, which is contained in the 250 bps/500 sps binary navigation data stream. MSAS is planning to expand bandwidths for L1 and L5. This implementation is under study, in accordance with the improvement schedule for the Wide-Area Augmentation System of the United States.

Performance Standards Versus Actual Performance

MSAS provides horizontal guidance for navigation, which is used in non-precision approaches. According to ICAO standards and recommended practices, in order to

satisfy these requirements: horizontal accuracy is less than 20 m, the observed value is less than 2.2 m (95%), integrity (probability of hazardous, misleading information) is less than 1×10^{-7}/h, fault tree analysis leads to 0.903×10^{-7}/h, and availability is more than 99.9%.

Services Provided and Provision Policies

MSAS is used for aircraft navigation and for that purpose offers three advanced functions. In the event of a GPS failure, the health status of GPS is transmitted via the integrity function of MSAS, while the differential correction function provides ranging error data. MSAS also employs a ranging function to generate GPS-like signals and enable aircraft to use MTSAT as an additional GPS satellite. In order to ensure the reliability of this function, MSAS monitors MTSAT/GPS signals, carries out ranging for the MTSAT satellite orbit and estimates of ionospheric delay on a 24-hours-a-day, 7-days-a-week basis.

Perspective on Compatibility and Interoperability

MSAS is compatible and can interoperate with other satellite-based augmentation systems.

Description of QZSS

QZSS is a regional space-based, all-weather, continuous positioning, navigation, and timing system that provides interoperable signals for GPS (L1, L2, and L5), a wide-area differential GPS augmentation signal called "L1-SAIF" and an experimental signal, "LEX," having a message that contains more data, at a shorter time of transmission. QZSS provides navigation services for East Asia, including Japan, and Oceania.

Space Segment

The space segment comprises the QZSS satellites, which function as celestial reference points, emitting precisely time-encoded navigation signals from space. The operational constellation of three satellites operates in 24-h orbits with an altitude of apogee of less than 39,581 km and a perigee of more than 31,911 km. Each of the three satellites is placed in its own separate orbital plane inclined 39–47° relative to the equator. The orbital planes are equally separated (i.e., phased 120° apart) and the satellites are phased so that there is always one satellite visible from Japan at a high elevation angle.

The satellite is a three-axis stabilized vehicle whose mass, without propellant, is approximately 1,800 kg, including a 320 kg-navigation payload. The major elements of its principal navigation payload are the atomic frequency standard for accurate timing, the onboard navigation computer to store navigation data, generate the ranging code and stream navigation messages, and the 1.2/1.6 GHz band transmitting antenna whose shaped-beam gain pattern radiates near-uniform power of signals at the four 1.2/1.6GHz band frequencies to users on or near the surface of the Earth. Figure 25.14 shows a QZSS satellite watching Japan from above.



Fig. 25.14 An artist representation of a QZSS satellite (Courtesy: JAXA)

Current and Planned Signals

The QZSS navigation signals transmitted from the satellites consist of five modulated carriers: two L1 carriers at center frequency 1,575.42 MHz (154f0), L2 at center frequency 1,227.6 MHz (120f0), L5 at center frequency 1,176.45 MHz (115f0), and LEX at center frequency 1,278.75 MHz (125f0) where f0 = 10.23 MHz. It should be noted that f0 is the output of the onboard frequency reference unit to which all signals generated are coherently related.

The L1 signal consists of four BPSK modulation signals. Two of them, the L1 C/A and the L1-SAIF, are modulated with two different pseudorandom noise spreading codes that are modulo-2 add sequences of the outputs of two 10-bit-linear-feedback-shift registers (10-bit-LFSRs) having a clock rate of 1.023 MHz and a period of 1 ms. Each of them is modulo-2 added to a 50 bps/50 sps or 250 bps/500 sps (i.e., Standard Positioning Service) binary navigation data stream prior to BPSK.

The other two signals, L1Cp and L1Cd, are modulated with two different spreading codes having a clock rate of 1.023 MHz and with two same square waves having a clock rate of 0.5115 MHz. Data stream is modulo-2 added to L1Cd. Only the L1-SAIF signal is transmitted through a separate horn antenna using a different L1 carrier wave.

The L2 signal is BPSK with an L2C spreading code. The L2C code has a clock rate of 1.023 MHz with alternating spreading codes having a clock rate of 0.5115 MHz: L2CM with a period of 20 ms and L2CL with a period of 1.5 s. A 25 bps/50 sps data stream is modulo-2 added to the code prior to phase modulation.

The L5 signal consists of two BPSK signals (I and Q) multiplexed in quadrature. The signals in both I and Q channels are modulated with two different L5 spreading codes. Both of the L5 spreading codes have a clock rate of 10.23 MHz and a period of 1 ms. A 50 bps/100 sps binary navigation data stream is

transmitted on the I channel and no data (i.e., a data-less "pilot" signal) on the Q channel.

The LEX signal is also BPSK modulated. A set of small Kasami Code sequences is employed for the spreading code having a clock rate of 5.115 MHz.

Performance Standards Versus Actual Performance

The specification of signal-in-space user range error is less than 1.6 m (95%), including time and coordination offset error to GPS. User positioning accuracy for QZSS is defined as positioning accuracy of the combined GPS L1 C/A and QZSS L1 C/A for a single frequency user and L1-L2 for a dual-frequency user. The figures of specification are 21.9 m (95%) and 7.5 m (95%) respectively. These specifications have already been verified by simulation using actual system design and parameters measured in an engineering model test. Simulation results for signal-in-space have shown user range error of 1.5 m (95%), a positioning accuracy of 7.02 m (95%) for a single frequency user and of 6.11 m (95%) for a dual-frequency user.

The L1-SAIF signal provides wide-area differential GPS correction data and its positioning accuracy is to be estimated 1 m (1 sigma rms) without large multipath error and ionospheric disturbance.

Services Provided and Provision Policies

GPS interoperable signals like L1 C/A, L2C, L5, and L1C are to be provided free of charge to direct users. Regarding GPS performance enhancement signals, such as L1-SAIF and LEX, a charging policy is under examination. In order to support the design of the QZSS receiver by the receiver manufacturers and its application by a positioning, navigation, and timing service provider, interface specifications for QZSS users were released at an early stage of system development. The document describes not only radio frequency properties, message structure, and definition but also system characteristics, service performance properties, and the concept of operation. Both Japanese and English versions of the document can be downloaded from the JAXA website.[16]

The Indian Regional Navigation Satellite System (IRNSS) and the Global Positioning System-Aided GEO-Augmented Navigation (GAGAN) System

The Indian Regional Navigational Satellite System, hereafter referred to as IRNSS, is an autonomous regional satellite navigation system being developed by the Indian Space Research Organization. The IRNSS would be operated by ISRO under control of the Indian government. The requirement of such a navigation system is driven by the fact that access to Global Navigation Satellite Systems such as GPS is not guaranteed in hostile situations. The IRNSS would provide two

services, the Standard Positioning Service (SPS) open for civilian use and the Restricted Service (RS), which will be encrypted, available only for authorized users (military).

System Description

The proposed system would consist of a constellation of seven satellites and a support ground segment. Three of the satellites in the constellation will be placed in geostationary orbit. These Geostationary orbit satellites will be located 34° E, 83° E, and 131.5° E longitude. The GSO satellites will be in orbits with a 24,000 km apogee and 250 km perigee inclined at 29°. Two of the satellites will cross the equator at 55° E and 111.5° E. Such an arrangement would mean all seven satellites would have continuous radio visibility with Indian control stations. The satellite payloads would consist of atomic clocks and electronic equipment to generate the navigation signals.

IRNSS signals will consist of a Special Positioning Service and a Precision Service. Both will be carried on L5 (1,176.45 MHz) and S band (2,492.08 MHz). The SPS signal will be modulated by a 1-MHz BPSK signal. The Precision Service will use Binary Offset Coding, BOC (5,2).

The navigation signals themselves would be transmitted in the S-band frequency (2–4 GHz) and broadcast through a phased array antenna to maintain required coverage and signal strength. The satellites would weigh approximately 1,330 kg and their solar panels generate 1,400 W.

The System is intended to provide an absolute position accuracy of better than 20 m throughout India and within a region extending approximately 2,000 km around it.

Current and Planned Signals

The IRNSS constellation transmits navigation signals in L5 and S bands. Standard Position Services (SPS) and authorized/restricted services (RS) that use encryption technologies are the basic services offered by IRNSS. The IRNSS standard position and restricted services are transmitted on L5 (1,164–1,215 MHz) and S (2,483.5–2,500 MHz) bands. The carrier frequencies and bandwidths of these signals are the following:

- SPS – L5 has a carrier frequency of 1,176.45 MHz and a bandwidth of 24 MHz.
- RS – L5 has a carrier frequency of 1,176.45 MHz and a bandwidth of 24 MHz.
- SPS – S has a carrier frequency of S 2,492.028 MHz and a bandwidth of 16.5 MHz.
- RS – S has a carrier frequency of 2,492.028 MHz and a bandwidth of 16.5 MHz.

The standard position service (SPS) signal is BPSK(1) modulated on both the L5 and S bands. The navigation data is at a data rate of 25 bps and is modulo 2 added to a pseudorandom noise code chipped at 1.023 Mcps identified for the standard position service. The CDMA-modulated code modulates the L5 and S carriers at 1,176.45 MHz and 2,492.028 MHz, respectively.

The restricted service is only available for authorized users. The restricted service signal is transmitted on L5 and S bands using binary offset coding (BOC). It has two channels: a "data" channel and a "pilot," or "data-less," channel. The navigation data at 25 bps is modulo 2 added with designated PRN code chipped at 2.046 Mcps in the data channel. The CDMA bit stream modulates the L5 and S carriers using BOC (5,2). The pilot channel is transmitted using primary and secondary codes without data modulation. The primary codes are chipped at 2.046 Mcps. The pilot carrier is in phase quadrature with the data channel.

GPS-Aided GEO-Augmented Navigation System (GAGAN)

The GPS-aided GEO-Augmented Navigation System (GAGAN) is a satellite-based augmentation system to enhance the use of the GPS signal. As an operational system, it is planned that the space segment will consist of two geostationary satellites, located at 82° E and 55° E, respectively. An additional on-orbit spare (located at 83° E) will also be added.

Services Provided and Provision Policies

Regarding the Indian Regional Navigation Satellite System, standard position services and authorized service/restricted service are the basic services offered by IRNSS. The standard position service is free for all users. The restricted service is encrypted and, as such, available only to authorized users.

GAGAN will provide a safety-of-life service that meets all the requirements of accuracy, integrity, continuity, and availability required by ICAO for the utilization by civil aviation for en route, non-precision, and precision approaches.

Conclusion

This chapter has sought to provide a detailed summary of the various satellite navigation and timing systems that are deployed or are being deployed around the world. Clearly, this a very dynamic period with the Russian GLONASS system being restored to a very high level of performance and new and improved systems being deployed by Chinese, European Japanese, and Indian initiatives in this rapidly evolving field. Although the US Navstar/GPS system along with the Wide-Area Augmentation System (WAAS) is the most widely used system – partly because of the global availability of low cost GPS receivers – the other systems described in this chapter are going to be more and more widely used. The compatibility among the GNSS will be key to the interoperability. Since this is a rapidly changing area with the deployment of many new satellites and system capability it is prudent to go to the various websites cited in this chapter to check on the very latest information that becomes available.

Cross-References

- ▶ Global Navigation Satellite Systems: Orbital Parameters, Time and Space Reference Systems and Signal Structures
- ▶ International Committee on GNSS
- ▶ Introduction to Satellite Navigation Systems

Notes

1. http://www.losangeles.af.mil/shared/media/document/AFD-081021-035.pdf
2. http://www.losangeles.af.mil/shared/media/document/AFD-081021-036.pdf
3. http://www.losangeles.af.mil/shared/media/document/AFD-081021-034.pdf
4. http://www.glonass-ianc.rsa.ru/en/
5. R.B. Langley, GLONASS Update Delves into Constellation Details, http://www.gpsworld.com/gnss-system/glonass/news/glonass-update-delves-constellation-details-10499
6. United Nations Office for Outer Space Affairs, ST/SPACE/50, "Current and Planned Global and Regional Navigation Satellite Systems and Satellite-Based Augmentations Systems", http://unoosa.org/pdf/publications/icg_ebook.pdf
7. http://www.glonass-ianc.rsa.ru/en/GLONASS/
8. C. Alan, "The System: A Healthy Constellation", GPS World, April 2008, http://www.gpsworld.com/gnss-system/compass/the-system-a-healthy-constellation-4226
9. Working Group Workshop on GNSS Interoperability, Vienna, Austria, 30–31 July 2009, http://www.unoosa.org/oosa/en/SAP/gnss/icg/wg/wga02/pres.html
10. http://www.beidou.gov.cn/attach/2011/12/27/201112273f3be6124f7d4c7bac428a36cc1d1363.pdf
11. http://www.esa.int/esaNA/galileo.html
12. http://ec.europa.eu/enterprise/policies/satnav/galileo/satellite-launches/index_en.htm
13. http://ec.europa.eu/enterprise/policies/satnav/galileo/files/galileo-os-sis-icd-issue1-revision1_ ?en.pdf
14. http://space.skyrocket.de/doc_sdat/mtsat-1.htm
15. http://space.skyrocket.de/doc_sdat/mtsat-2.htm
16. http://qzss.jaxa.jp/is-qzss/index_e.html and http://qz-vision.jaxa.jp/USE/is-qzss/index_e.html

References

D. Levin, "Chinese Square Off With Europe in Space", The New York Times, 23 March 2009, http://www.nytimes.com/2009/03/23/technology/23iht-galileo23.html?_r=2%26scp=1%26sq= chinese%20europe%20galileo%26st=cse

P. Marks, " China's Satellite Navigation Plans Threaten Galileo", New Scientist, 08 November 2006, http://www.newscientist.com/article/dn10472-chinas-satellite-navigation-plans-threaten-galileo.html

Section 3

Space Remote Sensing

Introduction and History of Space Remote Sensing

26

Scott Madry

Contents

Introduction 658
The Purpose of Remote Sensing and Geomatic Systems 659
History of Satellite Remote Sensing Services 659
Overview of Satellite Remote Sensing Services 663
Cross-References 665
Notes 665
References 666

Abstract

This chapter introduces the subject of remote sensing both in terms of its technology and its many applications. Remote sensing via satellite has become a key service that is used in many civil applications such as agriculture, forestry, mining (and prospecting for many types of resources), map making, research in geosciences, urban planning, and even land speculation. Perhaps, one of the most vital uses of remote sensing today is related to disaster warning and recovery. The first use of remote sensing was essentially for military purposes and this remains the case today, and, thus, this chapter addresses these applications as well. Remote sensing, Earth observation, related Geographical Information Systems (GIS), plus the interpretation and use of this type of data are today often referred to today as Geomatics.

This section starts with a history of remote sensing and then continues with a discussion of the technology and its applications. In a number of ways meteorological or weather satellites are essentially a specialized form of remote sensing satellites. Thus the history presented here covers the not only what are considered remote sensing satellites but meteorological satellites as well.

S. Madry
International Space University, Chapel Hill, North Carolina, USA
e-mail: Scottmadry@mindspring.com

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_37, © Springer Science+Business Media New York 2013

The meteorological satellites are discussed in much greater detail later in this Handbook.

Keywords

Corona • Cosmos • Earth resources technology satellite (ERTS) • Geographical information systems (GIS) • Geomatics • Geostationary orbiting environmental satellites (GOES) • Hyperspectral sensing • Infrared sensing • Landsat • LIDAR • Multispectral sensing • National ocean and atmospheric administration (NOAA) • Optical sensing • Pixel • Polar orbiting environmental satellites (POES) • Polar sun synchronous orbit • Radar sensing • Remote sensing • Resolution • Sensor • SPOT • Television and infra-red observation satellite (TIROS) • Zenit satellites of the Soviet Union

Introduction

Humans have always sought the high ground looking for food, exploring new lands, and watching for danger. Seeking the heights is an activity as old as the human species itself, as old as our urge to explore and discover. As technology has continually developed, we have found new ways to provide this important capability with greater precision and with greater coverage. As our capabilities have increased, we have also have found many new uses for that view from on high. Remote sensing can be defined as the art and science of acquiring information from sensors or systems that are not in direct contact with the objects being studied. The first practical remote sensing systems used photographic cameras, or simply the viewer's eyes, from balloons, kites, and even small cameras attached to pigeons. The development of aircraft and the rapid technological innovations of the first and second world wars created practical aerial remote sensing capabilities that have helped to map and understand our world.

The dawn of the space age was the next logical extension of this, by now quite important, aerial view of the Earth. Indeed, both the Soviet Union and the United States were quick to see the tremendous potential of space remote sensing for military reconnaissance, weather prediction, environmental analysis, and other applications. The original photographic systems used film and photographic prints to capture data. Over time more and more sophisticated sensors were developed to capture data across an ever broader range of spectra – both above and below visible light and the visual range available to the human eye. Over time an interpreter's skill in analyzing remote sensing data has given way to modern electro-optical systems and digital computers that help us to acquire, store, and analyze data about not only our Earth but our neighboring planets as well.

Over the decades since the first military remote sensing satellite launches in the 1950s, the technology has developed tremendously, and we are now in a very

different situation from the cold war era. Today, many nations are developing, launching, and operating their own remote sensing satellite systems for a wide variety of civil, commercial, and strategic applications.

Satellite remote sensing has changed the way that we think about our planet and has served as important stabilizing influences in the cold war era, and perhaps even more so in the post cold war era. This versatile technology has great technology transfer potential, and has served as the basis for our initial robotic exploration of the moon and other planets as well.

The Purpose of Remote Sensing and Geomatic Systems

This chapter of the handbook will cover what remote sensing is, how it has developed, how it is used, and where the technology and its many applications are going. It is particularly important to note that there is a new term of art called Geomatics. Geomatics refers to all forms of remote sensing that includes the sensing technology, the capturing and display of data on Geographical Information Systems (GIS), and the capturing and interpretation of all forms of remote sensing data.

History of Satellite Remote Sensing Services

The development of space remote sensing is an evolution of mapping and cartography, which has its origins in many cultures around the world. Accurate maps were created and used in ancient Mesopotamia and Egypt over 3,500 years ago. Eratosthenes (276–195 BC) measured the circumference of the Earth and divided the Earth into 60 even grids, and Hipparchos of Nicea (165–127 BC) defined our 360° system that is still in use around the world today. It was in Ptolemy's *Geography* that the world was first defined within an even more precise system of degrees, minutes, and seconds of longitude and latitude. In China, the Chou emperors (1,100 BC) had royal geographers and Phei Hsui (267 AD) had a detailed set of maps created that covered all of the lands within the Chinese domain. The era of European exploration in the sixteenth through eighteenth centuries created a flourishing of cartographic techniques and methods that laid the foundation for accurate mapping and analysis of our world.

Even before the development of aircraft, people were finding ways to look from above, and the first recorded aerial photo was taken of Boston, Massachusetts in 1860 by Samuel Archer King from a balloon. In the American civil war balloons were used for military observation, and Dr. Julius Neubronner patented a miniature pigeon camera system in 1903 that worked quite well. Gaspar Felix Tournachon, known as "Nadar," photographed Paris from his balloon in the 1860s and Alfred Nobel developed a photo-taking solid rocket in 1897. The US Army Signal Corps used a series of large box kites to photograph the damage of the April, 1906 great earthquake and fire in San Francisco (see Fig. 26.1), California, and the pictures



Fig. 26.1 Kite-based photographic image of San Francisco in 1906 (Image courtesy of the US Geological Survey)

were prominently used by newspapers throughout the USA to rally relief support, the first known use of remote sensing for disaster response, which is now a very common application (Madry and Pelton 2010).

The development of the airplane and the military needs of World War one caused a huge improvement in aerial reconnaissance, which was quickly followed by a growing commercial aerial mapping industry. The Second World War produced the next great leap in the development of advanced cameras, films (including the first infrared film to defeat camouflage), image interpretation, and other advanced systems. After the war there were thousands of surplus aircraft and aerial cameras, trained pilots and landing fields throughout the world that spurred the next leap in aerial mapping systems, and the development of the German A4/V2 rockets laid the foundation for space remote sensing systems.

Modern remote sensing systems all derive directly from cold war military needs and capabilities. The Soviet Union and the USA shared an intense desire to acquire reconnaissance capabilities capable of covering large areas deep inside the other's territory. The Soviets authorized their Zenit spy satellites in 1958, only months after the flight of Yuri Gagarin, which used the same Vostock capsule. A total of 10 of the first 20 Cosmos series launches were Zenit systems, carrying four cameras that could acquire 1,500 individual photographs of the Earth per mission. The entire capsule was then deorbited and the cameras refurbished and reused.[1]

At the same time, the USA approved their Corona system in 1958, with the first launch and film recovery in 1960. The Corona was one of the most secret of US military program, and suffered 13 failures in a row before the first useable film was recovered. The exposed film was deorbited in a reentry capsule that was snagged in midair, while the entire satellite was deorbited and burned up upon reentry. The Corona program ran through 1972 and mapped over 5.5 million square miles in over 100 missions, an amazing technological feat at that time.



Fig. 26.2 Image from Corona satellite of the US Pentagon (Image from US Defense Department Archives)

This satellite, in its time, was by far the highest resolution imaging satellite available in Earth orbit. While it was one of the most secret military programs of its day, the Corona images are now available online and are available for use for long-term environmental analysis (see Fig. 26.2). The direct descendants of these early systems are flying today, continuing to push the development of the technology and providing an important stabilizing factor in global geopolitics.[2]

The first explicitly civil satellite remote sensing system was the US TIROS, the Television and Infrared Observation Satellite. While it started as a defense department initiative, it was transferred to NASA in 1959 and launched in April of 1960. It was not in a polar orbit, as is standard now, and had very poor spatial resolution by our standards today, but it quickly demonstrated the tremendous potential that satellite imaging could have for a variety of applications including weather, ice monitoring, ocean studies, and more. Nine more TIROS satellites followed, and the program was later moved to NASA, and its descendants continue as the National Ocean and Atmospheric Administration's (NOAA's) Geostationary Operational Environmental Satellite (GOES) and Polar Operational Environmental Satellite (POES) systems.

The Landsat program was the first systematic moderate resolution civil remote sensing system. It was started in 1970 and the first satellite was launched on July 23, 1972, only 2 years later. The Landsat program was first conceived in 1966 as a direct outgrowth of the successful Mercury, Gemini, and Apollo astronaut photographs, and was originally named ERTS, for Earth Resources Technology Satellite. The Landsat system continues in operation and it has generated years of broad synoptic coverage of our planet and originated hundreds of scientific applications and commercial uses.[3]

France launched their very successful SPOT program in 1986 that became the first system to produce remote sensing data for commercial consumption.[4] Once this process of collecting remote sensing data by satellite for commercial distribution became established the idea caught on quickly. Today, there are over 30 nations around the world operating a wide range of satellite systems that fall into several general categories. The Geostationary weather satellites provide a broad, synoptic coverage of our planet and are primarily used for continental scale weather forecasting and severe weather monitoring. These are operated as governmental organizations as a public good, but many of the other remote sensing satellite systems now offer commercial services to a global customer base although some restrictions often apply.

The optical systems in polar orbit can be grouped into low (1,000–250 m), medium (100–30 m), high (30–5 m), and very high (less than 5 m) spatial resolution. Currently available commercial satellite data of as little as 35–50 cm spatial resolution can be purchased today. This is a truly amazing transition from the first highly classified military satellites to a global commercial market for high-resolution images of the Earth. Active microwave systems using radar are also operating as civil systems, providing a very different and useful view of the Earth. All of these systems share common parameters and provide us with a continuous stream of data about our planet.

The latest evolution is reflected in the deployment of remote sensing satellites with hyperspectral imaging sensors. These satellites instead of sensing a broad range of the spectrum, provides the ability to break down imaging into very narrow bands. The data output from these sensors is deployed in so-called “data cubes” with the “x” and “y” axis being able to depict a spatial area and the “z” axis in the data cube showing the results across narrow spectral bands.

Hyperspectral imaging presents a very difficult issue in terms of processing the data since it requires a broadband downlink to send the data back to Earth and it presents a very difficult issue as to how to analyze and interpret the mountain of data generated by such imaging. The amount of data provided in hyperspectral sensing (since it may be collecting data over 200 different narrow spectral ranges) is so large, one can really only process by computers. Hyper-spectral sensing can produce almost two orders of magnitude greater data than is the case with multi-spectral sensing. This is because data is collected in very narrow spectral slots (i.e. up to 200) rather than across five to ten broad bands.[5]

Overview of Satellite Remote Sensing Services

Remote sensing satellites now provide the world with an amazing variety of information about our spaceship Earth. Two primary orbits are used, a Sun-synchronous polar orbit, often between 500 and 800 km, provides high and moderate resolution coverage of the entire planet on a regular basis. This special orbit allows a satellite to pass over the same location on the Earth periodically at the same time of day and with the same solar illumination. The satellite has to shift its orbit by approximately 1° per day as the Earth orbits around the Sun to achieve this periodicity. These types of satellites provide data for uses as varied as land use planning, forestry, military reconnaissance, and natural resources monitoring. Weather satellites in the geosynchronous arc provide a constant view of a portion of the globe at a lower resolution, but with a constant view. A ring of these, operated by several nations, now provide a constant world-wide watch over our planet for severe weather and broad climate applications.

By themselves, each of these capabilities are useful, but the integration of different types of remote sensing data, along with other types of information for various sources, are driving many new scientific investigations, practical applications, and commercial markets. Geographic Information Systems, or GIS, allows the integration of remote sensing data with population and other demographic data, in situ environmental telemetry, GPS positioning, and other relevant data. GIS has now become such an important way of depicting information in a useful and integrated way that it now represents a larger commercial market than remote sensing.[6] There are important and developing political, legal, social, and dual-use issues with all of these technologies, and these will continue to evolve as newer and more powerful systems and applications become available. The fact that over 300 million people now have access to Google Earth is a powerful demonstration of the impact and utility of geospatial data combined with high-resolution remote sensing imagery.

Remote sensing is a very complex and technologically advanced process, and includes a complex chain of end-to-end data flow and image processing, in order to deliver useful data to the end user. It is quite an achievement that a sensor on a satellite, moving at 8.5 km/s at some 800 km above the Earth, can collect, store, and transmit data to the ground that can be processed into the images we have become familiar with today on Google Earth and other sources.

The fundamental process of extracting useful information from remote sensing data relies upon the fact that objects that are of similar physical makeup can be identified by analyzing the energy that is reflected or emitted from it. This "spectral signature" is unique to each type of matter, and this allows us to discriminate between clean and polluted water, mature or immature crops, healthy or diseased forests, etc.

At the heart of each remote sensing system is a detector, which is the heart of the system. Energy is emitted from the Sun and some small fraction enters the Earth's atmosphere, interacts with the surface, is reflected or emitted from the surface back

through the atmosphere, and falls upon the detector in a satellite after passing through the system's optical telescope. Most passive remote sensing detectors are designed in accordance with the photoelectric effect, and there will be an emission of negatively charged particles (electrons) when a negatively charged plate of an appropriate light-sensitive material is subjected to a beam of photons.

These electrons can then be made to flow from the plate, collected, and measured as an electrical current or signal. After some noise reduction and other preprocessing, the signal is digitized on the satellite and is transmitted to the ground, where an extensive ground processing system further processes the data to produce the data which are then available for use. Image processing and remote sensing specialists further process the data to extract useful information about our world. Most passive remote sensing data are in a raster, or two-dimensional numerical array format, with a single digital value for each area on the ground where the sensor was focused for some milliseconds. This instantaneous field of view (IFOV) is the pixel size of the data, and represents the spatial resolution of the image.

Other important resolution parameters include the temporal resolution (or how often the sensor samples the same location on the ground), the spectral resolution (or which part of the electromagnetic spectrum is measured), and the radiometric resolution (or how finely the signal can be measured). These resolution parameters define the characteristics of the sensor, and are used to determine which data are useful for a given application. Active radar and lidar systems work in a very different manner and are more complex, but, in the end, are processed into similar raster imagery for analysis or integration into a GIS.

There are many exciting new developments in the field of satellite remote sensing. These new developments include: new higher resolution systems operating down to 35 cm per pixel, hyperspectral sensors, high resolution radar systems, new artificially intelligent processing techniques, as well as a range of new applications. The entire field of geomatics is evolving both in terms of its technology, applications, and data display capabilities. The key remaining issue is the extent to which geomatics is to continue to evolve as largely a governmental service offer to citizens as a public good much like meteorological satellite services or whether it will evolve as a commercial service to support an array of newly evolving commercial markets. Clearly some vital remote sensing services such as to support disaster warning and recovery services and those related to coping with climate change will remain as a public service supporting the public good, but others new services may evolve on a strictly commercial basis.

Conclusion

The chapters that follow in this section seek to provide a comprehensive and interdisciplinary overview of satellite remote sensing technology, services, and applications. These chapters also seek to address to some degree the relevant markets, economics, operations, regulation, and future trends. The specifics of spacecraft design and technology for remote sensing and other applications satellites are not addressed in detail here. Many aspects of these technologies have already been addressed in the earlier section on telecommunications

satellites. Thus a consolidated approach related to all applications is taken by combining this information in Sections 6 and 7 of the Handbook.

Section 6 of this book thus addresses the common technical elements found in essentially all applications satellites in terms of power systems, spacecraft bus technology and structural design, and on-board thrusters systems. Specifically, Section 6 addresses solar, battery, and nuclear spacecraft power systems; thermal balancing and heat dissipation systems; orientation, pointing and positioning systems; structural design elements; diagnostic systems; tracking, telemetry and command systems; manufacturing and integration; and quality and reliability testing processes.

Section 7 addresses how the various applications satellites are launched by different rocket systems from various launch sites around the world and the differences between solid, liquid, and hybrid launching systems.

Remote sensing satellites represent the smallest of the various space applications in terms of market size, but the application of this technology is of utmost importance. Today remote sensing satellite are used to undertake disaster recovery, carry out environmental and climate change science, while remote satellite imaging is actively applied to an ever-increasing number of economic fields that include agricultural, forestry, fishing, mining, and industrial applications and services. The commercial market for satellite-based remote sensing may be small in comparison to a field like space telecommunications but the critical importance of the global applications makes this field enormously large.

Cross-References

- ▶ Astronaut Photography: Handheld Camera Imagery from Low Earth Orbit
- ▶ Digital Image Acquisition: Preprocessing and Data Reduction
- ▶ Digital Image Processing: Post-processing and Data Integration
- ▶ Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing
- ▶ Electro-optical and Hyper-spectral Remote Sensing
- ▶ Geographic Information Systems and Geomatics
- ▶ Lidar Remote Sensing
- ▶ Operational Applications of Radar Images
- ▶ Remote Sensing Data Applications

Notes

1. Zenit, *Encyclopedia Astronautica*, http://www.astronautix.com/craft/zenit.htm
2. Corona Program Summary, FAS Space Policy Project, Military Satellite Programs, http://www.fas.org/spp/military/program/imint/corona.htm
3. The Landsat Program, http://landsat.gsfc.nasa.gov/

4. J. Rao, Spot remote Sensing system, http://www.sphttp://www.space.com/6870-spot-satellites.html
5. Hyperspectral imaging, http://en.wikipedia.org/wiki/Hyperspectral_imaging
6. M. Frenkel, National Institute of Standards and Technology, *Global Information Systems in Science* http://pubs.acs.org/doi/abs/10.1021/je800877f

References

S. Madry, J. Pelton, Satellites in the service to humanity in *The Farthest Shore: A 21st Century Guide to Space*, ed. by J.N. Pelton, A. Bukley (Apogee Press, Burlington Ontario Canada, 2010)

Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing

27

M. J. Rycroft

Contents

Introduction 668
The Fundamentals..... 669
Reflection, Refraction, Diffraction, Interference, and Polarization: Important Properties of Electromagnetic Waves 670
The Doppler Effect..... 674
Multiwavelength Studies and Black Body Radiation 674
Another Effect Due to Photons, the Photoelectric Effect 679
References..... 681

Abstract

Here, we consider a topic which is absolutely central to the successful operation of all satellites and spacecraft, namely the basic principles and fundamental concepts of visible light in particular and of electromagnetic radiation in general. Both the wavelike nature of light (the speed of light being 300,000 km/s through free space) and its particle-like nature (as photons) are considered. We introduce its wave properties which explain the phenomena of reflection, refraction, diffraction, interference, polarization, and the Doppler effect. The photon properties explain blackbody radiation, continuous spectra, emission spectra, absorption spectra, and the photoelectric effect. We mention how electromagnetic radiation is used actively for radio communications with Earth-orbiting satellites and passively for remote sensing investigations not only of the atmospheres of the Earth and other planets but also of distant stars and the structure of Universe.

M.J. Rycroft
Cambridge Atmospheric and Space Activities and Research (CAESAR) Consultancy,
35 Millington Road, Cambridge, CB3 9HW, UK
e-mail: michaelrycroft@btinternet.com

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_38, © Springer Science+Business Media New York 2013

Keywords
Electromagnetic radiation • Electromagnetic waves • Frequency • Gamma-rays • GPS • Infrared • Light • Milky Way • Photons • Radar • Radio waves • Spectra • Sun • Ultraviolet • Universe • Velocity of light • Wavelength • X-rays

And God said, "Let there be light," and there was light
Genesis, chapter 1, verse 3, The Holy Bible, New International Version

Introduction

Our human eyes are tuned to receive radiation from the Sun, in the form of light. Light is one type of electromagnetic radiation which, from many different viewpoints, is crucial for space studies and space applications. Thus an understanding of both electromagnetic waves and the electromagnetic spectrum is essential for all readers of this volume. Electromagnetic radiation is used for both active and passive studies carried out from space – in the former, waves are transmitted and received (e.g., in a radar altimeter) and, in the latter, they are only received, for example, from a star or from the Sun, the star of our solar system. Electromagnetic radiation from the Sun, whose intensity peaks in the visible part of the spectrum, is the source of energy for almost all processes occurring on the Earth, and for all life on Earth, for the fauna and flora which abound. Light takes about 8 min to travel the $\sim$150 million kilometers from the Sun to the Earth; thus, when we observe the Sun we see it as it was 8 min ago.

Nowadays, electromagnetic waves give us much information on the Earth's atmosphere and its weather systems, and on the atmospheres of our neighboring planets. Such passive remote sensing studies are carried out using the observations made by diverse instruments on geostationary or polar-orbiting satellites or aboard spacecraft. Such studies also give us valuable information on the Earth's sea and land surfaces, such as its vegetation cover and the development of urban sprawl.

Radio waves are used to command satellite operations from the ground. They are also used in order to transmit scientific data from satellites, as modulated telemetry signals, to ground stations or to transmit very high frequency radio signals or TV broadcasts from space to different parts of the globe. Radio waves are essential for the operation of positioning and navigation services using satellites, such as provided by the GPS (Global Positioning System) signals now commonly received by the ubiquitous "sat nav" equipment in our cars. A thorough appreciation of the behavior and properties of electromagnetic waves is therefore crucial when discussing the performance and reliability of applications satellites.

The Fundamentals

Visible light is a wave motion, for which electric and magnetic fields oscillate extremely rapidly. Two cycles, or two periods, of a sinusoidally varying wave motion are illustrated in Fig. 27.1. In this diagram, the horizontal distance between two successive crests of the wave is one wavelength of the wave, which is denoted by the Greek letter λ. The vertical distance between the maximum and the minimum shown is twice the amplitude of the wave. The intensity, or power, of the wave is proportional to the product of the wave electric field amplitude and the wave magnetic field amplitude. Light carries energy radially away from a source, with the light rays traveling in straight lines. The wave electric and magnetic fields are perpendicular to each other, and they are both perpendicular to the direction of propagation of the wave. That result is obtained by solving Maxwell's four equations of electromagnetism (Grant and Phillips 1990).

Light travels at an incredibly high speed. In free space (a vacuum), the velocity of light, c, equals 3×10^8 meters per second (m/s), or 300,000 km/s. For yellow light, its wavelength λ has the value of 600 nm, that is, 6×10^{-7} m, or 0.6 μm. The velocity of light, c, is equal to the product of (i.e., is obtained by multiplying) the frequency f of vibration, or oscillation, of the wave (in cycles/s, now termed Hertz, Hz) and the wavelength λ (in m). The equation $f = c/\lambda$ is the fundamental relation connecting these three quantities.

Inserting the numbers for yellow light, $f = 5\times10^{14}$ Hz, 500 million million vibrations per second. Now blue light has a wavelength ~400 nm, and red light ~700 nm, which differ by nearly a factor of two – in between blue and red are all the colors of the rainbow. Their frequencies are $7.\times10^{14}$ Hz (blue), and $\sim4.3\times10^{14}$ Hz (red). We can only see radiation in the visible (Vis) part of the spectrum, with wavelengths between 400 and 700 nm. At shorter wavelengths, there is ultraviolet (UV) radiation, with infrared (IR) radiation at longer wavelengths, as shown in the third line down from the top of Fig. 27.2 that shows the entire electromagnetic spectrum.

The electromagnetic spectrum stretches from the lowest radio frequencies, 8 Hz (to the left of the bottom of Fig. 27.2), where the wavelength is equal to the circumference of the Earth (~40,000 km), through radio frequencies, ~100 MHz (or 10^8 Hz) where frequency modulated (FM) commercial radio stations operate, through the microwave (radar) part of the spectrum, at 10 GHz (10^{10} Hz, often termed X-band), having wavelengths ~3 cm, and up to the infrared. Beyond the visible, the electromagnetic spectrum stretches up in frequency another million times, through X-rays and gamma-rays, radiation whose wavelengths are as small as a millionth of the wavelength of visible light.

It might be helpful to some readers to provide a musical – acoustic – analog of the visible part of the electromagnetic spectrum. The note of middle C on the piano has a frequency of 256 Hz, and the C an octave above has a frequency twice that, namely, 512 Hz. The full range of a piano is from the lowest bass, three and a bit octaves below middle C, at ~25 Hz to the top treble (three octaves above 512 Hz)

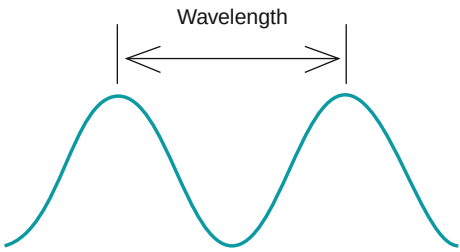


Fig. 27.1 Diagram showing one wavelength of a wave motion (From www.nrao.edu/index.php/learn/radioastronomy/radiowaves, courtesy of NRAO)

at ~4 kHz, that is, it covers a range of 160 times (1.6×10^2 times) in frequency. So the full electromagnetic spectrum from 10 to 10^{21} Hz, that is, over 20 orders of magnitude in frequency, is equivalent to having five more bass "pianos" and three more treble "pianos," as well as the actual piano! The electromagnetic spectrum indeed spans an incredibly wide frequency range.

At higher frequencies it is valuable to think of light as a particle as well as a wave. A beam of light is then represented as a stream of particles, each of which is called a photon. A photon is both massless and without an electric charge; it travels at the velocity of light. The energy of an individual photon, a quantum of energy, which is shown on the bottom line of Fig. 27.2, is a well defined quantity. It is given by the product of a fundamental constant known as Planck's constant (6.6×10^{-34} Joules.seconds, and abbreviated as Js) and the frequency of the electromagnetic wave of interest (in Hz). Therefore the energy of one photon is directly proportional to its frequency; for yellow light the energy of one photon is $\sim3\times10^{-19}$ J, which may also be expressed as ~2 electron Volts (eV). A photon of blue light has an energy that is almost twice that of a photon of red light. X-rays have very much greater photon energies, ~1–100 keV (thousands of eV), with gamma-ray photon energies exceeding 1 MeV, that is, >1 MeV (millions of eV).

Reflection, Refraction, Diffraction, Interference, and Polarization: Important Properties of Electromagnetic Waves

When a beam of light strikes a mirror, it is reflected. The angle that the beam makes to the perpendicular to the mirror surface (which is sometimes called the normal to the surface), called the angle of incidence (i), is equal to the angle which the reflected beam makes to the perpendicular, known as the angle of reflection (r_l). This relation shows a critical property of the phenomenon of *reflection*. For a rough surface, the reflected waves have different directions so that we often consider the waves to be scattered by the surface.

When electromagnetic waves travel through any material medium, they may be partially absorbed. For waves traveling through a transparent medium, such as glass, rather than through a vacuum, the velocity of light becomes less than c. In fact, we can write an equation for the velocity of propagation through a medium, v, as $v = c/\mu$,

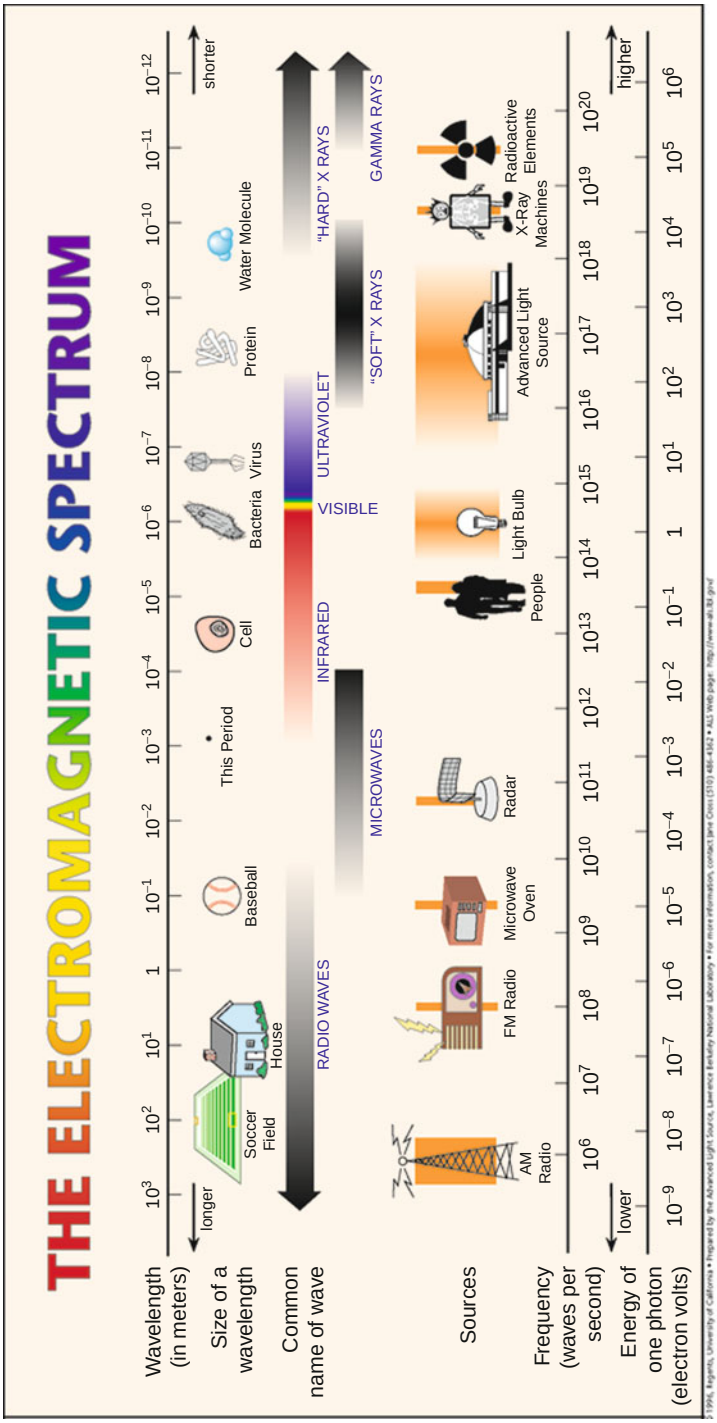


Fig. 27.2 Diagram showing the full electromagnetic spectrum, in all its glory; the wavelengths (*top*) range from 1 km on the left to 10^{-12} m on the right, and the corresponding frequencies, given at the *bottom*, go from radio waves of 300 kHz on the *left* to 3×10^{20} Hz gamma-rays on the right. The *second row from the top* illustrates different typical objects whose size is that of the corresponding wavelength. Different sources of electromagnetic radiation are shown above the frequency scale, which is itself above the scale specifying the energy (in electron Volts, eV) of one photon of that radiation (From www.lbl.gov/MicroWorlds/ALSTool/EMSpec/EMSpec2.html, courtesy of The Advanced Light Source, Lawrence Berkeley National Laboratory)

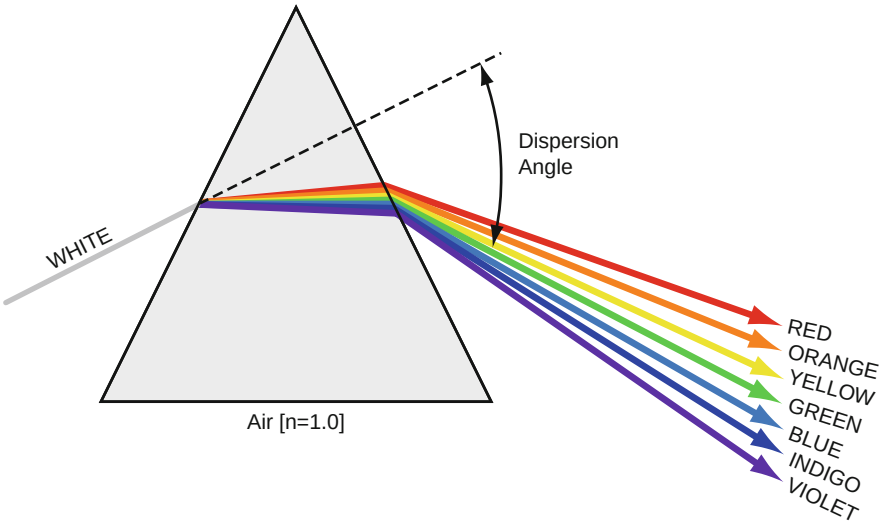


Fig. 27.3 Diagram illustrating how a glass prism separates a beam of white light into rays having all the colors of the rainbow. This is explained by the phenomenon of refraction (From www.thescienceclassroom.wikispaces.com, or www.heasarc.nasa.gov, courtesy of NASA)

where μ is called the refractive index of the medium; its value is always greater than 1. For glass, μ is $\sim$1.5, and its value is larger for blue light than for red light.

In the year 1665, Isaac Newton carried out experiments on passing a beam of sunlight, white light, through a glass prism. As Fig. 27.3 indicates, the light beam does not travel along the dashed line, but it is bent, that is to say it is refracted, so that it emerges from the prism at an angle to the dashed line; this is called the dispersion angle. The ray is bent toward the normal to the prism surface, making an angle to the normal of r_r. Going from air, whose refractive index is unity, to glass of refractive index μ, the relation $\sin i = \mu.\sin r_r$ applies. The blue light is dispersed more than the red light. The beam of white light is split into all the colors of the rainbow by the glass as Fig. 27.3 shows. Thus the phenomenon of *refraction* allows us to investigate the spectrum radiated by a light source of interest, such as a star.

Figure 27.4 shows the spectrum of light emitted by the Sun, from blue at 400 nm at the bottom to red at 700 nm at the top. Each of the 50 horizontal lines shows the spectrum for a width of only $\sim$6 nm. The Sun emits a broad – continuum – spectrum. On this continuum spectrum, dark – absorption – bands appear at generally very well-defined wavelengths. How these are formed will be mentioned later in this chapter.

Knowing about refraction makes it possible for us to design a lens which brings a beam of parallel light to a focus, for example, in the eyepiece of a telescope. With combinations of lenses, prisms, and mirrors we can design telescopes of several different types (e.g., Newtonian, Cassegrain, or Coudé) to view distant astronomical objects. These range in complexity and performance from the first telescope of Galileo Galilei made in 1609 to view the Sun, when he discovered sunspots that are dark regions on the solar surface, to the Hubble Space Telescope or the Chandra X-ray

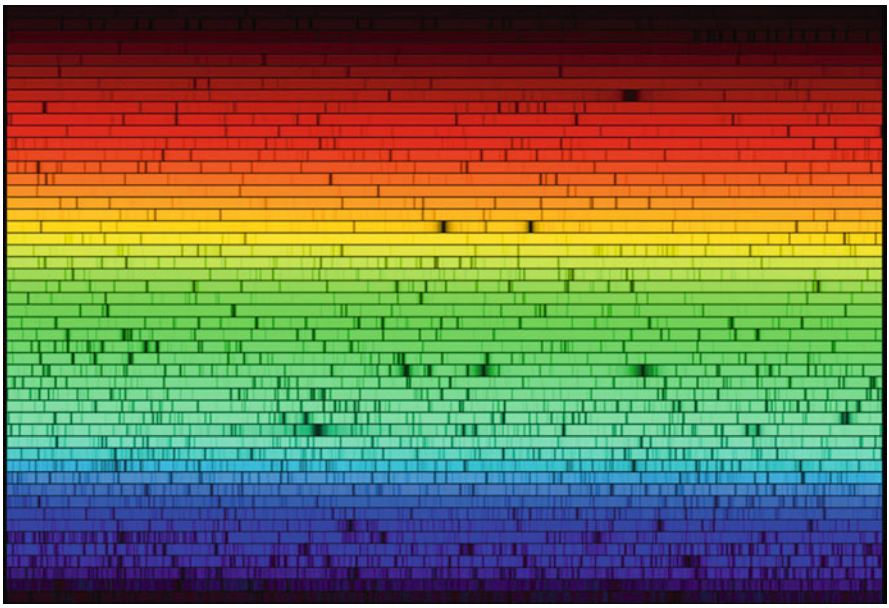


Fig. 27.4 The Sun's spectrum recorded by an instrument termed a spectrometer (From www.noao.edu, courtesy of NOAO/AURA/NSF)

telescope. Radio telescopes use metal parabolic reflectors to bring the radio beam from a satellite or a distant radio galaxy to a focus, where a sensitive receiver is placed.

The ionospheric plasma is the naturally occurring mixture of positive ions and electrons formed by the action of sunlight in the atmosphere at heights above ~80 km. The refractive index of this plasma is determined in part by the concentration of electrons. In order to travel through (i.e., not be reflected by) the ionosphere, the command and/or telemetry radio signals must have frequencies exceeding the largest value of the electron plasma frequency along the propagation path from the ground to the satellite, or vice versa. That means that their frequency must exceed ~30 MHz. If the radio frequency is much larger than that, say, ~1 GHz, the refractive index is only slightly larger than unity. In fact, there are two values of the refractive index; that is due to the presence of the Earth's magnetic field, which allows two types of wave to propagate. These are termed ordinary and extraordinary waves.

The performance of every telescope is limited by the phenomenon of *diffraction*. When light goes through a circular aperture or a slit, it does not travel straight through, but is diffracted by some small angle. This phenomenon of diffraction limits the resolving power of a telescope; it determines the angular separation between two nearby objects in the sky that can be distinguished from one another. The 2.4 m diameter of the Hubble Space Telescope determines that its resolving power is equivalent to resolving two pinpricks of yellow light only 1 mm apart at a distance of 2 km, an incredible achievement.

The phenomenon of diffraction and of *interference* between light waves enables a diffraction grating to be created. This acts like a prism, bending light. Diffraction gratings are often used as spectrometers.

Another important property of an electromagnetic wave is its *polarization*. For a wave whose electric field always lies in the same direction, the wave is said to be linearly polarized. Alternatively a wave whose plane of polarization rotates as the wave propagates is called a circularly (or elliptically) polarized wave. The rotation can be either clockwise or counterclockwise; this property is what causes ordinary and extraordinary waves to exist. During propagation, the plane of polarization rotates – this is called the Faraday rotation of the plane of polarization. It enables the total electron content along the radio path between a satellite and a ground station to be calculated.

The Doppler Effect

When an observer is moving at a velocity v relative to a source of light, or if the source is moving relative to the observer, then the observer will notice a change in the wavelength of the light. Motion along the line of sight, away from the observer, causes an increase of the wavelength – this is termed a red shift. However, if the motion is toward the observer, a decrease of the wavelength is caused, termed a blue shift. This phenomenon is known as the Doppler effect. The reader may be more familiar with the acoustic analog. The siren of a police car approaching the observer increases in pitch, or frequency, whereas when it is moving away the frequency decreases below the transmitted frequency.

A useful equation is that the magnitude of $\delta\lambda/\lambda = \delta f/f = v/c$, for values of v which are very much less than ($<<$) c. Measuring the Doppler frequency shift of well-known spectral lines (discussed in the next section) leads directly to an estimate of the source velocity. The application of this result has demonstrated that the Universe is expanding, and that the velocity of more distant objects is larger than that for nearer objects. A space applications example of the Doppler effect is the changing frequency of a radio signal transmitted by a polar-orbiting satellite traveling at 7 km/s; it is increased by up to $\sim$7/300,000, or $\sim$0.002%, as it approaches the ground station, and is decreased by that amount as the satellite moves away.

Multiwavelength Studies and Black Body Radiation

Dust, clouds, and various molecular gases, including water vapor, the most important "greenhouse gas," and carbon dioxide in the Earth's atmosphere absorb much of the Sun's infrared radiation. Molecular oxygen and ozone in the stratosphere absorb almost all the dangerous – to the DNA molecules in our bodies – ultraviolet radiation from the Sun. Therefore, the only way to carry out experimental gamma-ray, X-ray, ultraviolet or infrared astronomical studies is to put an instrument aboard a rocket, satellite, or spacecraft. Space technology is essential for such fundamental scientific studies.

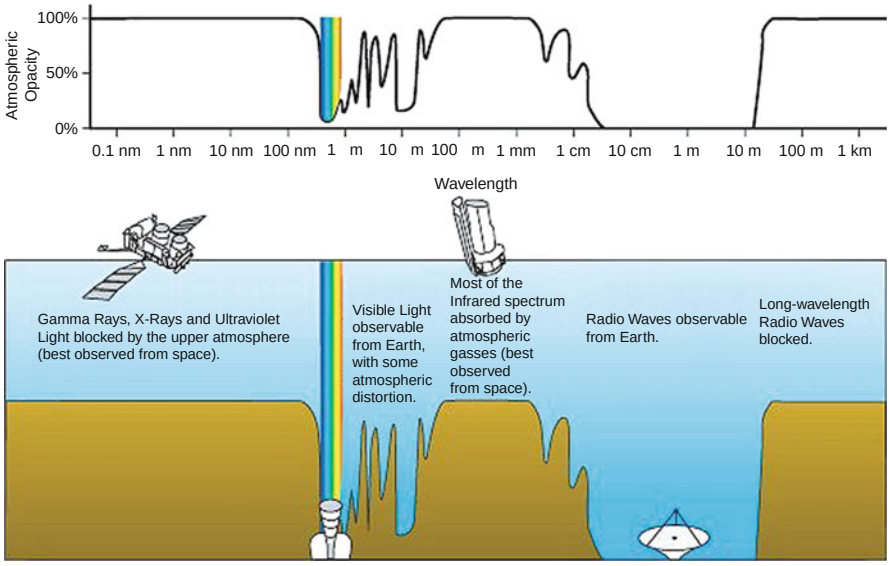


Fig. 27.5 The upper panel shows the percentage absorption (or opacity) due to the atmosphere, which is 100% for gamma-rays, X-rays, and for ultraviolet, infrared radiation and very long wavelength radio waves. The lower panel demonstrates why space technology is required for fundamental scientific studies in these spectral ranges (From http://www.newworldencyclopedia.org/entry/Electromagnetic_spectrum)

Where in the spectrum there is little absorption of radiation is termed a spectral window. The visible (Vis) part of the spectrum is one such region; that is clearly shown between 400 and 700 nm in the upper part of Fig. 27.5. There is a broad radio window at wavelengths between 0.1 and 10 m, which allows ground-based radio telescopes to operate, as illustrated in the lower part of Fig. 27.5. Longer wavelength radio waves cannot penetrate the ionosphere; they are reflected by the ionosphere.

Figure 27.6 plots the atmospheric opacity quantitatively as a function of frequency. An absorption of 1 dB (decibel), shown here as 1.E + 00, is rather negligible. However, 20 dB (two times 1.E + 01) shows absorption of the power of the radio signal by one order of magnitude (i.e., ten times); this would be noticeable. Absorption of 100 dB (i.e., by five orders of magnitude, a hundred thousand times, 1.E + 02) would be devastating for any communications system. Thus, this plot informs a communications engineer about those frequencies (such as around 60 GHz and near 118 GHz) which should definitely *not* be chosen for an effective communications system.

There are three basic types of spectra – the continuous spectrum, the emission (or bright-line) spectrum, and the absorption (or dark-line) spectrum. The *continuous spectrum*, an uninterrupted sequence of wavelengths, is a broadband spectrum; it is emitted by a hot gas at high pressure (e.g., the photosphere at the Sun's surface). The *emission spectrum* emitted by an atomic gas (e.g., hydrogen) at low pressure, that is, under rarefied conditions, is a set of discrete bright narrow lines. A particular

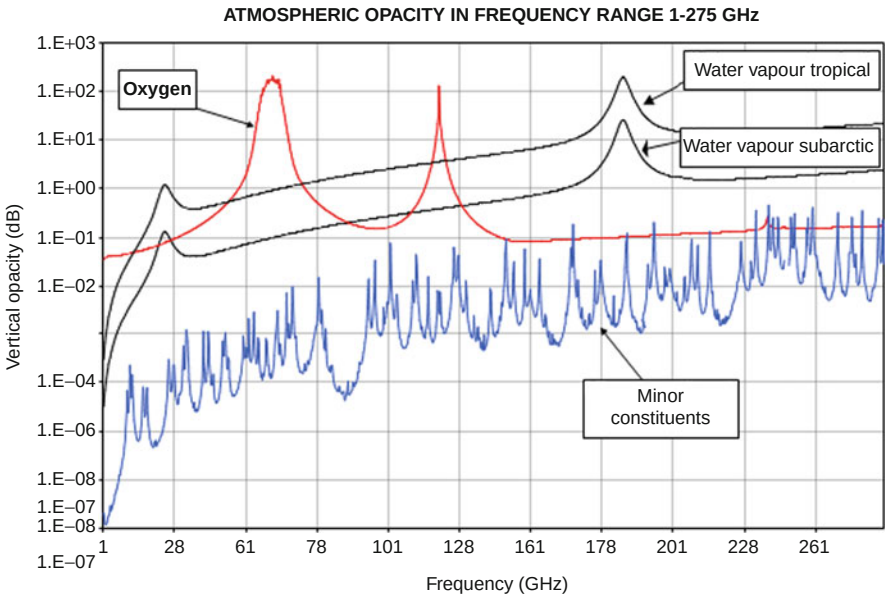


Fig. 27.6 The atmospheric opacity in the radio part of the spectrum from 1 to 275 GHz; here molecular oxygen and water vapor are the agents most responsible for the absorption (From ITU Report, RO7-SG07-C-0104!!MSW-E, page 11)

spectral line is radiated when an electron in a certain excited state falls back to a lower level state or to its ground level state. (An electron is said to have been excited when a photon collided with it, in an atom or molecule, and raised it to a higher energy level.) Spectral lines occur at well defined wavelengths; they are a characteristic of the particular chemical species radiating them. For molecular gases, the emissions are broader bands rather than narrow lines.

The *absorption spectrum* is observed when light from a bright source producing a continuous spectrum passes through a cooler gas which absorbs its characteristic line radiation. Radiation at these well defined wavelengths is removed from the spectrum. Figure 27.4 presented an example of an absorption spectrum for the Sun. Radiation is absorbed by the cooler gases in the chromosphere just above the solar surface (the photosphere). The thousand or so absorption lines evident in Fig. 27.4 are called Fraunhofer absorption lines. The most dominant pair of absorption lines is near the center of Fig. 27.4, in the yellow part of the spectrum, at wavelengths of 589.0 and 589.6 nm. These are due to absorption by atoms of sodium in the solar atmosphere.

The intensity – or brightness – spectrum of a continuum spectrum known as black body radiation, Planck's radiation law, is plotted as a function of wavelength in Fig. 27.7. A black body is an object that absorbs all the radiation which is incident upon it; it reradiates that energy as radiation which depends solely on the temperature of the object. Stefan-Boltzmann's law states that the total brightness of a black body (the area under the curve) is proportional to the fourth power of the absolute

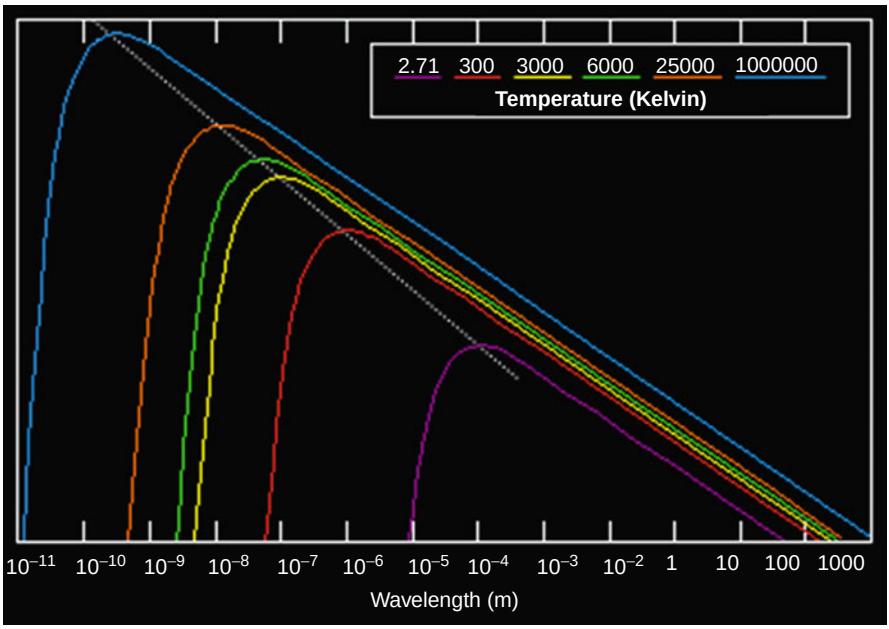


Fig. 27.7 The intensity (or brightness, on a vertical logarithmic scale of over 30 orders of magnitude) of the continuum spectrum (shown on the horizontal axis using a logarithmic scale from 10^{-11} to 1,000 m, a staggering wavelength range of 14 orders of magnitude) radiated by a black body at different temperatures. These range from microwaves approaching the far infrared, at 2.71° absolute (Kelvin, K), which is the temperature of the cosmic microwave background radiation remaining from the "big bang" origin of the Universe, to X-rays with wavelengths <1 nm at a temperature of a million degrees, which is the temperature of the Sun's outermost atmosphere, the corona (From http://sci.esa.int/science-e/www/object/index.cfm?fobjectid=35774&fbodylongid=1696, or from the encyclopedia of science, www.daviddarling.info)

temperatures of that body. The wavelength of the spectral peak also depends on the temperature, expressed in degrees on the absolute (Kelvin) scale. On this scale, the freezing point of water is 273 K and its boiling point is 373 K. The origin of the Kelvin temperature scale is at −273°C; a negative absolute temperature is impossible.

The range of wavelengths displayed on a logarithmic scale in Fig. 27.7 is enormous, 14 orders of magnitude. The range of intensities, also shown on a logarithmic scale (but not given quantitatively), is even larger, at least 30 orders of magnitude. The curve shown at 300 K is representative of the black body radiation emitted by the Earth and its atmosphere into space. Most of this radiation lies in the infrared part of the spectrum at wavelengths of ∼1–10 μ. The curve at 6,000 K is close to the Sun's continuous spectrum; here most radiation is emitted at wavelengths <1 μ. The curve at a million K peaks in the X-ray part of the spectrum. The Sun's outer atmosphere, termed the corona, together with compact and very energetic stars, such as neutron stars, or material falling into a black hole, all emit X-rays. X-ray telescopes in space such as the Chandra Observatory investigate

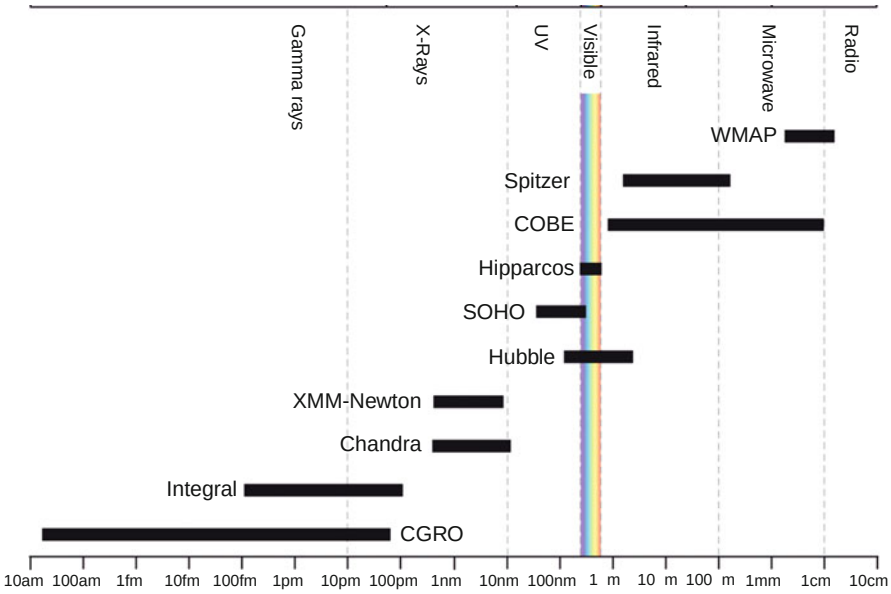


Fig. 27.8 Diagram indicating the wavelength ranges over 16 orders of magnitude (from 10 attometers, 10 am, or 10^{-18} m, to 0.1 m, or 10 cm) investigated by different space-borne instruments and missions. CGRO refers to the NASA Compton gamma-ray observatory, XMM to the ESA X-ray Multi-Mirror Newton observatory, and SOHO to ESA's solar and heliospheric observatory; COBE (Cosmic Background Explorer) and WMAP (Wilkinson Microwave Anisotropy Probe) both study the cosmic microwave radiation from the "big bang" origin of the Universe (From L.L. Christensen, R. Hurt, R. Fosbury, Hidden Universe, Wiley-VCH)

these areas. Figure 27.8 presents the spectral regions studied by the Chandra Observatory and by other space missions.

The curve below the Earth's spectrum in Fig. 27.7 is for black body radiation at 2.71 K. This radiation in the microwave part of the spectrum comes from the remnant of the "big bang" origin of the Universe 13.7 billion years ago. The dashed line in Fig. 27.7 shows Wien's displacement law; this law states that the wavelength of the most intense emission, the peak, is inversely proportional to the absolute temperature. The three fundamental laws of black body radiation introduced here can be proved only on the basis that the concept of radiation as streams of photons is valid.

Figure 27.9 presents a remarkable composite view of the color-coded intensities of electromagnetic radiation coming from all directions of our galaxy, the Milky Way, in different wavelength regions. The most energetic gamma-rays are shown at the bottom, with radiation at X-ray wavelengths (e.g., from neutron stars) the next up, then visible radiation, and then infrared radiation from stars that are cooler than our Sun (in three different wavelength regions). Further up in Fig. 27.9 are displayed intensity maps of radio emissions at various frequencies, including that for atomic hydrogen at 1.42 GHz (at the wavelength of 0.21 m)

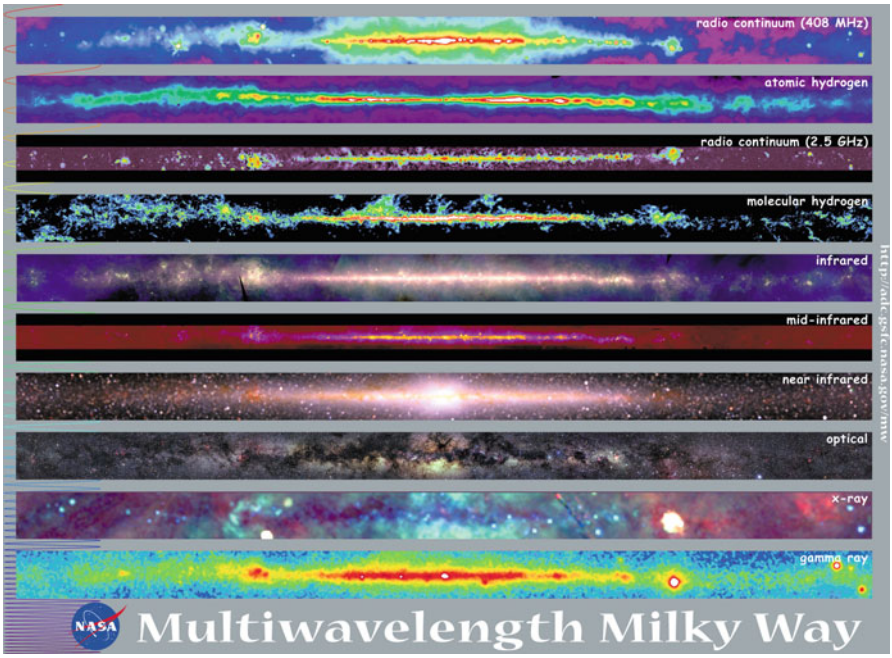


Fig. 27.9 A view of the entire Milky Way, our galaxy, shown in ten different wavelength regions; the most energetic processes are shown at the *bottom* and the least energetic at the *top*. Some observations are made from the ground and some from space. The central plane of the galaxy is clearly evident in all images, with the galactic center being especially evident in the top and bottom images (From http://son.nasa.gov/tass/content/emspec.htm, courtesy of NASA)

which is emitted in the cold interstellar medium; it is strongest in star-forming regions. The lowest energies (lowest frequencies, and longest wavelengths) are evident at the top. At all wavelengths, there is more radiation coming from the plane (disk) of the galaxy, where most of the $\sim 10^{11}$ stars are to be found. The radiation is especially strong toward the center of the galaxy, at the midpoint of these horizontal images.

Another Effect Due to Photons, the Photoelectric Effect

Heinrich Hertz observed in 1887 that a charged metal surface exposed to ultraviolet light lost its electric charge. He had found that the illuminated metal emitted electrons; this effect is called the photoelectric effect. In 1921, Albert Einstein was awarded the Nobel Prize for physics for his 1905 theory that explained the photoelectric effect. His theory requires that light is considered as a beam of photons, each of which has an energy equal to hf (in J), f being the frequency of the light and h Planck's constant. Electrons which are bound within metal atoms require at least this minimum energy, = hf, a few electron Volts (eV), to be ejected as photoelectrons from the

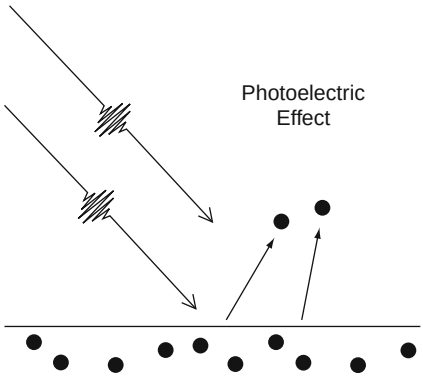


Fig. 27.10 Diagram showing two photons (wave packets) of ultraviolet light striking a metal surface and ejecting electrons (*dark dots*) from it, illustrating the photoelectric effect (From www.canadaconnects.ca/quantumphysics/1078/)

surface of the metal (Fig. 27.10), and to cause a photoelectric current to flow. A lesser photon energy causes no photoelectric effect whatsoever. A greater intensity of light whose energy exceeds the threshold energy causes a greater photoelectric current. Different metals (different elements) have different threshold energies.

The photoelectric effect causes a satellite or spacecraft having a metal surface which is exposed to the UV radiation and X-rays contained in sunlight to gain a positive electric charge, through the loss of negatively charged electrons. There is a corresponding negative charge in the surrounding plasma in regions of shadow. If this charge becomes large, an electric discharge – a mini lightning discharge – can occur between the two electric charges of opposite sign; that can damage sensitive electronic equipment aboard the satellite. Other sources of danger to equipment operating in space aboard satellites and spacecraft are considered in the ▶ Chap. 47, "Space Weather and Hazards to Application Satellites."

Conclusion

In this chapter, we have considered the principles and fundamental concepts of light and electromagnetic radiation as they pertain to the performance and reliability of applications satellites. The topic is a central and essential one. Visible light is the best known example of electromagnetic radiation. Its narrow spectrum from blue to red (with wavelengths varying by less than a factor of two) contrasts markedly with the enormously broad electromagnetic spectrum stretching from radio waves to gamma-rays (where the range of frequencies covered is by 20 orders of magnitude).

We have summarized both the wave and the particle (photon) properties of light and, more broadly, of electromagnetic radiation. We have discussed six different phenomena occurring when light interacts with matter, as well as the concept and basic properties of black body radiation. We have explained the Doppler effect. Almost all our knowledge of the Universe, and much of our understanding of the Earth's and other planetary atmospheres, derive from the passive reception of electromagnetic radiation emitted by atoms and molecules in stars, other celestial objects and the atmospheres of planets. Both one way and

two way radio communications between the ground and rockets, satellites, or space probes, and also the operation of GPS systems and radar altimeters, depend upon the active generation, transmission, and reception of electromagnetic waves.

References

I.S. Grant, W.R. Phillips, *Electromagnetism*, 2nd edn. (Wiley, Chichester, 1990), p. 542

Astronaut Photography: Handheld Camera Imagery from Low Earth Orbit

28

William L. Stefanov, Cynthia A. Evans, Susan K. Runco, M. Justin Wilkinson, and Kimberly Willis

Contents

Introduction 684
Historical Overview of Astronaut Photography 685
USA Space Program 685
USSR/Russian Space Program 688
Specifications of Astronaut Photography 689
Cameras, Film, and Digital Media 689
Platforms 706
Remote Sensing Using Astronaut Photography 710
Advantages and Limitations 710

W.L. Stefanov (✉)
Science Applications Research and Development, Code KA/ESCG, Astromaterials Research and Exploration Science Directorate, NASA Johnson Space Center, Houston, TX, 77058, USA
e-mail: william.l.stefanov@nasa.gov

C.A. Evans
Code KT, Astromaterials Research and Exploration Science Directorate, NASA Johnson Space Center, Houston, TX, USA
e-mail: cindy.evans-1@nasa.gov

S.K. Runco
Code KX, Astromaterials Research and Exploration Science Directorate, NASA Johnson Space Center, Houston, TX, USA
e-mail: susan.k.runco@nasa.gov

M.J. Wilkinson • K. Willis
Code KX/ESCG, Astromaterials Research and Exploration Science Directorate, NASA Johnson Space Center, Houston, TX, USA
e-mail: justin.wilkinson-1@nasa.gov, kim.willis-1@nasa.gov

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_39, © Springer Science+Business Media New York 2013

Public and Education Outreach Applications 718
Public Outreach 720
Education Outreach 721
References 723

Abstract

Photographic observations of the Earth by humans in low earth orbit, in contrast to unmanned orbital sensor systems, began during the 1960s as part of both the USA and former USSR manned space flight programs. The value of regularly repeated photographic observations of the Earth from orbit was demonstrated by later long-duration missions and led directly to the development of unmanned, multispectral orbital sensors such as the Multispectral Scanner and Thematic Mapper onboard the Landsat series of satellites. Handheld imagery of the Earth has been continually acquired during both USA and USSR/Russian space station and Space Shuttle programs, and represents a rich dataset that complements both historical and current unmanned sensor data for terrestrial studies. This chapter provides an overview of astronaut/cosmonaut imagery and development of specific data collection programs, then moves on to discussion of technical aspects of both the historical film and current digital cameras used in orbit with information on how to access online datasets. Case studies are presented to highlight varied applications of handheld imagery for terrestrial research and natural hazard response. The chapter concludes with discussion of future directions for digital handheld imagery of the Earth from manned orbital platforms such as the International Space Station.

Keywords

Apollo • Astronaut • Atmospheric science • Camera • Cosmonaut • Ecology • Film • Geography • Geology • Hazards • Human space flight • Hydrology • International space station (ISS) • Meteorology • Mir • National aeronautics and space administration (NASA) • Oceanography • Skylab • Space shuttle • Urban

Introduction

Remote sensing – the detection of surface-material properties such as composition and texture without physical interaction with the material – is an important analytical approach for investigating and monitoring planetary surface processes. For many Earth scientists, remotely sensed data is strongly associated with unmanned satellites. Remotely sensed data typically is collected by automated sensors on satellites in high polar and sun-synchronous orbits at approximately 700–900 km altitude above the earth's surface. Following the advent of civilian, government-supported orbital platforms in 1972 (e.g., Landsat) remotely sensed data has become an important resource for Earth scientists.

There is another spatially and temporally extensive remotely sensed dataset available for terrestrial studies and applications: systematic photographic images

of the earth taken by astronauts from the NASA Mercury missions of the 1960s to the present International Space Station (ISS) crews. The astronaut photography dataset covers much of the earth's land and coastal surface, as well as atmospheric phenomena such as hurricanes and aurora. Unlike the unmanned satellite-based sensors mentioned above, astronauts use off-the-shelf film and digital cameras to image the earth, rather than mission-specific instruments. Such equipment limits astronaut photographs to the visible and near-infrared wavelengths in three bands (red, green, blue, and/or near-infrared with appropriate filters and films), similar to what is collected by aerial photograph surveys. The majority of astronaut photographs were taken from altitudes of 300–400 km (185–250 miles) – the most notable exception being the Apollo missions to the Moon during 1969–1972. Currently the ISS is the primary manned platform for astronaut photography, which is acquired exclusively with digital cameras.

Today, the process of acquiring high-quality and scientifically useful images of the earth begins with astronaut training in the Earth Sciences. This training is provided for ISS crews by the Crew Earth Observations (CEO) team at the NASA Johnson Space Center in Houston, Texas (Stefanov 2008). Experts in a variety of disciplines provide ISS crewmembers with the science background appropriate to a variety of research image foci including glaciers, urban areas, long-term ecological monitoring sites, volcanoes, aurora, coral reefs, and deltas (both coastal and inland). This training provides the astronaut with the scientific context of a given image request. Experience with numerous crew members over the years indicates that intellectually engaged crew members tend to take higher quality data.

Collection of digital imagery from manned space vehicles such as the Space Shuttle and ISS is request-driven, with specific operational procedures and constraints. Each day CEO personnel examine the predicted ISS orbital path for the coming 24 h to determine which science targets may be visible to the crew. This list is then filtered by predicted cloud cover and illumination conditions, daily crew schedule, and the ISS ground track relative to the target's latitude/longitude position. The resulting target list is then augmented with additional instructions and data – such as landmark features to aid in locating a target, desired camera focal lens, additional imagery, etc. – and uplinked to the ISS crew. Returned imagery of targets is routinely reviewed to determine whether imagery objectives have been met. The ability to request targets of opportunity, such as volcanic eruptions and hurricanes, adds a near-real-time response capability to the ISS for hazard detection and management.

Historical Overview of Astronaut Photography

USA Space Program

Observing the Earth from space is one of NASA's longest-standing science experiments. For nearly 50 years, US astronauts have trained for and performed

observations of the Earth from orbiting spacecraft, starting with the Mercury missions and continuing today on the International Space Station. The legacy of NASA's Astronaut Earth Observations program is evident in the wealth of Earth observation data now available to the global citizenry. Viewing the Earth from space is now routine and fully integrated into our daily habits due to widespread Internet access, advances in both data visualization and serving technologies, and ready availability of sophisticated personal computing devices such as laptops, tablets, and smart phones.

Since the beginning of the US human spaceflight programs in the early 1960s, a broad community of scientists has been involved in defining Earth observations objectives and training crew members about Earth processes. Support for Earth observations increased rapidly during the 1960s to support development of new unmanned sensors, and was reinforced by the intensive geological training for the future Apollo program. Other authors have written extensively about these early programs, e.g., Amsbury (1989) and references therein; (Compton et al. 1989; Wilhelms 1993).

Handheld photography from space started in the Mercury program when payload volume expanded to include cameras and film for photographing the Earth. Structured crew Earth observations developed quickly during the Gemini program, and resulted in dedicated experiments to develop systematic capabilities. The new views of the Earth from space excited both the public and scientists, and were published in several technical reports (Anonymous 1968; Derr et al. 1972; Ewing et al. 1965; Foster 1967; Foster and Smistad 1967; Gill 1967; Lowman 1965, 1969b) and popular venues, e.g., Lowman (1966a). The early Apollo test flights were used to evaluate specific imagery applications (stereo photography, documentation of changes, weather photography) and also functioned as a test bed for demonstrating the capabilities of different films (color infrared and filtered multispectral grayscale) for future multispectral sensors on satellites, such as the US Landsat program (Amsbury 1989; Kaltenbach 1970; NASA 1970a). During this time the astronauts worked closely with investigators to obtain the best images for these projects, demonstrating the success of interactive teams of scientists and astronauts.

By the mid-1960s, crews were actively engaged in geology training in key locations in the western USA. Even as the training for future lunar missions accelerated, NASA maintained support for crew Earth observations to support ongoing scientific investigations. Crew Earth observations peaked during the Skylab missions, supporting the highly successful Earth Resources Experiment Program (EREP). EREP involved scores of scientists who selected sites, conducted field studies, trained the crews, and analyzed the orbital data (NASA 1974b, 1977). A fuller account of the early Earth observations training programs and additional references is contained in (Amsbury 1989).

The Skylab EREP program was succeeded by the Space Shuttle Earth Observations Program (SSEOP), established in the early 1980s. The early activities of the office supporting this program are described in (Helfert and Wood 1989). Astronaut training in the Earth Sciences was a prime function, with the goal of providing

critical scientific background to the crews to enable the acquisition of scientifically meaningful astronaut photography. The office staff included scientists from several disciplines in the Earth Sciences (meteorology, geology, oceanography, ecology, and geomorphology); they trained the crews, evaluated target requests, and planned and coordinated observation campaigns for each flight. Throughout the 1980s and 1990s, the SSEOP obtained a broad set of medium-resolution imagery using a variety of medium and large format cameras that complemented data from the unmanned Earth observations satellites of the time. Some flights supported specific data collection campaigns such as the Shuttle Imaging Radar missions (Cimino et al. 1986; Elachi et al. 1982; Evans et al. 1997) and the LIDAR In-space Technology Experiment, or LITE (Winker et al. 1996). Science targets and associated equipment (cameras, lenses, film complement) were selected based on the mission's orbit tracks, projected lighting, season, spacecraft altitude, and current events. Staff scientists provided last-minute updates to the crew a few days before flight, such as overviews of weather systems and noteworthy events (e.g., volcanic activity, biomass burning locations, tropical storms, and floods). During flight, the science team on the ground communicated with the crew daily about upcoming sites, global weather patterns, and additional targets based on current events.

The Shuttle-Mir program in the mid-1990s, followed by early missions to the ISS in 2000, allowed NASA's Earth scientists' insight into differences between shorter Space Shuttle flights and longer duration missions on a space station in terms of observational styles and appropriate scientific targets. In contrast to the SSEOP science strategy, based on the spaceflight parameters of short (less than 2 week duration) Space Shuttle missions and built around specifics of each flight (time of launch, season, mission objective, basic spacecraft attitude relative to Earth, orbital parameters including inclination and altitude), a modified scientific approach was adopted for astronauts and cosmonauts flying long-duration missions to the Russian Mir Space Station from 1996 to 1998. Building from both the Shuttle experience and Russian Earth Science approach (Glazovskiy and Dessinov 2000), the Shuttle-Mir Earth Observations strategy adjusted for the set orbital inclination, the changing views with seasonal shifts and changing solar illumination (Mir, like the current ISS, was not sun-synchronous). The Shuttle-Mir program segued directly to the current CEO program described above.

Over time, the ISS CEO program has increasingly favored an interdisciplinary Earth Science approach, integrating other remotely sensed data and new emphasis on coastal processes, human footprints, and environmental change into both crew training and target selection. Astronauts from ISS partner organizations, such as the Canadian Space Agency and European Space Agency, are occasionally trained in Earth observations as part of the CEO program as well. Recent NASA planning for potential return missions to the Moon, Mars, or Near-Earth Objects has revived more intensive geological and remote sensing training for incoming astronaut-candidate classes based in part on the Apollo-era training program, with the expectation of additional benefits to Earth observations (Evans et al. 2011).

USSR/Russian Space Program

The history of the Soviet Earth observations program during the early years (1960s) mirrored the US Earth Observations program, starting with Yuri Gagarin's flight in 1961. In the beginning much of the cosmonaut photography of the Earth was initiated by cosmonaut interest, rather than a systematic program-level approach. The early cosmonauts, using binoculars and handheld cameras, identified the potential for global, real-time observations of the planet (including dynamic events); general characterization of the Earth's surface from orbit, particularly study of geographically inaccessible regions; and the potential for comparative studies across different regions.

Glazovskiy and Dessinov (Glazovskiy and Dessinov 2000) detail the evolution of the cosmonaut-based Earth observations program in the Soviet space program through the 1960s, 1970s, and 1980s. The Earth Observations program structure was established by 1974 to support the Salyut 3–5 missions. The program was designed in collaboration with geographers at the Academy of Sciences, the Chief of Cosmonauts, the Soviet Air Force, and the RSC-Energia engineers and film experts. The cosmonauts were trained in the Earth Sciences and conducted experiments with powerful tracking binoculars equipped with cameras. Technologies and procedures were developed, including improvised motion tracking using attitude control mechanisms on the spacecraft. The capability for larger payloads enabled the use of large format cameras (e.g., KATE-140 with an 18 × 18 cm film format and longer lenses) complemented by conventional 35 mm and medium format cameras. Soviet scientists collaborated with scientists and engineers in the German Democratic Republic to develop the program and equip the spacecraft with camera systems.

The USSR Earth observations scientific program ramped up between 1977 and 1985 to support the long-duration missions on the Salyut 6 and 7 space stations. The program collaborated with dozens of institutions, and included extensive cosmonaut training in state-of-the-art facilities at the Gagarin Training Center outside of Moscow. Earth observation training flights in Tupolev-134 aircraft employed a strategy that involved close work between cosmonauts and scientists. Cosmonauts on the first Salyut 6 mission were tasked with building and refining the Earth observations procedures and testing the equipment, including fixed high-resolution nadir-looking cameras (KATE-140 and MKF-6 M cameras), powerful binoculars, and handheld cameras. The cosmonaut crews were actively involved in calculating manual controls for motion tracking by maintaining spacecraft attitude with thrusters as a means of damping out the relative ground motion. Through this method, they collected high-resolution photography of designated earth targets. During these early space station missions, cosmonauts started observing ephemeral features and events on the Earth's surface, conducted studies on changes in human visual acuity in space, and conducted simultaneous observations with aircraft surveys.

The Mir program followed in 1986, lasting through the last mission in 2000. The Earth observations program on Mir followed the strategy developed for the Salyut program, although financial support for scientists and equipment dropped during

the collapse of the USSR and the following difficult economic times lasting through the mid-1990s. The Earth observing assets on the Mir included the KATE and MKF-6 M fixed cameras, conventional handheld cameras, medium format cameras, and a variety of films including a unique color infrared film. The Priroda module, a dedicated Earth Resources facility, was launched to the Mir Station in 1996 with an international complement of remote sensing equipment. However, heavy power constraints precluded full operations of many of the Priroda instruments, and the capabilities were not realized.

Cosmonaut crews on the International Space Station are currently performing a variety of Earth observations tasks and experiments, including the "Uragan" program that began in 2001 (I. Sorokin, S.P. Korolev, 2010, personal communication). This program is similar in many ways to the early NASA Apollo and Skylab EREP programs, in that it uses digital still cameras to optimize design and test of multispectral sensor configurations for monitoring and forecast of land surface processes and hazards from space.

Specifications of Astronaut Photography

Handheld Earth imagery has been acquired by a variety of film and digital cameras over the past 50 years, from both short- and long-duration missions of spacecraft in low earth orbit. We focus our overview of astronaut photography specifications on NASA missions, spacecraft, and equipment due to the availability of extensive and accessible documentation. A discussion of USSR/Russian handheld photographic equipment up to 2000 is presented in Glazovskiy and Dessinov (2000).

In contrast to purpose-built and mission-specific multispectral and hyperspectral imaging systems for airborne and orbital platforms, the majority of cameras, binoculars, etc. used for handheld astronaut Earth photography has been commercial off-the-shelf equipment that has seen little or no modification prior to launch and on-orbit use. Cameras or imagers meant for use during extravehicular activities (e.g., Apollo and ISS astronaut suit cameras) were modified both for space hardening – reduced outgassing and operation in the vacuum and hard radiation environments – and for the operational challenges posed by space suit design and ergonomics, e.g., operating the camera with pressurized suit gloves in low or zero gravity. A full discussion of lunar orbital and surface astronaut photography, training, and equipment is outside the scope of this chapter, but the interested reader is directed to Jones et al. (2010) and Woods (2009) for more information.

Cameras, Film, and Digital Media

Film cameras have been used in all of the NASA historical and current manned spaceflight programs, including on the Space Shuttle and International Space Station, with a variety of film types used for Earth observations (Tables 28.1–28.6). The Astronaut Office, Orbiter Photo/TV Working Group, and the SSEOP, in

Table 28.1 Specifications of handheld earth observations equipment and activities during the NASA Mercury program

Mission (dates flown)	Camera/lenses/other equipment	Films/filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Mercury 3 (5 May 1961)	Maurer model 220 G 70 mm camera (fixed camera system)	Super Anscochrome	None	None	None	General information photography	More than 150 photographs taken of sky, clouds, and ocean	Lowman (1964, 1965, 1966a, b), Short and Lowman (1973), Swenson et al. (1966), Underwood (1968)
Mercury 4 (21 July 1961)	Mauer model 220 G 70 mm camera (fixed camera system)	Super Anscochrome	None	None	None	General information photography	Photography taken over the Atlantic Ocean, North and Central Africa	Grimwood (1963), Lowman (1964, 1965, 1966a, b), Short and Lowman (1973), Swenson et al. (1966), Underwood (1968)
Mercury 6 (20 February 1962)	Ansco Autoset 35 mm camera/ 55 mm lens	Eastman color negative type 5250	None	None	None	General information photography of Earth, meteorological photography planned	48 photographs taken of clouds, oceans, northwest Africa	Cortright (1968), Grimwood (1963), Kuehnel (1972), Lowman (1964, 1965, 1966a, b), Short and Lowman (1973), Swenson et al. (1966), Underwood (1968)
Mercury 7 (24 May 1962)	Robot recorder 35 mm camera	Eastman color negative type 5250	None	None	None	Photograph selected areas for terrain	155 photographs taken of Earth and Earth's limb	Cortright (1968), Grimwood (1963), Kuehnel (1972), Lowman (1964, 1965, 1966a, b), Short and Lowman (1973), Swenson et al. (1966), Underwood (1968)
Mercury 8 (3 October 1962)	Hasselblad 500C 70 mm camera (NASA modified)/ Zeiss 80 mm planar f/ 2.8 lens	Super Anscochrome	None	None	None	Synoptic terrain photography experiment	Photography obtained of poor quality due to general overexposure	Grimwood (1963), Kuehnel (1972), Lowman (1964, 1965, 1966a, b), Short and Lowman (1973), Swenson et al. (1966), Underwood (1968)
Mercury 9 (15–16 May 1963)	Same as Mercury 8	Ultraspeed Anscochrome (fpc-289)	None	None	None	Same as Mercury 8	29 photographs taken of Earth and cloud features, including observations on visibility and color of Earth features	Cortright (1968), Grimwood (1963), Kuehnel (1972), Lowman (1964, 1965, 1966a, b), Short and Lowman (1973), Swenson et al. (1966), Underwood (1968)

Table 28.2 Specifications of handheld earth observations equipment and activities during the NASA Gemini program

Mission (dates flown)	Camera/lenses/other equipment	Films/Filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Gemini 3 (23 March 1965)	Hasselblad 500C 70 mm camera (NASA modified)/Zeiss 80 mm planar f/2.8 lens	70 mm, Ektachrome MS SO-217/Haze filter	None	None	None	General terrain and meteorology information photography	219 Earth-looking photographs taken	Carter and Stone (1974), Giddings (1975), Grimwood (1969), Kuehnel (1972), Lowman (1966a, 1968), Lowman and Tiedemann (1971), NASA (1966, 1967a, b), Short and Lowman (1973), Underwood (1968)
Gemini 4 (3–7 June 1965)	Hasselblad 500C 70 mm camera (NASA modified)/Zeiss planer 80 mm f/2.8 lens	70 mm Ektachrome MS SO-217 and 3401	None	None	None	Synoptic terrain photography experiment S005, Synoptic weather photography experiment S006	219 Earth-looking photographs taken	Badgley et al. (1969), Carter and Stone (1974), Cortright (1968), Giddings (1975), Grimwood (1969), Kuehnel (1972), Lowman (1966a, 1968, 1972), Lowman and Tiedemann (1971), Lowman et al. (1967), NASA (1966, 1967a, b, c), Short and Lowman (1973), Stevenson et al. (1968), Underwood (1968), Wobber (1968), Zeitler et al. (1971)
	Zeiss Contarex 35 mm camera/250 mm lens	Anscochrome D-200 film/Haze filter						
	Maurer 16 mm motion picture camera	Ektachrome film						

(*continued*)

Table 28.2 (continued)

Mission (dates flown)	Camera/lenses/other equipment	Films/Filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Gemini 5 (21–29 August 1965)	Hasselblad 500C 70 mm camera (NASA modified)/Zeiss planer 80 mm f/2.8 lens	70 mm Ektachrome MS SO-217 and 3401	None	None	None	Same as Gemini 4, with visual acuity experiment S008, Surface photography experiment D006	250 Earth-looking photographs taken	Carter and Stone (1974), Cortright (1968), Giddings (1975), Grimwood (1969), Kuehnel (1972), Lowman (1966a, 1968, 1972), Lowman and Tiedemann (1971), NASA (1966, 1967a, b, c), Pesce (1968), Short and Lowman (1973), Stevenson et al. (1968), Underwood (1968), Wobber (1968), Zeitler et al. (1971)
	Zeiss Contarex 35 mm camera/250 mm lens, 1,200 mm Questar lens	Anscochrome D-50 and D-200/ Haze filter						
	Maurer 16 mm motion picture camera with spotmeter	Ektachrome film						
Gemini 6-A (15–16 December 1965)	Hasselblad 500C 70 mm camera (NASA modified)/Zeiss planer 80 mm f/2.8 lens	70 mm Ektachrome MS SO-217 and type 2475/Haze filter	None	None	None	Same as Gemini 4	192 Earth-looking photographs taken	Carter and Stone (1974), Cortright (1968), Giddings (1975), Grimwood (1969), Kuehnel (1972), Lowman (1966a, 1968, 1972), Lowman and Tiedemann (1971), NASA (1966, 1967a, b, 1968d), Short and Lowman (1973), Underwood (1968), Wobber (1968), Zeitler et al. (1971)
	Maurer 16 mm motion picture camera with spotmeter	Ektachrome film						

Gemini 7 (15–16 December 1965)	Hasselblad 500C (NASA modified)/Zeiss planar 80 mm f/2.8 lens	70 mm Ektachrome MS SO-217 and Ektachrome infrared 8443	None	None	None	Experiment S005, S006, S008. Landmark contrast measurement experiment M412	429 Earth-looking photographs taken	Carter and Stone (1974), Cortright (1968), Giddings (1975), Grimwood (1969), Kuehnel (1972), Lowman (1966a, 1968, 1972), Lowman and Tiedemann (1971), NASA (1966, 1967a, b, 1968d), Pesce (1968), Short and Lowman (1973), Underwood (1968), Stevenson et al. (1968), Wobber (1968), Zeitler et al. (1971)
	Zeiss Contarex 35 mm camera/250 mm lens	Panatomic X 3400 and type 2475/ Haze filter						
	Maurer 16 mm motion picture camera with spotmeter	Ektachrome film						
Gemini 8 (16 March 1966)	Hasselblad 500C (NASA modified)/Zeiss planar 80 mm f/2.8 lens	70 mm Ektachrome MS SO-217/Haze filter	None	None	None	General Terrain and meteorology information photography	19 photographs taken, one of Earth's limb, six Earth-looking oblique photographs	Carter and Stone (1974), Cortright (1968), Grimwood (1969), Kuehnel (1972), NASA (1967a, 1968d), Short and Lowman (1973), Underwood (1968)
	Maurer 16 mm motion picture camera with spotmeter	Ektachrome film for Maurer cameras						
	Maurer 16 mm motion picture camera							
Gemini 9-A (6 June 1966)	Hasselblad 500C (NASA modified)/Zeiss planar 80 mm f/2.8 lens	70 mm Ektachrome MS SO-217/Haze filter	None	None	None	Same as Gemini 8	362 Earth-looking photographs taken	Cortright (1968), Giddings (1975), Grimwood (1969), Kuehnel (1972), Lowman (1968), Lowman and Tiedemann (1971), NASA (1967a, 1968d), Short and Lowman (1973), Stevenson et al. (1968), Underwood (1968), Zeitler et al. (1971)
	Hasselblad super wide angle-C (NASA modified)/Zeiss Biogon 38 mm, f/4.5 lens							
	Maurer 70 mm Space Camera/Xenotar 80 mm f/2.8 lens							
	Maurer 16 mm motion picture camera with spotmeter	Ektachrome film						

(*continued*)

Table 28.2 (continued)

Mission (dates flown)	Camera/lenses/other equipment	Films/Filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Gemini 10 (18–21 July 1966)	Hasselblad super wide angle/Zeiss Biogon 38 mm, f/4.5 lens Maurer 70 mm Space Camera/Xenotar 80 mm, f/2.8 lens Maurer 16 mm motion picture camera with spotmeter	Same as Gemini 9-A	None	None	None	Experiments S005, S006, M412. Color patch photography experiment M410	371 Earth-looking photographs taken	Carter and Stone (1974), Cortright (1968), Giddings (1975), Grimwood (1969), Kuehnel (1972), Lowman (1968), Lowman and Tiedemann (1971), NASA (1967a, b, 1968d), Short and Lowman (1973), Stevenson et al. (1968), Underwood (1968), Zeitler et al. (1971)
Gemini 11 (12–15 Sep 1966)	Hasselblad super wide angle/Zeiss biogon 38 mm, f/4.5 lens Maurer 70 mm space camera/Xenotar 80 mm, f/2.8 lens Maurer 16 mm motion picture camera with spotmeter	70 mm Ektachrome MS SO-368 film/Haze filter Ektachrome film	None	None	None	Experiments S005, S006	238 Earth-looking photographs taken	Carter and Stone (1974), Cortright (1968), Giddings (1975), Grimwood (1969), Kuehnel (1972), Lowman (1968), Lowman and Tiedemann (1971), NASA (1967a, 1968d), Short and Lowman (1973), Stevenson et al. (1968), Underwood (1968), Zeitler et al. (1971)
Gemini 12 (11–15 Nov 1966)	Same as Gemini 11	Same as Gemini 11	None	None	None	Experiments S005, S006 and ocean features photography	415 Earth-looking photographs taken	Carter and Stone (1974), Cortright (1968), Giddings (1975), Grimwood (1969), Kuehnel (1972), Lowman (1968), Lowman and Tiedemann (1971), NASA (1967a, b, 1968d), Short and Lowman (1973), Underwood (1968), Stevenson et al. (1968), Wobber (1968), Zeitler et al. (1971)

Table 28.3 Specifications of handheld earth observations equipment and activities during the Apollo program

Mission (dates flown)	Camera/lenses/other equipment	Films/filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Apollo 4 (9 Nov 1967)	Maurer Model 220 G 70 mm Sequence camera/ Ektar 76 mm f/2.8 lens (exposure setting 1/500 of a second)	70 mm Ektachrome MS So-368	None	None	None	Acquisition and return of highest altitude photography (~18,520 km) taken of Earth	712 photographs of Earth (2,136 Earth-looking photographs taken with the 70 mm camera during the Apollo 4, 6, 7, 8, 9, 10, 11 and 12 missions)	Dornbach (1968), Underwood (1968)
Apollo 6 (4 Apr 1968)	Maurer Model 220 G 70 mm sequence camera/ Ektar 76 mm f/2.8 lens	Same as Apollo 4	None	None	None	To obtain sequenced vertical photographs of Earth in stereo	Vertical or near-vertical photographs of Earth between ~183 and 306 km altitude. 372 Earth-looking photographs taken	Amsbury et al. (1969), Badgley et al. (1969), Dornbach (1968), Kaltenbach et al. (1969a), Lowman (1972), NASA (1968a, b, 1970b), Short and Lowman (1973)
Apollo 7 (11–22 Oct 1968)	Hasselblad model 500 C (NASA modified)/Zeiss planar 80 mm f/2.8 lens Maurer 16 mm sequence cam-Era/75 mm lens	70 mm Ektachrome MS SO-368, SO-121, and B&W type 3400/ Wratten 2A, 25A (red) and 58 (green) filters used with SO-121 film	Leitz monocular telescope (10X magnification, objective diameter 40 mm)	TV equipment, hand held or mounted. (Black & White)	Data tape recorder. Speed 9.5, 38, 305 cm/s, Tape 686 m	Synoptic terrain photography experiment S005, Synoptic weather photography experiment S006	Photographs taken of Earth between ~159 and 367 km altitude. 502 Earth-looking photographs taken. Hardware training for general photography and experiments given for each Apollo Mission	Amsbury et al. (1969), Carter and Stone (1974), Kaltenbach et al. (1969b), Kuehnel (1972), NASA (1968c, 1969a, 1970b, 1973a), Short and Lowman (1973)

(*continued*)

Table 28.3 (continued)

Mission (dates flown)	Camera/lenses/other equipment	Films/filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Apollo 9 (3–13 Mar 1969)	Hasselblad model 500EL (NASA modified)/Zeiss planar 80 mm f/2.8 lens	70 mm Ek- tachrome MS SO-368/Haze filter	Same as Apollo 7	Same as Apollo 7 (first color TV flown on Apollo 10)	Same as Apollo 7	Multispectral photography experiment SO65. Terrain, weather, air glow, and Earth's limb sites of opportunity were planned	127 vertical or near-vertical photographs taken of Earth in each of the four bands of the SO65 Experiment. 1,157 Earth-looking photographs taken with the handheld Hasselblad 70 mm camera. (A number of excellent 70 mm handheld color photographs were taken of the Earth during Earth orbit and enroute to and from the Moon on the Apollo 8, 10, 11, 12, 13, 14, 15, 16, and 17 missions)	Amsbury et al. (1969), Badgley et al. (1969), Bannert (1972), Carter and Stone (1974), Kaltenbach (1970), Kuehnel (1972), Lowman (1969a, 1972), NASA (1969b, c, 1970b, 1973a), Nicks (1970), Schowengerdt and Slater (1972), Short and Lowman (1973)
	SO65 Experiment: Hasselblad Model 500EL/ Zeiss planar 80 mm, f/2.8 lens/Photar 15 filter	70 mm Ektachrome IR, SO-180						
	Hasselblad Model 500EL/ Zeiss planar 80 mm f/2.8 lens/Photar 58B filter	70 mm B&W Panatomic X, type 3400						
	Hasselblad Model 500EL/ Zeiss planar 80 mm, f/2.8 lens/Photar 89B filter	70 mm B&W Infrared SO-246						
	Hasselblad Model 500EL/ Zeiss Planar 80 mm, f/2.8 lens/Photar 25A filter	70 mm B&W Panatomic X, type 3400						
	Maurer 16 mm data acquisition camera (DAC)/ 75 mm lens	Ektachrome SO-368 for Maurer DAC						

Table 28.4 Specifications of handheld Earth observations equipment and activities during the Skylab program

Mission (dates flown)	Camera/lenses/other equipment	Films/filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Skylab 2 (25 May to 22 June, 1973)	Multispectral photographic facility (MPF, s190): Multispectral photographic camera (S190A). Six Itek 70 mm boresighted cameras/ high precision f/2.8, 21.2 degree FOV lenses	MPF: 70 mm film: Station1-IR B & W (0.7–0.8 μm)	Leitz Trinovid 10 × 40 binoculars (space modified)	Westing house color 25–150 mm zoom. Focus range 20″ to infinity	Same as Apollo 7	Obtain data with the Earth resources experiment package (EREP) sensors. Obtain general information photography of Earth	35,000 frames cycled through S190A during Skylab 2, 3, and 4. (Not all of these were exposed or Earth-looking)	Kaltenbach et al. (1974), Kenney (1974, 1975), Lockwood and Sauer (1975), Lowman et al. (1973), Mcniel and Devalcourt (1974a), NASA (1973a, b, 1974b), Stevenson (1974), Underwood and Holland (1973a)
		Station 2-IR B & W (0.8–0.9 μm)			Video tape recorder, recording time 30 min			
		Station 3 Aerochrome IR color type EK2243 (0.5–0.88 μm)					32,994 Earth-looking photographs taken with S190A and 5,569 Earth-looking photographs taken with S190B during Skylab 2,3,and 4	
		Station 4- color SO-356 (0.4–0.7 μm)						
		Station 5- Panatomic-X B & W SO-022 (0.6–0.7 μm)					302 Earth-looking photographs taken with the handheld Hasselblad 70 mm camera, and 355 Earth-looking photographs taken with the handheld Nikon 35 mm cameras	
		Station 6- Panatomic-X B & W SO-022 (0.5–0.6 μm)						

(*continued*)

Table 28.4 (continued)

Mission (dates flown)	Camera/lenses/other equipment	Films/filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Skylab 2 (25 May to 22 June, 1973)	Actron Earth Terrain Camera (AETC, S190B)/Five inch format, f/4 lens, 18 inch focal length 16 mm DAC boresighted on Visual tracking system infrared spectrometer (S191) field of view	AETC: color SO-242 (0.4–0.7 μm) B & W EK3414 (0.5–0.7 μm) Ektachrome SO368					A number of Earth-looking TV passes were performed Note: The S190A, S190B and S191 are instruments in the Earth Resources Experiment Package (EREP) which obtained information of the Earth in numerous spectral regions during the three manned Skylab flights.	Kaltenbach et al. (1974), Kenney (1974, 1975), Lockwood and Sauer (1975), Lowman et al. (1973), Mcniel and Devalcourt (1974a), NASA (1973a, b, 1974b), Stevenson (1974), Underwood and Holland (1973a)
	Hasselblad Model 500EL (electric data camera system with Reseau plate), 70 mm camera/100 mm lens	70 mm Ektachrome SO-368						
	Nikon 35 mm Cameras (5)/ 55 mm and 300 mm lenses	Ektachrome SO-368, SO-168, 2485 and 2443 type						
	Maurer 16 mm DAC/5, 10, 18, 25, 75, & 100 mm lenses	Ektachrome SO-368 and SO-168						
Skylab 3 (28 July to 25 Sep 1973)	S190A, S190B, Hasselblad, Nikon, and DAC's same as Skylab 2	S190A, Hasselblad, Nikon, and DAC's film and filters same as Skylab 2	Same as Skylab 2	Same as Skylab 2	Same as Skylab 2	Obtain data with the Earth resources Experiment package (EREP) sensors.	642 Earth-looking photographs taken with the handheld Hasselblad 70 mm camera, and 537 Earth-looking photographs taken with the handheld Nikon 35 mm cameras.	Kaltenbach et al. (1974), Kenney (1974, 1975), Lockwood and Sauer (1975), Lowman et al. (1973), Mcniel and Devalcourt (1974b), NASA (1973a, b, 1974b), Stevenson (1974), Underwood and Holland (1973b)
						Obtain general information photography of Earth.		
	Earth Terrain Camera (S190B)	SO-242 (0.4–0.7 μm)						
		EK3414 (0.5–0.7 μm)					A number of Earth-looking TV passes were performed	
		Infrared color EK3443 (0.5–0.88 μm)						

Skylab 4 (16 Nov 1973 to 8 Feb 1974)	S190A, S190B, Hasselblad, Nikon, and DAC's same as Skylab 2 & 3	S190A, Hasselblad, Nikon, and DAC's film and filters same as Skylab 2 & 3	Leitz Trinovid 10 × 40 binoculars (Space modified)	Same as Skylab 2	Same as Skylab 2	Obtain data with the Earth resources Experiment package (EREP) sensors. Obtain general information photography of Earth	1,260 Earth-looking photographs taken with the handheld Hasselblad 70 mm camera, and 1,255 Earth-looking photographs taken with the handheld Nikon 35 mm cameras	Kaltenbach et al. (1974), Kenney et al. (1974, 1975), Lockwood and Sauer (1975), Lowman et al. (1973), NASA (1973a, b, 1974b), Stevenson (1974), Underwood and Holland (1974)
	Earth Terrain Camera (S190B)	SO-242 (0.4–0.7 μm)					A number of Earth-looking TV passes were performed	
		EK3414 (0.5–0.7 μm)						
		Infrared color (high resolution) EK3443 (0.5–0.88 μm)						

Table 28.5 Specifications of handheld Earth observations equipment and activities during the Apollo–Soyuz project

Mission (dates flown)	Camera/lenses/other equipment	Films/filters	Binoculars, minoculars, spotting scopes	Television	Tape recorders	Experiments	Comments	References
Apollo-Soyuz Test Project (15–24 Jul 1975)	Two 70 mm Hasselblad Model 500EL Cameras: One reflex camera with 50 mm and 250 mm lenses	SO-242/Wratten 2A; SO-368 (QX807)/film coated with equivalent of Wratten 2A	Leitz Trinovid 10 × 40 Binoculars (Space modified)	ASTP Television System (camera, lens, monitor)	Modified Sony Cassette-Coder Model TC-SS for voice recorder and playback	Earth Observations and photography Experiment MA-136	1,916 Earth-looking photographs taken; 751 good exposures obtained	NASA (1974a, 1977, 1976)
	One electric Data Camera System with Reseau plate with 60 mm and 100 mm lenses	Infrared color, Type 2443 equivalent/ Wratten 12		Operated in Apollo and Soyuz. Zoom 6–1 and 3–1.	Video tape recorder recording time 30 min. Data tape recorder same as Skylab 2		A number of Earth-looking TV passes were performed	
	One 35 mm Nikon camera/35 mm lens (interior photography), 300 mm lens (exterior photography)	Ektachrome		Range 25–150 mm, 9–27 mm. F-stop 4.4–44, 3.5–35				
	16 mm Data Acquisition Camera/ 10 mm, 25 mm, and 75 mm lenses	Ektachrome		Focus 51 cm-inf., 30.5 cm - inf. (modified Westinghouse)				
	Spotmeter							

conjunction with the JSC photo lab, Kodak, and camera hardware vendors conducted a variety of tests on various film and hardware combinations, with the end goal of producing the highest resolution images with the greatest dynamic range and the best color reproduction, and minimized image artifacts (e.g., film graininess or image vignetting). The tests were driven by factors that included market and environmental considerations (film production and streamlined film

Table 28.6 Specifications of handheld Earth observations film cameras and films used during the Space Shuttle and ISS programs

Cameras	Lenses	Film types
Hasselblad 500 EL/M, 70 mm, NASA modified	Zeiss 40 mm	Fuji, natural color positive, Velvia 50, CS-135-36, ASA 32, standard base
	Zeiss 50 mm	
	Zeiss 80 mm	
	Zeiss 100 mm	Kodak
	Zeiss 110 mm	B & W, Plus-X Aerographic
	Zeiss 250 mm	Color negative, Vericolor III, 70 mm unperforated, process C-41
	Zeiss 350 mm	
		Natural color positive, Ektachrome Professional 5017, ASA 64, standard base
		Natural color positive, Ektachrome 5036, 200 Professional, ASA 200, standard base
		Natural color positive, Ektachrome Professional SO-117, ASA 400, standard base
		Natural color positive, Lumiere 100/5046, ASA 100, standard base
		Natural color positive, Lumiere 100×/5048, ASA 100×, standard base
		Natural color positive, Ektachrome MS, ASA64, thin base, fine grain
		Color positive, Ektachrome X Professional, ASA 64, standard base
		Color positive, Ektachrome SO-368, fine grained, with yellow dye layer equivalent to Wratten 2A
		Color positive, Ektachrome, high speed, ASA 400
		Color positive, Ektachrome 64, 220 Roll format
		Color positive, Ektachrome 64
		Color positive, Aerochrome II color-reversal, ISO-A 32, , process EA-5, standard base
		Color positive, Elite 100 S, E6 reversal
		Color infrared, Aerochrome 2443, ASA 160, standard base
		Russian color infrared, 2 dye layer, estimated ASA 64

(*continued*)

Table 28.6 (continued)

Cameras	Lenses	Film types
Linhof Aero Technika, 100 × 120 mm, NASA modified	Linhof 90 mm	Kodak
	Linhof 250 mm	Natural color positive, Ektachrome QX 868, ASA 64, 5017 emulsion, thin base
		Natural color positive, Lumiere 100/5046, ASA 100, standard base
		Natural color positive, Lumiere 100×/5048, ASA 100×, standard base
		Color positive, Elite 100 S, E6 reversal
		Color positive, Ektachrome X Professional, ASA 64, standard base
		Color positive, Ektachrome 64, 220 roll format
		Color positive, Aerochrome II color-reversal, ISO-A 32, standard base
		Color positive, Aerochrome II Duplicating Film, 70 mm, process EA-5
		Color infrared, Aerochrome 2443, EA-5 process through June 1999, E-6 process afterward, thin base
Rolleiflex, 70 mm, NASA modified	Zeiss 50 mm	Kodak, natural color positive, Ektachrome Professional 5017, ASA 64, standard base
	Zeiss 80 mm	
	Zeiss 100 mm	
	Zeiss 250 mm	
Nikon F3 35 mm, NASA modified	All lenses are interchangeable and autofocus:	Fuji
		Natural color negative, NHG, ASA 400, standard base
		Color negative, 35 mm, ASA 800
Nikon F4 35 mm, NASA modified	16 mm	Natural color positive, Velvia 50, CS-135-36, ASA 32,standard base
	20 mm	Kodak
	28 mm	
	35 mm	B &W positive, Technical Pan Film 2415 Estar AH Base, ASA 100
	35–70 mm	
	zoom	Natural color negative, Ektar 100–3101, ASA 125, standard base
	55 mm	
	60 mm	Natural color negative, Ektapress 5030, ASA 1600, standard base
	70 mm	
	85 mm	Natural color negative, Vericolor III 5026, ASA 160, standard base
	180 mm	
	300 mm	Color negative, Kodacolor VRG/100, ASA 100, standard base
	400 mm	
		Color negative, Ektar 25 Professional Film, ASA 25
		Color negative, Pro 400, 35 mm, ASA 400

(*continued*)

Table 28.6 (continued)

Cameras	Lenses	Film types
Nikon F4 35 mm, NASA modified	2X doubler	Color negative, Pro PMZ 1000, 35 mm or 120 mm, process C-41
		Natural color positive, Ektachrome Professional 5017, ASA 64, standard base
		Natural color positive, Ektachrome Professional 5074, ASA 400, standard base
		Natural color positive, Ektar 25 Professional, ASA 25, standard base
		Color positive, Vericolor 400 Prof (VPH), ASA 400, standard base
		Color positive, Ektachrome X Professional, ASA 64, standard base
		Color positive, Ektachrome 64 T Professional Film, ASA 64
		Color positive, Elite 100 S, E6 reversal
		Color positive, EXR 500
		Portra
		Color negative, 160NC, 35 mm, ASA 160
		Color negative, 400NC, 35 mm, ASA 400
		Color positive, 400VC, 35 mm, ASA 400

Table 28.7 Digital SLR cameras used in the Space Shuttle and ISS programs (Nikon D2Xs, D3, D3X, and D3S cameras are currently in use onboard the ISS)

Manufacturer/model	Original image size (mm)	Original image size (pixels)
Sony HDW-700 high-definition television camcorder	–	1,920 × 1,035 interlaced
Kodak DCS460, RGBG array	27.6 × 18.5	3,060 × 2,036
Kodak DCS660, RGBG array	27.6 × 18.5	3,060 × 2,036
Kodak DCS760, RGBG array	27.6 × 18.5	3,060 × 2,036
Nikon D1, RGBG imager color filter	23.6 × 15.5	2,000 × 1,312
Nikon D2Xs, RGBG imager color filter	23.7 × 15.7	4,288 × 2,848
Nikon D3	36.0 × 23.9	4,256 × 2,832
Nikon D3X	35.9 × 24.0	6,048 × 4,032
Nikon D3S	36.0 × 23.9	4,256 × 2,832

development processes), image resolution results (for example, using faster ISO 100 films and faster lenses, enabling shorter exposure times), on-orbit operational flexibility (e.g., using auto exposures rather than calculating appropriate f-stops) and, ultimately, new digital technologies.

Both the Shuttle and ISS programs – and international partners – transitioned to exclusive use of digital 35 mm Single Lens Reflex (SLR) and video cameras in 2004 (Table 28.7). Coupled with improvements in operational uplink/downlink of

Table 28.8 Calculated ground resolutions for representative film and digital still cameras used on the International Space Station

		Ground resolution in m/pixel at image center	
		Minimum altitude = 368 km asl	Maximum altitude = 386 km asl
Camera	Lens (mm)		
Hasselblad 70 mm	110	35.4	37.1
	250	15.6	16.3
	350	11.1	11.6
Nikon 35 mm	300	13.0	13.6
	400	9.7	10.2
Kodak DSC	300	11.0	11.6
	400	8.3	8.7
	800	4.2	4.4

Altitudes are given as km above sea level (asl), and calculated resolutions do not include effective ground motion blurring caused by the high orbital velocity of the ISS (~27,500 km/h) relative to the Earth's rotation

data between ground stations and orbiting spacecraft, the use of digital cameras has both increased the volume of imagery and decreased the lag time between acquisition and public availability of data. A variety of interchangeable fixed focal length and zoom lens have also been (and continue to be) used with both the film and digital cameras, producing a wide range of ground resolutions in the handheld imagery dataset (Table 28.8).

More recently, improvements in camera technology, crew photographic technique, and improved lenses have enabled collection of images with center pixel resolutions approaching 2 m/pixel. Most handheld astronaut photography is also oblique – has a viewing angle relative to the earth surface less than 90° – to some degree which introduces varying ground resolution per pixel across a given image. For these reasons, it is difficult to assign a standard per pixel ground resolution to the astronaut photography dataset as is common with other orbital sensor datasets.

Original film negatives of handheld Earth observations imagery are archived in a cold-storage facility at the NASA Lyndon B. Johnson Space Center (JSC) in Houston, TX. Much of the historical film archive has been digitally scanned into standard formats (such as JPEG and TIFF) and is available online for review and download (see discussion below on online access to data). Many frames of Earth observation data, particularly for the early manned missions, have not been scanned in a systematic fashion however. Scanning of individual film frames of interest can be requested through the JSC Media Services Center.

Processing, Archiving, and Accessing Astronaut Photography

Digital camera imagery taken by astronauts is stored onboard the Space Shuttle or ISS and periodically transmitted to the ground using NASA or Russian downlink systems (both geosynchronous satellite network and line of sight). Handheld

camera imagery taken by NASA or partner agency astronauts, whether on physical film rolls or downlinked digital files, is initially processed by JSC archivists to determine the image subject category, e.g., Earth observation, with each image assigned a unique NASA identification number.

Imagery taken by cosmonauts is typically entered into the NASA system after receipt from the Rocket Space Corporation Energia, contractor to the Russian Space Agency. Once received on the ground and processed, the imagery is delivered to the CEO group – or historically, to a precursor group – for cataloging and entry into an online database, the Gateway to Astronaut Photography of Earth (http://eol.jsc.nasa.gov). If the imagery was downlinked in a raw camera format, it is converted into a full-resolution JPEG format for public access through the database in order to reduce server and network loads; the raw camera files (if available) can be requested through JSC Media Services.

The image cataloging process begins with determination of the image centerpoint coordinates. In contrast to data collected by current automated orbital sensors, there is no geolocation data embedded in astronaut photographs – recall that off-the-shelf cameras are used. The time of image acquisition, if accurate, can be used with the known orbital position of the spacecraft (e.g., the ISS) to determine a nadir-viewing ground coordinate useful as an initial estimate of the image's centerpoint coordinates. This estimate typically needs refinement however due to the unconstrained nature of handheld astronaut photography – all viewing angles, from essentially nadir to highly oblique, are available within the constraints of the spacecraft window, and the image centerpoint may be some distance from the spacecraft nadir point. This lack of standard viewing angles requires manual determination of the image centerpoint geographic coordinates by analysts using other georeferenced data – such as Landsat scenes or cartographic maps – with a typical location error of 0.1° in latitude and longitude.

Once the centerpoint is determined, descriptive metadata is generated for the image that includes major visible features or landmarks, estimated cloud cover, and the calculated viewing angle/direction relative to the orbital position of the spacecraft. Together with the camera file metadata for each frame (digital camera data only), the image is added to the Gateway to Astronaut Photography of Earth database. While all received ISS and Space Shuttle Earth observations imagery is currently entered into the database, it is prioritized for purposes of cataloging.

The entire digital collection of astronaut photography (more than one million images) is accessible using both map- and metadata-based search tools, and imagery can be downloaded at various resolutions free of charge. Keyhole Markup Language (kml) files for cataloged data also can be generated on the fly for direct input into geospatial browsers such as Google Earth. Digital images that have not been cataloged may also be queried using spacecraft nadir ground location coordinates, times of data acquisition, and lens focal length. Selected astronaut photographs and descriptive content are also available as part of the NASA content layer in the free geospatial browser Google Earth (http://earth.google.com).

Platforms

Manned spacecraft launched by the USA, Russia/former USSR, and other countries have uniformly employed asynchronous inclined elliptical orbits around the Earth. This is due to the relative ease of establishing and maintaining this type of orbit (e.g., less fuel needed to attain orbit and perform attitude adjustments), minimization of loss-of-signal periods for radio communication, and accessibility to landing sites on both land and at sea (Green and Lopez 2009). While this type of orbit limits the degree of nadir-viewing land and sea surface that can be observed to the degree of inclination, it also provides for a wider range of illumination and viewing conditions than typically available for sensors on sun-synchronous, polar-orbiting platforms. NASA retired the remaining Space Shuttles from service in 2011, leaving the International Space Station as the only platform for regular handheld Earth observations for at least the coming decade. While China has announced plans to complete construction of a crewed space station in low earth orbit by 2020, any plans to include collection of handheld Earth imagery as a formal component of science activities are unknown.

International Space Station (ISS)

Originally conceived as the US Space Station Freedom, the ISS orbital inclination was initially chosen at 28° to facilitate launching from NASA Kennedy Space Center in Florida. With expansion of the program to include international partners the inclination was increased to 51.6° to accommodate the Russian launch site at Baikonur Cosmodrome (Eppler and Runco 2001). This inclination allows the ISS to overfly the temperate and tropical regions of the Earth – covering approximately 75% of the Earth's land area and approximately 95% of the Earth's population (Fig. 28.1). The ISS orbit varies in altitude from approximately 350–455 km asl; due to atmospheric drag, reboosting of the ISS to maximum altitude is required approximately every 90 days. Due to the westward precession of orbit tracks, the ISS has an approximate repeat time over the same location every 3 days, with similar lighting conditions being repeated every 3 months not correcting for seasonal lighting shifts. For purposes of Earth observation, the CEO team generally limits daily target selections to regions with at least a 20° sun angle (elevation above local horizon) in order to have adequate illumination of ground targets. This constraint combined with ISS orbit precession and seasonal precession of the Earth's orbit around the Sun also produces intervals when only Northern or Southern Hemisphere ground targets meet the illumination criteria (Fig. 28.1). This constraint does not hold for atmospheric or nighttime ground targets.

The (ISS) Destiny Laboratory Window

The US Destiny Laboratory Module of the ISS has a window port built into its nadir facing side (NASA 2006). The window consists of three panes of Corning 7940 fused silica which are approximately 56 cm in diameter, providing an approximately 51 cm clear aperture. The ISS program agreed to upgrade the glass in the Destiny window to a set of stringent optical performance requirements in 1996.

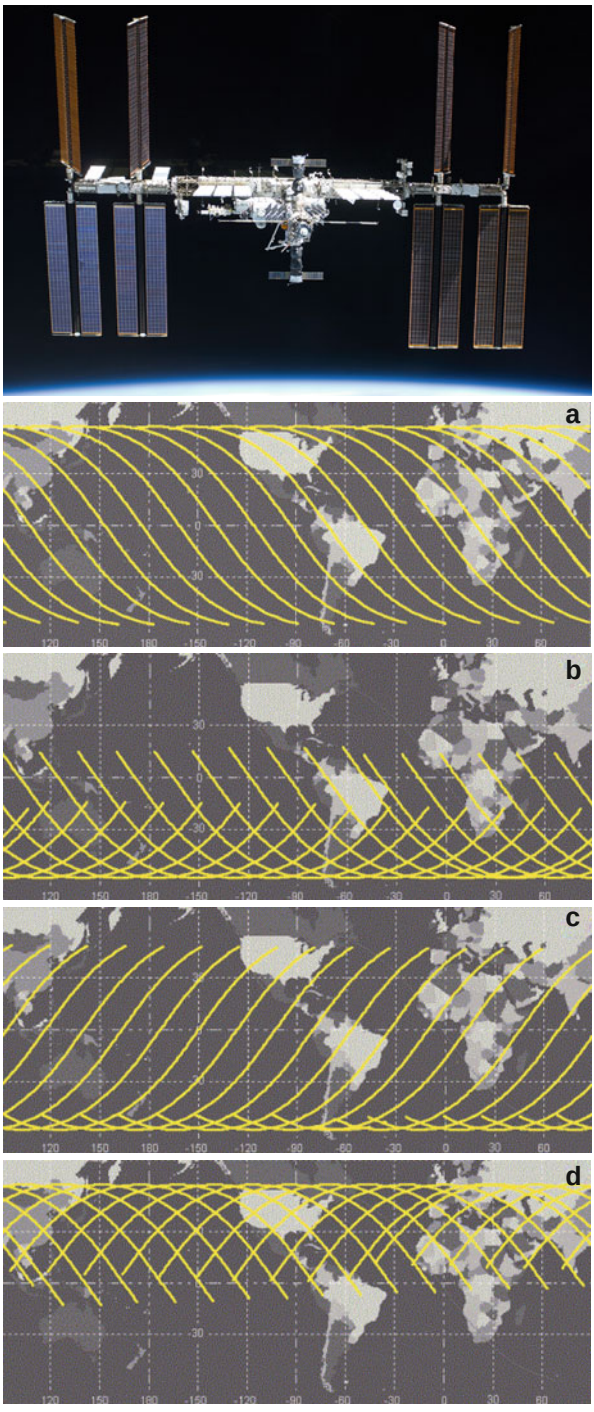


Fig. 28.1 International Space Station as viewed from Space Shuttle Atlantis during the STS-132 mission (Image S132-E-13221, taken 23 May 2010). Westward orbit ground track precession for the ISS at left. (**a**) Successive descending orbit tracks, daylight illumination in both North and South Hemispheres; (**b**) Daylight illumination in South Hemisphere only; (**c**) Ascending orbit tracks, daylight illumination in both North and South Hemispheres; (**d**) Daylight illumination in North Hemisphere only

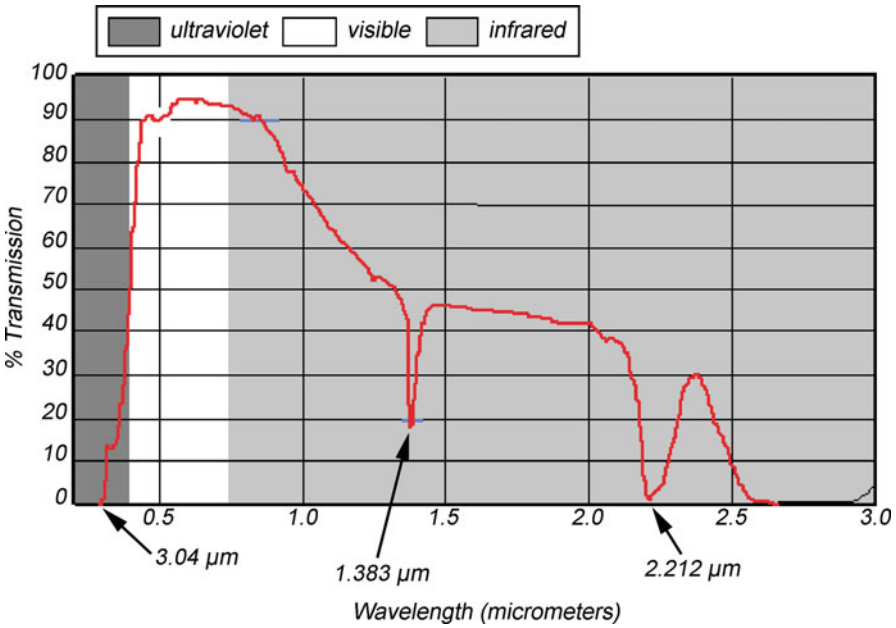


Fig. 28.2 Transmission curve for the center of the as-flown Destiny laboratory window. Cutoff below 0.304 μm is caused by a reflective coating

The Destiny window has a wavefront error of λ/15 peak-to-valley over a 15.2 cm aperture relative to a reference wavelength of 0.6328 μm, which allows the use of up to a 30 cm telescope with no degradation of wavefront due to the glass. These properties give the Destiny window the highest performance of any window ever flown on a crewed vehicle (Eppler and Runco 2001).

With the support of the ISS program, the flight article window was radiometrically calibrated prior to installation in the Destiny Laboratory in May of 2000. This calibration indicated that the window had better than 95% transmittance in the visible region, with a steep drop-off in the ultraviolet (due to a reflective coating) and a gradual drop-off into the infrared wavelengths (Fig. 28.2). Window transmittance decreases to 50% or less at approximately 1.3 μm.

In addition to the Destiny window, there are a number of other viewing ports on the ISS that are frequently used for Earth observations. Windows in the Russian Service Module (or Zvezda), albeit of lesser optical quality, provide both nadir and oblique viewing opportunities of the Earth. A moveable viewing module, the Cupola (Fig. 28.3), was transported to the ISS in 2010 during the STS-130 mission of Space Shuttle Endeavour and attached to the Tranquility Module. The Cupola includes seven optical-quality fused silica and borosilicate glass windows that provide 360° viewing capability, and was designed to support vehicle docking, remote manipulator arm operations, and Earth and space observations (European Space Agency 2011).

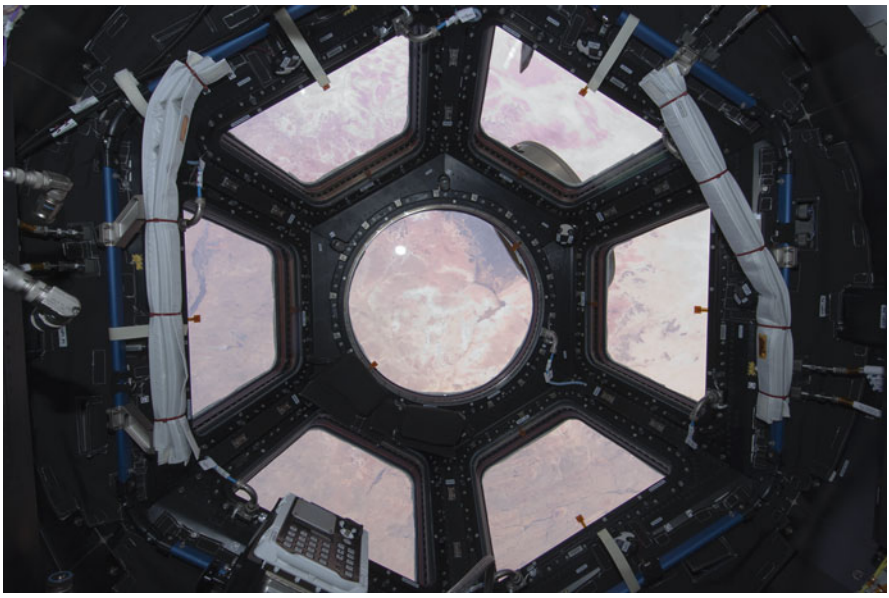


Fig. 28.3 Astronaut photograph ISS022-E-66972 of the Sahara Desert as seen through the cupola onboard the ISS (Image acquired 17 February 2010)

Window Observational Research Facility (WORF)

Complete utilization of the optical performance of the Destiny window would be impossible without a facility to allow stable positioning of research payloads in the window. Design and fabrication of the Window Observational Research Facility, or WORF, began in 1998. The WORF facility was transported to the ISS aboard the Space Shuttle Discovery in 2010 during the STS-131 mission.

The WORF is based on an ISS-standard Express rack (NASA 2006) to capitalize on the Express philosophy in accommodating sub-rack payloads. The WORF is essentially an Express rack with an approximate cubic meter-sized space in the middle (Fig. 28.4), centered on the Destiny window. This space – the payload volume – provides mounting surfaces for window payload hardware, including mounting on a stiff lower payload shelf that is designed to minimize transmission of ISS vehicle vibrations into the optical components of the payload. The interior of the WORF is sealed by means of an aisle-side hatch. The interior of the payload volume is painted flat black to minimize stray light and allow investigations of faint upper atmosphere phenomena such as aurora and noctilucent clouds.

The WORF can provide power, data, and cooling water for up to three payloads simultaneously by interfacing with existing ISS systems. At present, the WORF can provide an average downlink data rate on the order of 2 Mbps, although this may be improved with proposed communications infrastructure improvements to ISS. Investigators can operate their payloads autonomously at their institution, with up- and downlink data going through the Huntsville Operations Support Center at



Fig. 28.4 The Window Observational Research Facility (WORF) installed over the US Destiny Laboratory module window onboard the ISS. The Destiny window is visible through the WORF payload volume at image center (Image S131-E-8619 was taken 10 April 2010)

Marshall Space Flight Center in Huntsville, AL. The general design philosophy of WORF favors autonomous payloads, but crewmembers can operate payloads from the Destiny laboratory aisle using an externally mounted laptop computer.

It is an axiom in the US manned space program that a trained crewmember is one of the best analytical tools Earth observations can employ. The WORF is designed to accommodate crew stabilization devices and brackets to allow vibration-free operation of still cameras and video recorders. In addition, the aisle side hatch allows crew members to interface with a piece of equipment dubbed the "kayak shroud," which allows access to the interior of the WORF without glare from the Destiny Laboratory aisle interfering with Earth photography.

Remote Sensing Using Astronaut Photography

Advantages and Limitations

The most significant distinction of astronaut photography as a research dataset is the most obvious one – these images are framed and acquired by a human being rather than an automated sensor system. The astronaut can make on-the-fly decisions about image targets (e.g., pointing), resolution of data collected (through selection of cameras lenses), and whether conditions are favorable for taking

imagery (e.g., acceptable illumination and cloud cover) that are beyond the capabilities of automated sensor systems. A full orbit of the earth takes approximately 90 min, during which time the ISS crosses both illuminated and dark portions of the globe. The astronaut photograph dataset contains great variability in illumination conditions, look angle, spatial resolution (typically 4–40 m/pixel), and repeat imagery of a given location on the earth's surface.

The majority of astronaut photographs of the Earth have been acquired using film and digital cameras sensitive to the visible blue, green, and red wavelengths of the electromagnetic spectrum (0.38–0.72 μm), with specific experiments conducted using filters and film sensitive to narrow bandpasses within this range and in the near-infrared wavelengths (~0.72–1.30 μm; see ▶ Chap. 26, "Introduction and History of Space Remote Sensing" – for discussion of the historical development of remote sensing).

The use of off-the-shelf digital SLR cameras is somewhat limiting for quantitative spectral analysis of digital astronaut photography due to the generally broad bandpasses of the visible blue, green, and red channels, particularly in images taken with moderate to low illumination (Fig. 28.5). Without application of wavelength-limiting filters, these broad bandpasses make application of traditional spectral analysis and classification techniques difficult due to potential contamination of spectral information from adjacent channels. High- to medium-resolution digital astronaut photographs however contain significant spatial information content as recorded by pixel-to-pixel brightness variations, and this information can be capitalized on using sophisticated classification techniques such as object-based image analysis (OBIA).

Object-based image analysis uses image segmentation to identify homogeneous image objects at several different scales, rather than the classical pixel-based (and single-scale) classification approaches. Membership of pixels in a given image object is determined by rule-based analysis using fuzzy classification algorithms incorporating spectral character, shape, and neighborhood relationships across the class levels (Baatz et al. 2008; Blaschke et al. 2004). The application of OBIA to high-resolution digital astronaut photographs was explored for land cover classification of both urban and coastal ecosystems, and found to perform comparably to classifications derived from orbital multispectral data (Stefanov and Vande Castle 2006).

Astronaut photography presents challenges with regard to preprocessing and quantitative analysis, but these challenges are not insurmountable with currently available image-processing software. Variable look angles, acquisition times, and resolutions inherent to the astronaut photography dataset makes it a highly useful addition to more traditional sources of remotely sensed data, particularly for time-series analysis and change detection (Gebelein and Eppler 2006; Robinson et al. 2002; Stefanov et al. 2003). Most recently, astronaut photography from the ISS supported investigations of atmospheric phenomena such as aurora and polar mesospheric (or noctilucent) clouds; sea ice transport and plankton blooms; and snow cover during the International Polar Year of 2007–2008 (Evans et al. 2006). The following examples illustrate applications of astronaut photography to

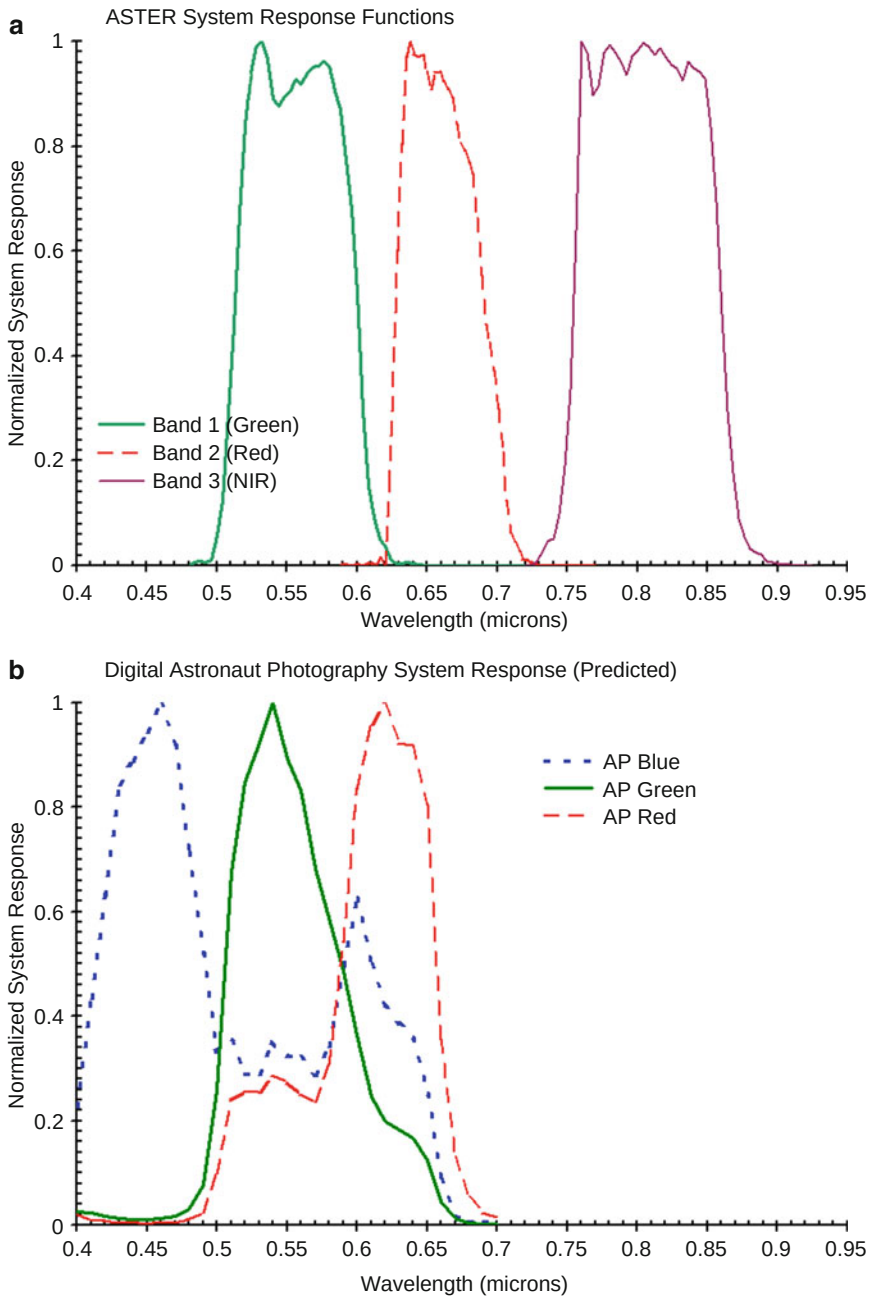


Fig. 28.5 Normalized system response curves for the visible *blue*, *green*, and *red* bands of the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) sensor (**a**) and predicted response for Kodak DCS 760 camera flown on the ISS (**b**). Note the overlap below 0.6 system response between *blue*, *green*, and *red curves* for the Kodak camera

geomorphic mapping, urban ecology, and volcano monitoring and response. These examples are intended to convey a sense of the potential of astronaut photography for Earth remote sensing, and are by no means all-inclusive.

Mapping of Megafans ("Inland Deltas")

Astronaut imagery has made an unexpected contribution to the study of world landscapes. It revealed the existence of numerous *megafans*, also known (inaccurately) as inland deltas, hundreds of km in radius, in many parts of the world, where only a few had been reported in local regional geological literature (Fig. 28.6). Because megafans are dramatic in their size – up to 650 km in radius, with a 200,000 km^2 area for the largest – there had been several claims in the literature for the "largest fan" on the planet. The global astronaut view suggested that a systematic global survey was needed, especially since claims for the largest fan *excluded* the largest that had actually been described.

By showing unknown examples of these features in various parts of the world, handheld imagery provoked a wider systematic search, which proved productive scientifically. Combined with other imagery, the global view revealed the significance of large fans. First, large fans can be claimed to be a significant landform because it is widespread (Hartley et al. 2010). Prior regional studies had shown the existence of a few of these large features. Low local numbers perhaps lulled geologists and geomorphologists into thinking these features were merely the insignificant tail end of the alluvial fan population (*alluvial fans* are fan-shaped cones of sediment that accumulate at the foot of mountains, being very common in the American Southwest, with small radii, usually <25 km [Blair and McPherson 1994]). Cataloging modern megafans worldwide showed their presence on all continents, with >160 very large fans (radii >100 km: Wilkinson et al. 2010) now known between 55°N and 55°S. Megafans can be so large that they constitute major features on many continental surfaces, especially when nested – there are an estimated 1.2 million km^2 in South America alone. The significance of this landform is suggested by the fact that almost every flat zone (with areas >100,000 km^2) on the continents is occupied by megafans (Wilkinson et al. 2011). We show an example from the roughness map of the eastern Sahara Desert and plains of central South America (Fig. 28.7).

Several aspects of this "find" in the astronaut imagery dataset are intriguing. Geologists and geomorphologists had assumed they knew the broad makeup of continental surfaces fairly well, especially in terms of major features, and especially in cloud-free more desertic zones where remotely sensed imagery had promoted familiarity with the landscapes. But geologists and geomorphologists were not *expecting* to see these features, and therefore did not see them – or at least did not accord them any particular significance. The variable perspective and ground resolutions of the astronaut imagery revealed an entire *set* of unforeseen landscapes and landscape relationships. It needed this unusual imagery to clarify what now seems a simple point, that rivers lay down vast quantities of material on continents, far from the oceans. Controls are an erosional zone, a river, and a neighboring basin to accomodate the river sediment. If the basin is wide the fans will be large.



Fig. 28.6 Okavango megafan (*outlined*), northwest Botswana, is one of the most visible and photogenic large fans, with its vegetated fingers radiating from an apex, where the Okavango River (a highly reflective section appears in sunglint above the apex in this northwest-looking view) spreads laterally (*arrows*). The radius of this fan is 140 km. This is a well-known visual cue to astronauts circling over southern Africa (NASA image STS043-151-32 was acquired on 8 August 1991)

Research based on the new understanding that megafan formation is a normal midscale component of river behavior on continental surfaces has been published in diverse fields. The megafan model is the basis for a new theory of Amazonian landscape development (Wilkinson et al. 2010), a set of new theories of fish speciation (Wilkinson et al. 2006), and the basis of a fluvial theory of evolution of the Martian surface where rover *Opportunity* has been performing its investigations (Wilkinson et al. 2011).

Future study of astronaut imagery is likely to yield other scientifically important finds, because it has the potential to show us new features and processes on the Earth's surface. Also, the human brain "behind the lens" can accomplish the critical function of selection, thereby reducing the amount of imagery to be processed. A new view, even of the multiscale oblique kind, can lead to a significant change in what we see. The megafan advance in understanding would not have been predicted.

Urban Geography and Ecology

Large urban areas are easily recognizable from orbit, and cities have been a frequent subject of astronaut photographs since the Gemini program. The rich historical record of city photographs acquired during the Apollo, Skylab, and Shuttle-Mir programs was used by Robinson et al. (2000) to compare urban growth

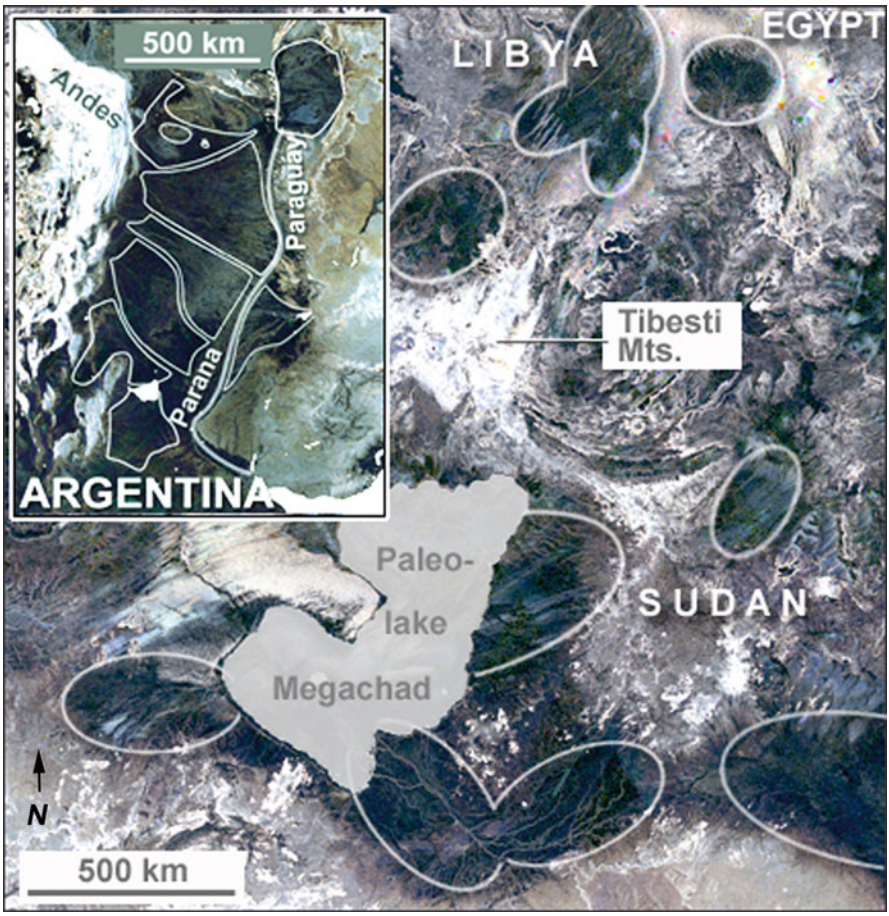


Fig. 28.7 *Darkest tones* in this roughness map of the Eastern Sahara Desert indicate areas of low slope, smooth topography covering very wide areas (note *scale bars*). All are surfaces formed by nested megafans (*outlined*). The inset figure is a similar map for the extensive plains of central Argentina and Paraguay provided for comparison. Rougher surfaces such as mountains and dune fields are light-toned

patterns and rates with population change for six North American metropolitan areas (Vancouver, Chicago, San Francisco, Dallas/Fort Worth, Las Vegas, and Mexico City). Photographs spanning a 28-year period were digitized, co-registered, and resampled to uniform pixel size to enable mapping of urban areal extent and measurement of change (Fig. 28.8). The results of the study provided insight into the variability of rates of urban land cover expansion relative to urban population growth in large North American metropolitan areas during 1969–1999. The first 2 years of this period predate publically available multispectral data from automated orbital sensors and highlight the temporal extent of the astronaut photography dataset.



Fig. 28.8 Change in Mexico City urban extent from 1969 (*inner gold polygon*) to 1996 (*outer red polygon*) – a 112.3% increase in built-up area. The Mexico City metropolitan area experienced less increase of built area relative to population growth over this time period compared to other North American cities examined by Robinson et al. (2000). Base image NM22-741-54B was acquired in December 1996

The transition from film to digital cameras during the first decade of the twenty-first century facilitated the application of digital image processing and analysis techniques to astronaut photography. A comparison of true-color (RGB) digital camera imagery of the Paris, France metropolitan area, obtained from the ISS, with visible and near-infrared data of similar spatial resolution from the Advanced Spaceborne Thermal Emission and Reflection (ASTER) sensor onboard the Terra satellite (acquired 2 weeks earlier) was performed by

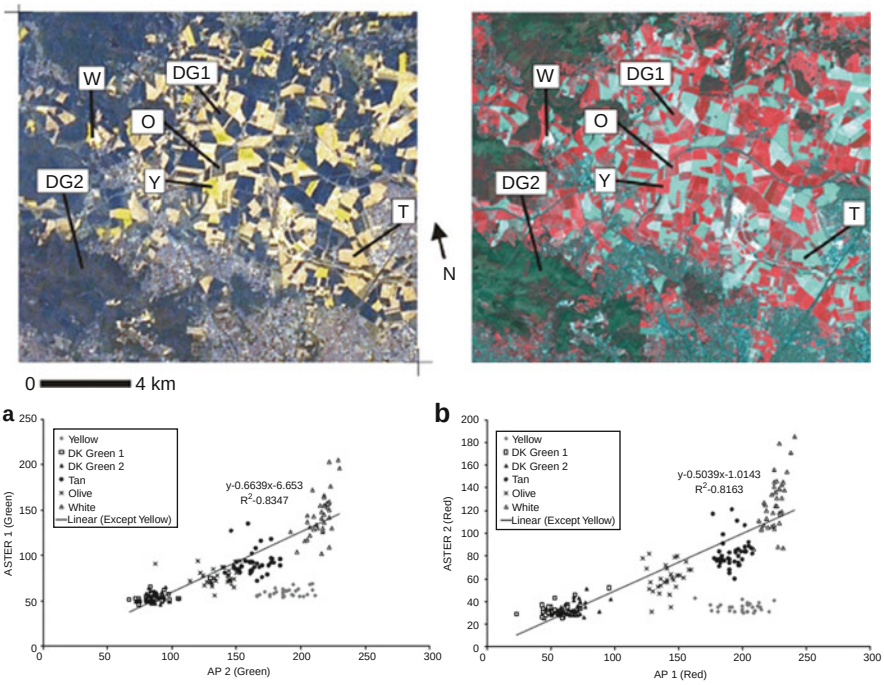


Fig. 28.9 Comparison of visible RGB digital astronaut photograph (AP; *left*) and visible to near-infrared (bands 4, 3, 2 as RGB) ASTER data (*right*) for agricultural fields in the northwestern Paris, France, metropolitan area. Y – *yellow*: vegetation, high productivity; DG1 – *dark green* 1: vegetation, moderate to low productivity; DG2 – non-canopied vegetation; T – *tan*: bare soil; O – *olive* 1: sparsely vegetated soil; W – *white*: light colored soils and built materials. (**a**) green band correlation; (**b**) red band correlation

Stefanov et al. (2003). The primary goal of this study was to identify the nature of yellow agricultural fields observed in the astronaut photograph that had no comparable signature in the ASTER data. Digital number values (DN, representing pixel brightness) recorded by both ASTER and the digital camera (Kodak 760 DCS) in the green and red bands were highly correlated for a number of vegetation and soil classes (Fig. 28.9). While the results of the study indicating that the data from both imagers could be reliably compared, the difference in coloration of the agricultural fields resulted from the flowering of rapeseed (*Brassica* sp.), a common commercial crop in the region. The digital astronaut photograph (image ISS004-E-10414, acquired on 24 April 2002) captured the phenological change of the rapeseed which had not occurred at the time of the ASTER overpass 2 weeks earlier. The study highlights the power of combining temporally variable astronaut photography with the more regularly acquired automated sensor datasets to obtain denser time series for urban ecological and agricultural monitoring purposes.

Volcanic Eruptions and Hazard Monitoring

The inclined equatorial orbit of the ISS allows observation of active volcanoes located within (or near to) approximately 52° North or South latitude at variable times, providing the potential to capture data on eruptive activity outside the repeat frequency of polar-orbiting satellite sensors. Frequent communication between ISS crews, ground controllers, and even the general public through Internet social networking sites also provides a rapid-response capability for potentially hazardous eruptions. This capability was dramatically demonstrated on 23 May 2006, when ISS Expedition 13 Flight Engineer Jeff Williams observed volcanic activity at Cleveland Volcano, located within the Aleutian chain of islands extending westward from Alaska (Fig. 28.10). Eruptions of Aleutian volcanoes can pose hazards to transcontinental airline flights because volcanic ash can disable jet engines. Cleveland Volcano is not heavily instrumented by the United States Geological Survey (USGS), and Williams was the first person to see (and report to the USGS) the eruptive activity from his vantage point in orbit.

The frequently serendipitous location of the ISS within observing distance of erupting volcanoes has also provided unique imagery of value to the volcanological community. The striking astronaut photograph of a pileus cloud forming during the large 2009 eruption of Sarychev Peak Volcano in the Kuril Islands chain is one of the best records of this unusual feature (Fig. 28.11). The pileus cloud was formed by rapid lifting of a moist air mass by the eruption column, and sequence photography of the eruption allows for analysis of the cloud formation and plume dynamics (Lockwood and Hazlett 2010; Venzke et al. 2009).

The value of high-resolution digital astronaut imagery of active volcanoes obtained from the ISS has been recently recognized by NASA, which has funded efforts to include photographs of volcanoes recognized as particularly hazardous, e.g., Mount Merapi in Indonesia, in the ASTER Volcano Archive (http://ava.jpl.nasa.gov). In addition to increasing astronaut situational awareness of active volcanoes and adding new imagery to the Archive, the potential value of fusing high-resolution astronaut photographs with low-resolution thermal infrared data from ASTER is being investigated in order to improve geologic hazard maps and response plans.

Public and Education Outreach Applications

It can be posited that NASA fundamentally changed humanity's perspective of the planet Earth, and our own importance within the cosmos, with an astronaut photograph. That photograph was taken in 1972 during the Apollo 17 mission as the astronauts looked back at the small "blue marble" of Earth, surrounded by the vastness of seemingly empty and dark space, on their way to the Moon. Some historians credit this image, and others like it, as helping to galvanize public opinion that led to important environmental legislation (such as the Clean Air and Water Acts) in the USA during the 1970s (Green and Jackson 2009). The public's fascination with looking at Earth from the vantage point of an astronaut on an orbiting platform continues today on the International Space Station (Fig. 28.12).



Fig. 28.10 Ash plume eruption from Cleveland Volcano on Chuginadak Island in the Aleutian Islands chain as seen from the International Space Station (The image (ISS013-E-24184) was acquired by the Expedition 13 crew on 23 May 2006)

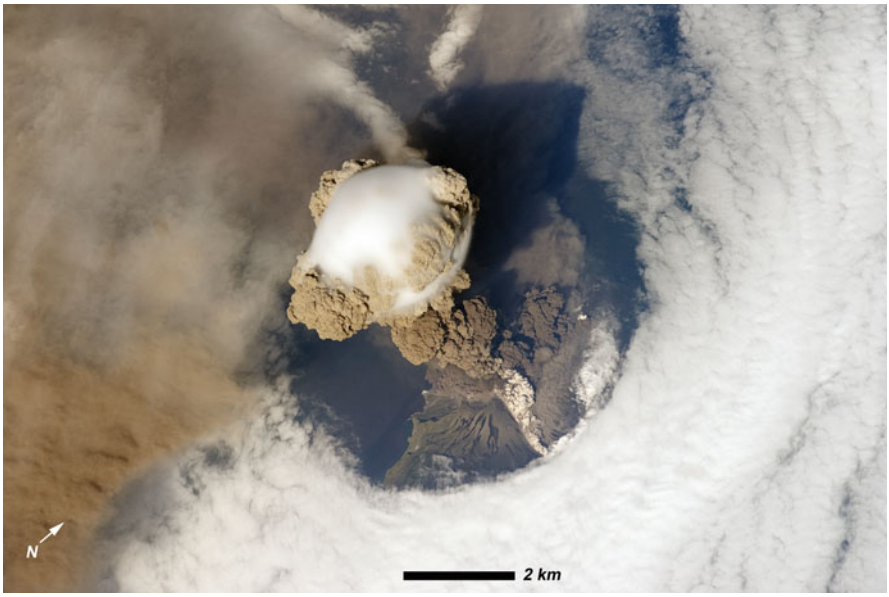


Fig. 28.11 Ash plume eruption, pyroclastic flows, and pileus cloud (image center) from Sarychev Peak Volcano, Matua Island in the Kuril Islands chain as seen from the International Space Station (The image (ISS020-E-9048) was acquired by the Expedition 20 crew on 12 June 2009)



Fig. 28.12 Apollo 17 "blue marble" view of the Earth showing the African and Antarctic continents (Image AS17-148-22727 was taken on 7 December 1972)

Public Outreach

Public access and distribution of handheld imagery of the Earth has evolved in step with the technological advances since the 1960s. Initially, the only way for the public to view astronaut imagery was to travel to the Earth Resources Observation and Science (EROS) Data Center in Sioux Falls, South Dakota, the Earth Data Analysis Center (EDAC) at the University of New Mexico in Albuquerque, or to the Media Service Center at NASA Johnson Space Center in Houston, Texas. With ready access to the Internet, today the global public can view and download astronaut imagery of the Earth through the Gateway to Astronaut Photography of Earth, or GAPE, web site (described above). Public dissemination of astronaut photography now includes social media outlets, including "tweeting" images directly from ISS via the Twitter web site (http://twitter.com/). Several astronauts have shared their unique views of the Earth by tweeting images along with commentary to the general public in near real time, providing a direct connection with humans working and living in space. These same images can also be accessed through the GAPE web site at various resolutions after working through the downlink and database entry procedure described above.

In recent years, the value of enlisting "citizen scientists" to participate in basic data analysis of large datasets has been proven by numerous online projects, such as Stardust@home and Galaxy Zoo (Wiggins and Crowston 2010). A public cataloging module is now in development for addition to the GAPE. This module will take advantage of the public's interest in astronaut images and ready access to geospatial

data (e.g., Google Earth) to provide an opportunity to participate in enhancing the metadata for the astronaut imagery of the Earth. The citizen scientist will be able to review the dataset of existing Earth imagery, select images, and add metadata such as the location of the center point of the image, cloud percentage, and identification of major features.

Education Outreach

Originally eight NASA Teacher Resources Centers (now called Educator Resource Centers – the number has expanded to 11) distributed slides, prints and 12-in. laser disks, of astronaut imagery on request to educators, museums, etc. Astronaut photographs of Earth have also appeared in textbooks, classroom lectures, and teacher workshops because they can easily be blended into existing classroom curricula without requiring sophisticated image processing or understanding of remote sensing – the visible wavelength, true-color imagery is intuitively interpreted by students. NASA Education Specialists from Oklahoma State University-Stillwater created a curriculum notebook based on astronaut handheld photography and satellite imagery to teach remote sensing as a tool for detecting Earth's features and changes. Themed slide sets of Earth phenomena such as weather/clouds and terrestrial impact craters from existing Shuttle imagery were made available to the public and to educators from the Lunar and Planetary Institute in Houston, Texas. In the mid-1990s, the Houston Museum of Natural Science integrated digital astronaut images into an interactive computerized display in the museum where students could direct virtual flyovers of anywhere in the world. As with public outreach, the education outreach applications of astronaut imagery have also evolved over time.

An initial effort to provide online resources for educators was the Earth from Space photograph collection, now incorporated into the GAPE web site. The content and format of the collection was informed by interactions with educators from all over the USA. Specifically, educators wanted searchable and themed imagery (e.g., cities, earth landscapes, hurricanes, and weather) with extensive captioning, including an indication of the north direction on the image. Earth from Space was originally populated with Space Shuttle imagery and is today updated with ISS imagery; there are now over one million Earth images in the collection.

Direct interaction between educators, NASA scientists, and students are facilitated through the accessibility of astronaut photography. The Crew Earth Observation (CEO) group and Teacher From Space Office, in cooperation with the NASA Aerospace Education Services Project-Montana/Nevada, Earth Observing System Education Project, University of Montana, and the Geospatial Research Group, Dept. of Geology, University of Montana developed an educational project called "Lewis and Clark's Travels: The Astronaut View" in 2003. The CEO office worked with the ISS Expedition 7 crew to take images along the 1804 path of Lewis and Clark to commemorate the 200th anniversary of their expedition. The ISS crew actively participated in this historical event, and often drew parallels between the

ISS mission and the Lewis and Clark Expedition. The images, highlighting locations along Lewis and Clark's westward trek to the Pacific coast were used by educators and historians and incorporated into museum displays, educational curriculum, and public web sites across the USA from Philadelphia and Ohio through Montana to Oregon.

Most recently, NASA has funded efforts to engage students in Science, Technology, Engineering and Math (STEM)-related authentic research that uses astronaut photography as the primary dataset. The Expedition Earth and Beyond (EEAB) Project is a student involvement program that allows teachers and their students in grades 5–14 to be actively involved in the process of science. Student-driven, authentic research projects are used to study the Earth and in some cases to be compared to other planetary bodies. Student teams are mentored by scientists at the NASA JSC and in select cases may request new imagery be acquired from the ISS. Information on the EEAB Project is available at http://ares.jsc.nasa.gov/ares/eeab/index.cfm.

Conclusion

Astronauts have taken images of the Earth using handheld cameras from the early days of space travel, and it is likely that they will continue to do so even with the development of ever more capable automated sensor systems. Indeed, the availability of regularly collected, standardized, and publically accessible remotely sensed data of Earth systems from orbital satellite platforms was enabled by early astronaut photography. Collection of handheld imagery of the Earth progressed from the essentially informal activities of USSR and USA astronauts in the 1960s through more formal programs and experiments during the 1970s and 1980s. Beginning in the late 1990s through today, collection of astronaut photography from long-duration orbital missions has enabled the use of this dataset for true time-series monitoring of selected targets as well as contributing to disaster response and management. The digital nature of current astronaut photography facilitates the application of standard image processing and analysis techniques, increases the volume of data acquired, and decreases the lag time between collection of data and availability to the public.

The human decision-making and response element inherent to astronaut photography continues to provide value and relevance within the current paradigm of Earth-observing automated sensors on satellite platforms. The International Space Station platform and the unconstrained pointing capability of handheld cameras allow for the collection of a dataset that is fundamentally different, yet fully complementary, to polar-orbiting sensor data due to the range of illumination conditions, viewing angles, and ground resolutions available. This enables the acquisition of several unique datasets, including high-resolution nighttime imagery of urban areas and detailed imagery of eruption plumes. The ready accessibility of digital astronaut imagery through online databases and collections fosters use by academic, government, and nonprofit organizations for a variety of purposes.

We expect that handheld astronaut photography will continue to integrate with, and complement, further enhancements of remote sensing capability on the International Space Station. As off-the-shelf digital camera technology improves, so do the opportunities for modification and use in orbit to collect new, unique datasets. Addition of high-definition video cameras to the ISS presents the opportunity to capture and target dynamic events such as wildfire plumes, hurricanes, volcanic eruptions, etc. The emerging human spaceflight programs of other nations may choose to develop their own handheld Earth observation programs; if so, the historical and current programs of the USA and USSR/Russia will provide useful models for data collection and dissemination.

References

D.L. Amsbury, Geological comparison of spacecraft and aircraft photographs of the Potrillo Mountains, New Mexico, and Franklin Mountains, Texas. in *Proceedings of the Sixth International Symposium on Remote Sensing of Environment*, vol. 2 (Environmental Research Institute of Michigan, Ann Arbor, 1969), pp. 493–515

D.L. Amsbury, United States manned observations of Earth before the space shuttle. Geocarto Int. **4**(1), 7–14 (1989)

Anonymous, Earth Photography from Gemini VI through XII: National Aeronautics and Space Administration Special Publication 142, 135 p (1968)

M. Baatz, C. Hoffmann, G. Willhauck, Progressing from object-based to object-oriented image analysis, in *Object-Based Image Analysis*, ed. by T. Blaschke, S. Lang, G.J. Hay (Springer, Berlin, 2008), pp. 29–32

P.C. Badgley, L. Miloy, L.F. Childs, *Oceans from Space* (Gulf, Houston, 1969)

D. Bannert, Plate Drift in the Afar and Issas Territory (French Somalia) and Eastern Ethiopia as Seen on Space Photography. NASA TN-D-6277 (1972), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19720020695_1972020695.pdf. Accessed 15 Nov 2011

T.C. Blair, J.G. McPherson, Alluvial fans and their natural distinction from rivers based on morphology, hydraulic processes, sedimentary processes, and facies assemblages. J. Sediment. Res. **A64**(3), 450–489 (1994)

T. Blaschke, C. Burnett, A. Pekkarinen, New contextual approaches using image segmentation for object-based classification, in *Remote Sensing Image Analysis: Including the Spatial Domain*, ed. by F. De Meer, S. de Jong (Kluwer, Dordrecht, 2004), pp. 211–236

L.D. Carter, R.O. Stone, Interpretation of orbital photographs. Photogrammetric Eng. **40**, 193–197 (1974)

J. Cimino, C. Elachi, M. Settle, SIR-B-The second shuttle imaging radar experiment. IEEE Trans. Geosci. Rem. Sens. **GE-24**(4), 445–452 (1986)

W.D. Compton, Where no man has gone before: a history of Apollo Lunar Exploration Missions. NASA History Series NASA SP-4214, (National Aeronautics and Space Administration, Washington, DC, 1989), 428 p

E.M. Cortright, *Exploring Space With a Camera*. NASA SP-168 (1968), http://history.nasa.gov/SP-168/sp168.htm. Accessed 15 Nov 2011

A.J. Derr, Photography equipment and techniques: a survey of NASA developments. NASA History Series NASA SP-5099, (National Aeronautics and Space Administration, Washington, DC, 1972), 25 p

J.E. Dornbach, Analysis of Apollo AS-501 Mission Earth Photography. NASA TM-X-58015 (1968), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19680026995_1968026995.pdf. Accessed 15 Nov 2011

C. Elachi, W.E. Brown, J.B. Cimino, T. Dixon, D.L. Evans, J.P. Ford, R.S. Saunders, C. Breed, H. Masursky, J.F. McCauley, G. Schaber, L. Dellwig, A. England, H. Macdonald, P. Martin-Kaye, F. Sabins, Shuttle imaging radar experiment. Science **218**(4576), 996–1003 (1982)

D. Eppler, S. Runco, Earth observations capabilities of the Window Observational Research Facility on board the International Space Station. in *Proceedings of the American Institute of Aeronautics and Astronautics Conference on International Space Station Utilization, contribution AIAA-2001-91401* (American Institute of Aeronautics and Astronautics, Reston, Virginia, 2001)

European Space Agency, Cupola (2011), http://www.esa.int/esaHS/ESA65K0VMOC_iss_0.html. Accessed 15 Nov 2011

D.L. Evans, J.J. Plaut, E.R. Stofan, Overview of the Spaceborne imaging radar-C/X-band synthetic aperture radar (SIR-C/X-SAR) missions. Rem. Sens. Environ. **59**(2), 135–140 (1997)

C.A. Evans, D.R. Pettit, S. Runco, G. Byrne, K. Willis, J. Heydorn, W.L. Stefanov, M.J. Wilkinson, M. Trenchard, *International Polar Year Observations from the International Space Station*. EOS Transactions of the American Geophysical Union, vol. **87**(52), Fall Meeting Supplement, Abstract IN41A-0878 (American Geophysical Union, Washington, DC, 2006)

C.A. Evans, M.J. Wilkinson, W.L. Stefanov, K. Willis, Training astronauts to observe the Earth from the Space Shuttle and International Space Station, ed. by W.B. Garry, J.E. Bleacher, in *Analogs for Planetary Exploration* (Geological Society of America Special Paper, 2011 in press)

G.C. Ewing, in *Oceanography from Space*, (Woods Hole Oceanographic Institution, Woods Hole, 1965). Ref. No. 65–10, 469 p

W.B. Foster, Earth photographs form Gemini III, IV, V: National Aeronautics and Space Administration Special Publication 129, 266 p (1967)

W.B. Foster, O. Smistad, Gemini Experiments Program Summary: National Aeronautics and Space Administration Special Publication 138, pp. 221–230 (1967)

J. Gebelein, D. Eppler, How Earth remote sensing from the International Space Station complements current satellite-based sensors. Int. J. Rem. Sens. **27**(13), 2613–2629 (2006)

L.E. Giddings, in *Index Maps for Gemini Earth Photography*. NASA JSC-09581, (National government publication, Houston, 1975)

J.R. Gill, Science Experiments Summary: National Aeronautics and Space Administration Special Publication 138, pp. 291–305 (1967)

N.F. Glazovskiy, L.V. Dessinov, Russian visual observations of earth: historical perspective, in *Dynamic Earth Environments: Remote Sensing Observations from Shuttle-Mir Missions*, ed. by K.P. Lulla, L.V. Dessinov, C.A. Evans, P.W. Dickerson, J.A. Robinson (Wiley, New York, 2000), pp. 15–24

K. Green, M.W. Jackson, Timeline of key developments in platforms and sensors for Earth observations, in *Earth Observing Platforms & Sensors, Manual of Remote Sensing*, ed. by M.W. Jackson, vol. 1.1, 3rd edn. (American Society for Photogrammetry and Remote Sensing, Bethesda, 2009), pp. 1–48

K. Green, C. Lopez, Basics of remote sensing systems, in *Earth Observing Platforms & Sensors, Manual of Remote Sensing*, ed. by M.W. Jackson, vol. 1.1, 3rd edn. (American Society for Photogrammetry and Remote Sensing, Bethesda, 2009), pp. 49–106

J.M. Grimwood, in *Project Mercury: A Chronology*. NASA SP-4001, (Washington, DC, 1963), http://history.nasa.gov/SP-4001/cover.htm. Accessed 15 Nov 2011

J.M. Grimwood, B.C. Hacker, P.J. Vorzimer, in *Project Gemini Technology and Operations – A Chronology*. NASA SP-4002, (Washington, DC, 1969), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19690027123_1969027123.pdf. Accessed 15 Nov 2011

A.J. Hartley, G.S. Weissmann, G.J. Nichols, G.L. Warwick, Large distributive fluvial systems: characteristics, distribution, and controls on development. J. Sediment. Res. **80**, 167–183 (2010)

M.R. Helfert, C.A. Wood, The NASA space shuttle earth observations office. Geocarto Int. **4**(1), 15–23 (1989)

E.M. Jones, K. Glover (eds.), in *Apollo Lunar Surface Journal* (NASA, Washington, DC, 2010), http://history.nasa.gov/alsj/. Accessed 15 Nov 2011

J.L. Kaltenbach, in *Science Report of the 70-millimeter Earth Photography of the Apollo 6 Mission*. NASA Technical Note S-217 (NASA Manned Spacecraft Center, Houston, Texas, 1969a)

J.L. Kaltenbach, in *Science Screening Report of the Apollo 7 Mission 70-millimeter Photography and NASA Earth Resources Aircraft Mission 981 Photography*. NASA TM-X-58029 (1969b), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19690021200_1969021200.pdf. Accessed 15 Nov 2011

J.L. Kaltenbach, in *Apollo 9 Multispectral Photographic Information: National Aeronautics and Space Administration, Technical Memorandum X-1957*, Washington, DC (1970), 34 p

J.L. Kaltenbach, W.B. Lenoir, M.C. McEwen, R.A. Weitenhagen, V.R. Wilmarth, in *Skylab 4 Visual Observations Project Report. NASA TM-X-58142*, Houston (1974), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19760009472_1976009472.pdf. Accessed 15 Nov 2011

G.P. Kenney, in *Skylab Program, Earth Resources Experiment Package Sensor Performance Report*, vol. 7 (S190B): SL2, SL3 and SL4 Evaluations. NASA MSC-05528 (NASA Johnson Space Center, Houston, Texas, 1974), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19750008509_1975008509.pdf. Accessed 15 Nov 2011

G.P. Kenney, in *Skylab Program, Earth Resources Experiment Package Sensor Performance Evaluation,* vol. 1 (S190A). NASA CR-144563 (NASA Johnson Space Center, Houston, Texas, 1975)

H.A. Kuehnel, in *Apollo Experience Report: Photographic Equipment and Operations During Manned Spaceflight Programs*. NASA TN D-6972 (NASA, Washington, DC, Houston, Texas, 1972)

J.P. Lockwood, R.W. Hazlett, *Volcanoes: Global Perspectives* (Wiley-Blackwell, Chichester, 2010)

H.E. Lockwood, G.E. Sauer, Processing corrections for Skylab photographic imagery. Photogramm. Eng. Rem. Sen. **41**(4), 523–532 (1975)

P.D. Lowman Jr., in *A Review of Photography of the Earth from Sounding Rockets and Satellites*. NASA TN D-1868, Washington, DC (1964)

P.D. Lowman Jr., Space photography – a review. Photogramm. Eng. **XXXI**, 76–86 (1965)

P.D. Lowman Jr., The Earth from Orbit: National Geographic Magazine, pp. 644–671 (1966a)

P.D. Lowman Jr., Photography from space-geological applications. Ann. N. Y. Acad. Sci. **140**, 99–106 (1966)

P.D. Lowman, Jr., in *Geologic Orbital Photography – Experience From the Gemini Program*. NASA X-644-68-228, Greenbelt (1968), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19680018143_1968018143.pdf. Accessed 15 Nov 2011

P.D. Lowman Jr., in *Apollo 9 Multispectral Photography: Geologic Analysis*. NASA GSFC-X-644-69-423 (NASA Goddard Space Flight Center, Greenbelt, Maryland, 1969a)

P.D. Lowman Jr., Geologic orbital photography: experience from the Gemini program. Photogrammetria **24**(3–4), 77–106 (1969)

P.D. Lowman Jr., *The Third Planet* (Weltflugbid, Zurich, 1972)

P.D. Lowman Jr., H.A. Tiedemann, in *Terrain Photography from Gemini Spacecraft Final Report*. NASA X-644-71-15, Greenbelt (1971), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19710008933_1971008933.pdf. Accessed 15 Nov 2011

P.D. Lowman Jr., J.A. McDivitt, E.H. White, in *Terrain Photography on the Gemini IV Mission – Preliminary Report*. NASA TN-D-3982, Washington, DC (1967), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19670017945_1967017945.pdf. Accessed 15 Nov 2011

P.D. Lowman Jr., H.V. Frey, W.E. Shenk, L. Dunkelman, in *Manual for 70 mm Hand-Held Photography From Skylab*. NASA GSFC X-644-73-147, Greenbelt (1973), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19730016748_1973016748.pdf. Accessed 15 Nov 2011

J.L. Mcniel, C.C. Devalcourt, in *Skylab-2 Handheld Photography Alphabetized Geographical Features List*. NASA CR-134245, Houston (1974a), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740014486_1974014486.pdf. Accessed 15 Nov 2011

J.L. Mcniel, C.C. Devalcourt, in *Skylab-3 Handheld Photography Alphabetized Geographical Features List*. NASA CR-140244, Houston (1974b), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740024691_1974024691.pdf. Accessed 15 Nov 2011

National Aeronautics and Space Administration, in *Gemini MidProgram Conference – Including Experiment Results*. NASA SP-121, Washington, DC (1966)

National Aeronautics and Space Administration, in *Earth Photographs from Gemini III, IV, and V*. NASA SP-129 (1967a), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19670017312_1967017312.pdf. Accessed 15 Nov 2011

National Aeronautics and Space Administration, in *Gemini Program Flight Summary Report: Gemini Missions I Through XII*. NASA MSC-GR-66-5 (Revision A) (1967b)

National Aeronautics and Space Administration, in *Gemini Summary Conference*. NASA SP-138, Washington, DC (1967c), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19680005472_1968005472.pdf. Accessed 15 Nov 2011

National Aeronautics and Space Administration, in *Apollo AS-502 Mission Data and Information List, 70 mm Color Photography*. NASA Mapping Sciences Branch, MSC (1968a), http://apollo.sese.asu.edu/SUPPORT_DATA/ap06_index.pdf. Accessed 15 Nov 2011

National Aeronautics and Space Administration, in *Apollo 6 Mission Report*. NASA MSC-PA-R-68-9 (1968b), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19800074520_1980074520.pdf. Accessed 15 Nov 2011

National Aeronautics and Space Administration, in *Apollo 7 Mission Report*. NASA MSC-PA-R-68-15 (1968c), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19760072144_1976072144.pdf. Accessed 15 Nov 2011

National Aeronautics and Space Administration, in *Earth Photographs From Gemini VI Through XII*. NASA SP-171, Washington, DC (1968d)

National Aeronautics and Space Administration, in *Apollo 7 Mission Data and Information List, 70 mm Color Photography*. NASA Mapping Sciences Branch, MSC (1969a), http://apollo.sese.asu.edu/SUPPORT_DATA/ap07_index.pdf. Accessed 15 Nov 2011

National Aeronautics and Space Administration, in *Apollo 9 Mission Report*. NASA MSC-PA-R-69-2 (1969b), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19760073064_1976073064.pdf. Accessed 15 Nov 2011

National Aeronautics and Space Administration, in *Apollo 9 Photographic Plotting and Indexing Report* (Including Aircraft Underflights). NASA Mapping Sciences Laboratory, MSC (1969c), http://earth.jsc.nasa.gov/sseop/metadata/Apollo9.htm. Accessed 15 Nov 2011

National Aeronautics and Space Administration, *Earth Resources Program Synopsis of Activity* (Manned Spacecraft Center, Houston, 1970)

National Aeronautics and Space Administration, in *Ecological Surveys From Space*. NASA SP-230 (1970b), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19700017671_1970017671.pdf. Accessed 15 Nov 2011

National Aeronautics and Space Administration, in *Handbook of Pilot Operational Equipment for Manned Space Flight*. NASA CD42-A/SL-997 (1973a), http://history.nasa.gov/alsj/JSC-07210PltOpsEquip.pdf. Accessed 15 Nov 2011

National Aeronautics and Space Administration, in *Skylab Program-EREP Investigator's Information Handbook*. NASA MSC-07874 (1973b)

National Aeronautics and Space Administration, in *Apollo-Soyuz Test Project Visual Observations Debriefing*. NASA JSC-09920 (1974a)

National Aeronautics and Space Administration, in *Skylab Earth Resources Data Catalog*. NASA JSC-09016 (1974b), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19750012726_1975012726.pdf. Accessed 15 Nov 2011

National Aeronautics and Space Administration, in *Apollo-Soyuz Test Project: Preliminary Science Report*. NASA TM-X-58173 (1976), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19760015986_1976015986.pdf. Accessed 15 Nov 2011

National Aeronautics and Space Administration, in *Skylab Explores the Earth*. NASA SP-380 (1977), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19820004619_1982004619.pdf. Accessed 15 Nov 2011

National Aeronautics and Space Administration, in *Reference Guide to the International Space Station*. NASA SP-2006-557 (2006), http://www.nasa.gov/pdf/508318main_ISS_ref_guide_nov2010.pdf. Accessed 15 Nov 2011

O.W. Nicks, in *This Island Earth*. NASA SP-250 (1970), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19710003091_1971003091.pdf. Accessed 15 Nov 2011

A. Pesce, *Gemini Space Photographs of Libya and Tibesti: A Geological and Geographical Analysis* (Petroleum Exploration Society of Libya, Tripoli, 1968)

R.A. Schowengerdt, P.N. Slater, in *Final Postflight Calibration Report on Apollo 9 Multiband Photography Experiment S065, Technical Memorandum 3* (Optical Sciences Center, University of Arizona, Tucson) (1972), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19720020819_ 1972020819.pdf. Accessed 15 Nov 2011

J.A. Robinson, B. McRay, K.P. Lulla, Twenty-eight years of urban growth in North America quantified by analysis of photographs from Apollo, Skylab, and Shuttle-Mir, in *Dynamic Earth Environments: Remote Sensing Observations From Shuttle-Mir Missions*, ed. by K.P. Lulla, L.V. Dessinov (Wiley, New York, 2000), pp. 25–41

J.A. Robinson, D.A. Liddle, C.A. Evans, D.L. Amsbury, Astronaut-acquired orbital photographs as digital data for remote sensing: Spatial resolution. Int. J. Rem. Sens. **23**(20), 4403–4438 (2002)

N.M. Short, P.D. Lowman Jr., in *Earth Observations From Space: Outlook for the Geological Sciences*. NASA X-650-73-316 (1973), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740004044_1974004044.pdf. Accessed 15 Nov 2011

M.J. Wilkinson, H.E. Newsom, C.C. Allen, D.Z. Oehler, M.R. Salvatore, J.B. DuBois, Fluvial architecture at different scales as possible analogs for the Ridge-Forming Unit, Sinus Meridiani, Mars, ed. by J.P. Grotzinger, R.E. Milliken, in *Proceedings of the First International Meeting on Stratigraphy and Sedimentology of Mars, Special Publication of the SEPM* (Society for Sedimentary Geology, 2011 in press)

W.L. Stefanov, Astronaut photography: hands-on remote sensing of the Earth. Phi Kappa Phi Forum **88**(1), 2–7 (2008)

W.L. Stefanov, J.A. Robinson, S.A. Spraggins, Vegetation measurements from digital astronaut photography. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. **34**(7/W9), 185–189 (2003)

W.L. Stefanov, J. Vande Castle, Ecological landscape classification using astronaut photography. Eos Transactions of the American Geophysical Union 87(52), Fall Meeting Supplement, Abstract B41A-0155 (2006)

R.E. Stevenson, Observations from Skylab of mesoscale turbulence in ocean currents. Nature **250**, 638–640 (1974)

R.E. Stevenson, R.M. Nelson, in *An Index of Ocean Features Photographed From Gemini Spacecraft*. Contribution 253, (Bureau of Commercial Fisheries Biological Laboratory, Galveston, 1968)

L.S. Swenson Jr., J.M. Grimwood, C.C. Alexander, in *This New Ocean: A History of Project Mercury*. NASA SP-4201 (1966), http://history.nasa.gov/SP-4201/toc.htm. Accessed 15 Nov 2011

R.W. Underwood, Color photography from space, in *Manual of Color Aerial Photography*, ed. by J.T. Smith Jr., A. Anson, 1st edn. (American Society of Photogrammetry, Falls Church, 1968), pp. 365–379

R.W. Underwood, J.W. Holland, in *Skylab 2: Photographic Index and Scene Identification*. NASA JL12-601 (1973a), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740007997_1974007997.pdf. Accessed 15 Nov 2011

R.W. Underwood, J.W. Holland, in *Skylab 3: Photographic Index and Scene Identification*. NASA JL12-602 (1973b), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740007948_1974007948.pdf. Accessed 15 Nov 2011

R.W. Underwood, J.W. Holland, in *Skylab 4: Photographic Index and Scene Identification*. NASA JL12-603 (1974), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19750019464_1975019464.pdf. Accessed 15 Nov 2011

E. Venzke, S.K. Sennert, R. Wunderman, Reports from the Smithsonian's global volcanism network, June 2009. Bull. Volc. **71**(10), 1211–1212 (2009)

A. Wiggins, K. Crowston, Developing a conceptual model of virtual organisations for citizen science. Int. J. Organ. Des. Eng. **1**(1–2), 148–162 (2010)

D.E. Wilhelms, *To a Rocky Moon: A Geologist's History of Lunar Exploration* (University of Arizona Press, Tucson, 1993)

M.J. Wilkinson, L.G. Marshall, J.G. Lundberg, River behavior on megafans and potential influences on diversification and distribution of aquatic organisms. J. S. Am. Earth Sci. **21**, 151–172 (2006)

M.J. Wilkinson, L.G. Marshall, J.G. Lundberg, M.H. Kreslavsky, Megafan environments in northern South America and their impact on Amazon Neogene ecosystems, in *Amazonia, Landscape and Species Evolution*, ed. by C. Hoorn, F.P. Wesselingh (Blackwell, Chichester, 2010), pp. 162–184

D.M. Winker, R.H. Couch, M.P. McCormick, An overview of LITE: NASA's LIDAR in-space technology experiment. Proc. IEEE **84**(2), 164–180 (1996)

F.J. Wobber, Orbital photography: applied Earth survey tool. Photogr. Appl. Sci. Technol **2**(7), 21–29 (1968). 56

D. Woods, in *Apollo Flight Journal* (2009), http://history.nasa.gov/afj/. Accessed 15 Nov 2011

E.O. Zeitler, T.G. Rogers, in *The Gemini Program Physical Sciences Experiments Summary*. NASA TM X-58075 (1971), http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19720003207_1972003207.pdf. Accessed 15 Nov 2011

Electro-optical and Hyper-spectral Remote Sensing

29

Scott Madry and Joseph N. Pelton

Contents

Introduction 730
Introduction to Hyperspectral Imaging 731
Early History of Hyperspectral Sensing 732
Electro-optical Sensing 736
Notes 738

Abstract

Remote sensing satellites have become increasingly sophisticated in terms of increased spatial, radiometric, and temporal resolution. Over the past few decades sensing devices have become more sophisticated with not only higher spatial resolution but have also now become more capable at capturing data in much more precisely defined bandwidths or frequency ranges. This provides the ability to identify particular vegetation, forestry, wildlife and fish, minerals – even camouflage – with greater precision.

This evolution of sensor capabilities has, however, led to new needs on the ground in terms of interpreting the data. The new interpretative needs – because much more data is captured – involve requirements for new and faster processing techniques on the ground. Or it has led to the need for "preprocessing of data" (i.e., discarding noncritical or nonmeaningful data) before being down loaded from the satellite. The point is that the more capable sensors that collect a larger amount of data serves to alter the way the torrent of data downloaded from the

S. Madry (✉)
International Space University, Chapel Hill, North Carolina, USA
e-mail: Scottmadry@mindspring.com

J.N. Pelton
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: peltonjoe@gmail.com

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_42, © Springer Science+Business Media New York 2013

sky is processed. This chapter explains the transition that is rapidly occurring in terms of the transition from multispectral imaging to the much more precise and data intensive hyperspectral sensing – also called imaging spectroscopy.

Much more capable electro-optical arrays (usually using charge coupled devices (CCDs) allow the capturing of hyperspectral data much more efficiently. In the past with multispectral sensing data was collected in perhaps five or perhaps as many as ten broad frequency bands. Now data can be collected in much more precise and narrower frequency bands in the infrared, near infrared, visible spectrum, and even ultraviolet bands.

This chapter discusses the transition from multispectral to hyperspectral sensing that is now in full swing. It notes that the first uses of hyperspectral sensing were for military and defense-related purposes but now hyperspectral sensing – using the latest electro-optical arrays – is becoming central to civil earth observation programs. This transition has not only meant a change in the imaging process and the types of sensor devices included on remote sensing satellites, but it has also signaled the shift in data processing formats with data being processed as "data cubes." In this format spatial data is provided along the (X,Y axis) while the various frequency bands are displayed on the vertical or Z axis.

Keywords

Charge coupled devices (CCD) • Data cubes • Data sets • Electro-optical sensors • Hyperspectral sensing • Imaging spectroscopy • Multispectral sensing • Post-processing • Preprocessing • Radiometric resolution • Signal-to-noise performance levels • Spatial resolution • Spectral band • Spectral resolution • Spectrometer

Introduction

Recent advances in remote sensing sensor devices and improved data processing techniques have led the way for the development of two key capabilities in the field. These newer capabilities are: (1) much more capable electro-optical sensors (typically using charge coupled device (CCD) array technology) that can collect data in precise and narrow frequency bands with great precision and (2) hyperspectral sensing, or imaging spectroscopy that is aimed at collecting data in hundreds of different spectra that ranges from infrared, to near infrared to the visible light band and up to ultraviolet light.

Initially hyperspectral sensing was used for military purposes such as to distinguish camouflage from real forestry or vegetation or the detection and identification of particular ore, minerals, or vegetation. Electro-optical sensing, which benefits from technology developed for modern digital cameras and other types of ground-based technology, has also developed rapidly in the past few years and is also being used for an expanding number of purposes as well.

Imaging spectroscopy has been used in the laboratory by physicists and chemists many years for identifying the composition of materials. Spectroscopy can be used

to detect individual absorption features due to specific chemical bonds in a solid, liquid, or gas and thus it is a highly versatile tool. The idea to employ hyperspectral techniques in satellite remote sensing began in the mid-1980s and was initially used by geologists for locating and identifying minerals and rock structures that might help identify petroleum as well as valuable ores.

Hyperspectral remote sensing combines imaging and spectroscopy into a single system which often includes large data sets and requires new processing methods. Hyperspectral data sets are generally composed of as many as 150 to over 300 spectral bands of relatively narrow bandwidths (5–10 nm), whereas, multispectral data sets are usually composed of about 5–10 bands of relatively large bandwidths (70–400 nm). Thus, as noted above, hyperspectral data sets require much more intensive and demanding data processing. This is further complicated by the fact that detection of mineral or ores is dependent on many different factors in combination. These factors include spectral resolution and coverage, the signal-to-noise performance (or net effective gain) of the spectrometer, the abundance of the material and the strength of absorption features for that material in the wavelength region measured. When these factors are taken all together the data processing requirements can be orders of magnitude larger than is the case with multispectral remote sensing.

Electro-optical sensors depending on capturing photons from the imaged area and converting them into an "electronic image" that "mirrors" the imaged area. This type of imaging is most frequently now achieved using a charge coupled device (CCD) or a larger CCD array. A CCD is an integrated circuit etched onto a silicon surface so as to form light sensitive elements represented by what are called pixels. Photons incident on this surface generate a charge that can be read by electronics and turned into a digital copy of the light patterns falling on the device. CCDs come in a wide variety of sizes and types and are used in many applications from remote sensing satellites to digital cameras. A very high resolution CCD array – used for hyperspectral imaging – can produce a huge amount of data for processing. In some cases this can lead to a form of "preprocessing" that can serve to eliminate or reduce data that is considered not to be of particularly importance so that the data downloaded to remote sensing data centers is scaled down to levels so that the incoming data does not overload the processing center's capacity.

In the sections that followed the technology and application for both hyperspectral sensing and electro-optical imaging will be presented in greater detail.

Introduction to Hyperspectral Imaging

In the early days of remote sensing satellites the initial approach ways to undertake what is called multispectral sensing. This was to have sensors that collected images across a wide range of spectrum that ranged from infrared through the visible light and up to the ultraviolet frequencies. This imaging typically was across 5–10 broad bands that might be as wide as 400 nm. This produced a significant amount of data and important results. The infrared imaging could help map human and animal activity, identify heat pollution in the oceans, disease in vegetation and forests, and

meteorological applications. The visible light imaging was useful for a wide range of applications from weather tracking to a wide range of industrial, agricultural, and environmental applications, while ultraviolet sensing complemented the images captured at in the lower frequency ranges.

Hyperspectral imagery, however, represented a whole new range of capabilities that were much more targeted and precise. The much more narrow and targeted bandwidths associated with this type of imaging allows specific minerals or ores to be identified. With multispectral sensing one could distinguish trees from meadows and lakes and ponds. With hyperspectral sensing one can distinguish lead ore from copper ore or pitchblende from quartz or in defense-related applications one can identify specific sites of a strategic nature. The refinement in terms of much more narrowly targeted frequencies, greater spatial resolution, high gain spectrometers, and other enhancements, however, means that hyperspectral sensing gathers much greater amounts of data.

This acquisition of much greater amounts of data ultimately requires a much more sophisticated means to process the collected information within a reasonable amount of time or the value of the data tends to be depreciated. In order to process this data and display it efficiently it is typically collected (and represented) as a data cube with spatial information collected in the X-Y horizontal "spatial" plane, and spectral information is represented in the Z-direction. This works quite well in flat areas to "see" what the hyperspectral sensors are "seeing" and it can also allow for "preprocessing" in terms of discarding data from areas that are not considered of particular interest or from areas that have previously been successfully imaged.[1]

In recent years, hyperspectral sensors have become increasingly sophisticated in terms of being perfected to achieve higher and higher performance characteristics. Thus hyperspectral sensors have increased in the following regards:

- Increased spatial resolution
- Greater spectral coverage (within narrowed bandwidth slots across a wider spectrum)
- Higher net gain (signal-to-noise) by the sensing spectrometers
- Greater material absorption levels

All of these factors combine to make the "data cube" that is downloaded from remote sensing satellites designed to carry out hyperspectral imaging much more "dense" or to contain a much greater amount of data.

Early History of Hyperspectral Sensing

In the early 1980s the NASA Jet Propulsion Laboratory in California developed what was called the Airborne Visible-Infra Red Imaging Spectrometer (AVIRIS) and flew it on a test aircraft. Before AVIRIS, technological limitations prevented spectrometers from being used on moving platforms and produce an accurate hyperspectral image that could be processed and displayed with reliable results. In time this allowed the US Air Force to adapt this technology to be used on a test remote sensing satellite. On August 1, 2000, shortly after its July 9, 2000, launch the Fourier Transform Hyperspectral sensor aboard the US Air Force MightySat II.1

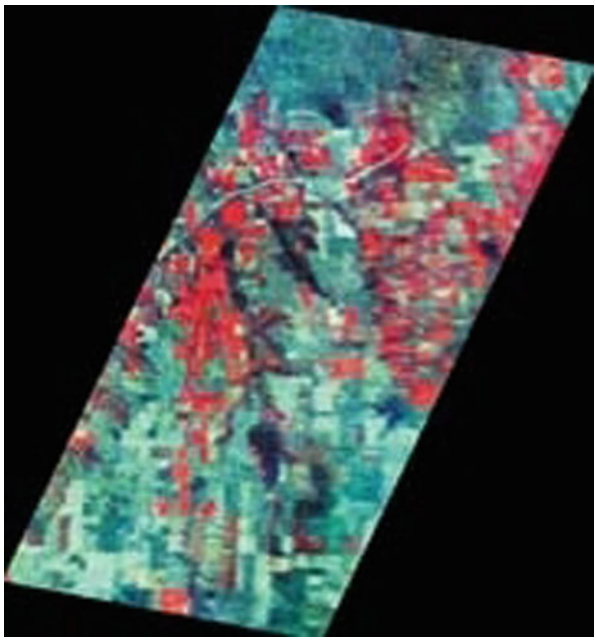


Fig. 29.1 Initial hyperspectral image from space captured in year 2000 (Image courtesy of US Air Force)

sent back its first "hypercube" dataset from orbit. This began a new phase in remote sensing from space. This first image showing a portion of rural Colorado is shown in Fig. 29.1 above.[2]

From this image, data processors on Earth could determine highway payment, ponds and lakes, and various types of vegetation and forestry. In terms of defense-related applications it was able to distinguish materials that might be used to camouflage military bases, airfields, or fields used to grow drugs. The success of this test of a hyperspectral sensor has led many other parallel applications in civil space programs.

The European Space Agency developed the Proba-1 remote sensing satellite with an on board hyperspectral spectrometer known as CHRIS (see Fig. 29.2). This sensor, launched in 2001, was designed to monitor 62 bands but had a spatial resolution of only 17 m. This was still sufficient to produce highly effective images. Today data from this spacecraft supports many practical and scientific applications, such as land surface, coastal zone, and aerosol monitoring and its imaging is actually being now used by 300 different scientific units in some 50 countries. One of the images from Proba-1 below shows with some clarity what is called the French Frigate Shoals. These shoals are an atoll about 800 km northwest of Hawaii (see Fig. 29.3).

NASA at the same time developed and launched the Hyperion hyperspectral sensor on its Earth Observing-1 satellite. With the success of these two hyperspectral sensors a number of new projects to develop much higher capabilities in terms of bands and resolution were undertaken around the world.

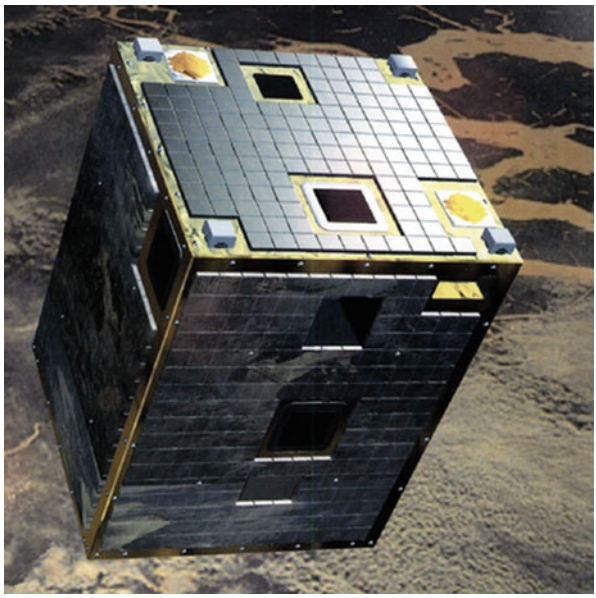


Fig. 29.2 Probia-1 low earth orbit satellite by ESA with CHRIS spectrometer on board (Image courtesy of European Space Agency)

Thus after decades of research and development hyperspectral imaging is becoming increasingly mature. Today one can say that the transition from multispectral sensing to hyperspectral sensing is happening fairly rapidly. This relatively new technology is now greatly improving our ability to characterize the state of Earth, monitor climate change and pollution, and locate both resources and problems around the world. In short it can probably be said with confidence that hyperspectral sensing has entered the mainstream of satellite remote sensing.[3]

Several new capabilities in the field are anticipated. One of these initiatives is the PRISMA satellite of Italian Space Agency (ASI). This is an earth observation system which will employ innovative electro-optical instrumentation for the hyperspectral imaging. The PRISMA mission combines a hyperspectral sensor with a panchromatic, medium-resolution camera. The electro-optical capability is important both in terms of sensing from space as well as processing on the ground.

The advantages of the combination of the hyperspectral electro-optical sensor and panchromatic camera are that in addition to the classical capability of observation based on the recognition of the geometrical characteristics of the scene (i.e., the panchromatic camera), there is the contrasting scene that is offered by hyperspectral sensors. The hyperpectral sensor can determine the chemical-physical composition of anything being viewed and do so regardless of whether it is solid, liquid, or gaseous.

The combined panchromatic and hyperspectral image offers the scientific community and users many applications in the field of environmental monitoring,



Fig. 29.3 Spectrometer image from Proba-1 satellite developed by ESA (Image courtesy of European Space Agency)

resource management, crop classification, pollution control, and mineral prospecting. Prisma is currently anticipated to launch as of the end of 2013.[4]

Closely following the PRISMA smaller-scale demonstration project, will be the full-scale German designed EnMAP hyperspectral satellite in 2014. This satellite will seek to map the entire surface of the world and do so in 200 narrowband frequency channels using the latest electro-optical techniques.

The primary goal of the German EnMAP satellite as designed by the German space agency DLR has been formally stated by EnMAP Project Scientist Prof. Herrmann Kaufmann to be "to offer accurate, diagnostic information on the state and evolution of terrestrial ecosystems on a timely and frequent basis, and to allow for a detailed analysis of surface parameters with regard to the characterization of vegetation canopies, rock/soil targets, and coastal waters on a global scale."

EnMAP has been designed to record biophysical, biochemical, and geochemical variables that will be displayed in hyperspectral data cubes such as shown below in Fig. 29.4. It is hoped that this unique ability to display of the hyperspectral data in such a manner will serve to increase our understanding of the world's chemical and biological make up as never before possible.

Further in 2015, NASA plans to launch the so-called HyspIRI mission. This satellite will be able to obtain images in 210 separate spectral bands. The focus of HyspIRI will be to study volcanoes and volcanic eruption, the water status, and nutrients associated with various types of vegetation, deforestation, and new scientific information that might be able to help provide early warning of droughts. This satellite will supplement the findings from EnMap and other hyperspectral satellites.

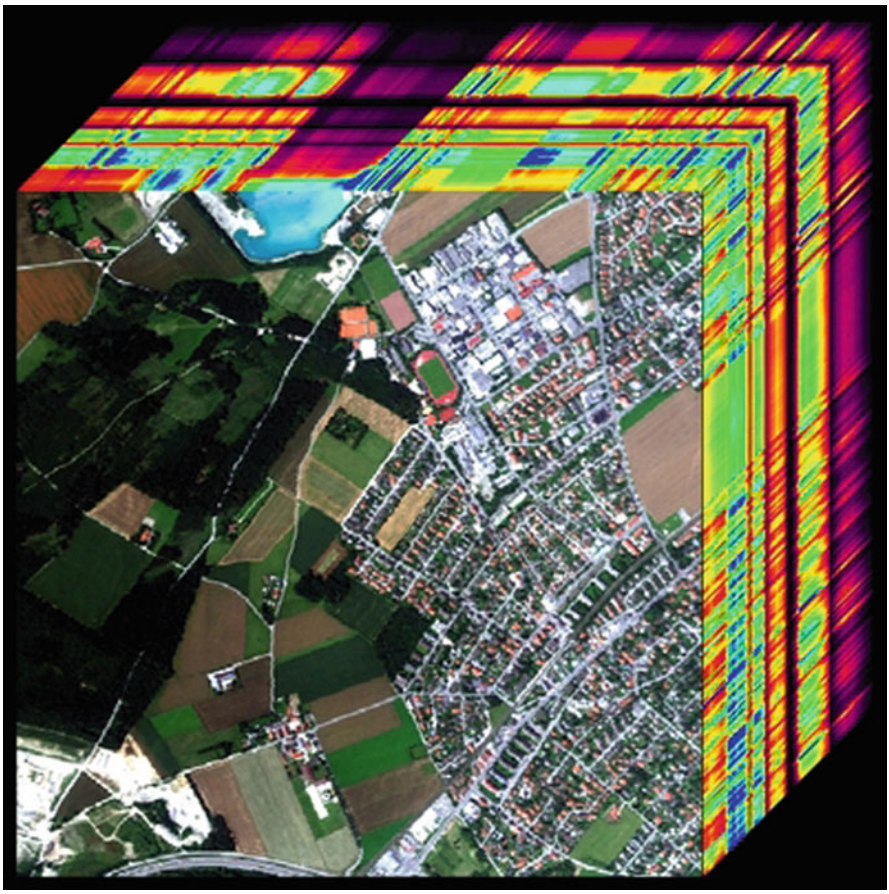


Fig. 29.4 A hyperspectral data cube that shows the spectra of the sensed spatial image (Image courtesy of DLR –The German Space Agency)

Electro-optical Sensing

The first hyperspectral imaging was restricted to the use of spectrometer technology that are scientific instruments that have a number of constraints in terms of being able to collect data, effectively send it back to earth for processing, as well as high costs associated with the equipment, the sensing and acquisition of data and the final processing and display of the data. The development of electro-optical sensing and particularly charge-coupled devices (CCDs) and CCD arrays have served to lower the cost of hyperspectral sensing.

In short the advent of CCD devices brings many advantages to satellite-based hyperspectral sensing in terms of costs and ease of data display. The various satellites described above such as PRISMA, EnMap, and HyspIRI all use

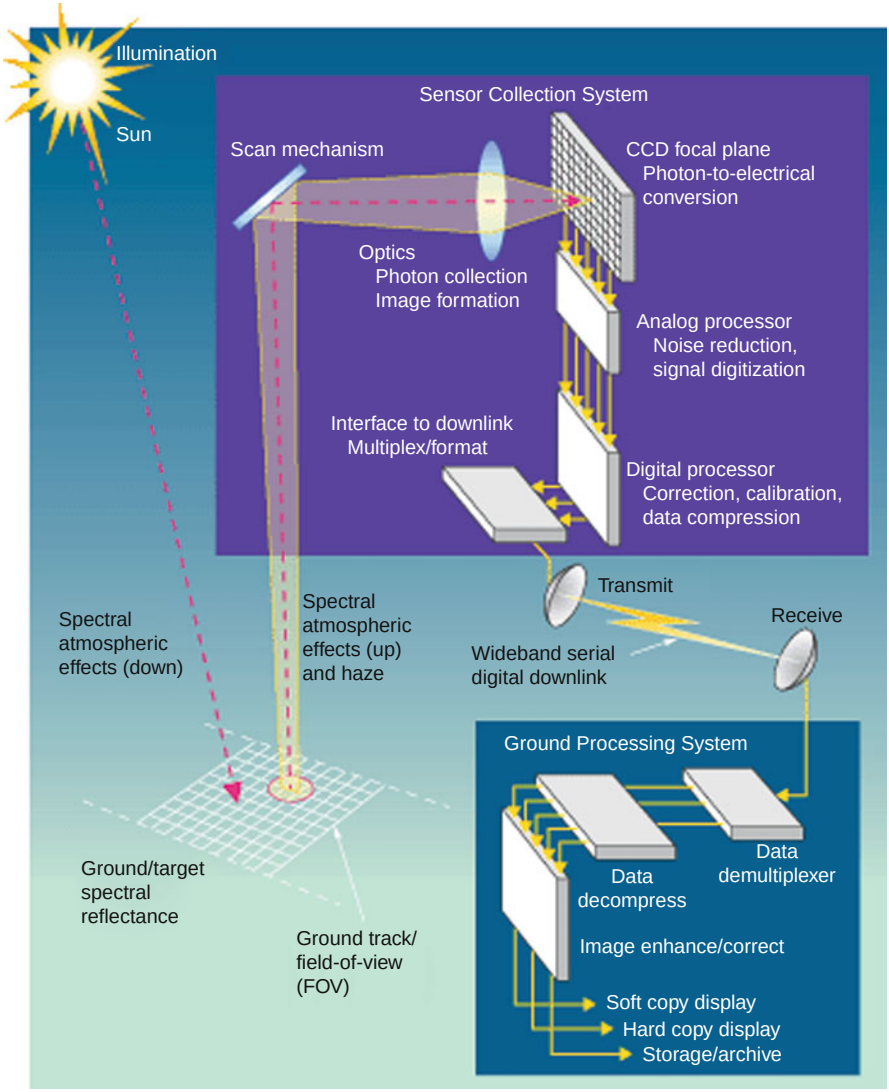


Fig. 29.5 Remote sensing using CCD for electro-optical imaging[5]

electro-optical CCD devices to gather hyperspectral images and then send them back to Earth as electronically coded information for processing.

In time remote sensing satellites designed to gather this type of information may also be designed with preprocessing capabilities to "discard" or "sort through" various images that have already been captured or where changes are not taking place.

Figure 29.5 above shows how that a satellite equipped with a CCD array designed for hyperspectral sensing actually works to capture images into data sets that can then be efficiently sent back to earth.

Conclusion

The shift from multispectral sensing to hyperspectral sensing is still in the process of occurring. The value that comes from this new way of capturing and displaying data as hyperspectral data cubes is rather universally understood to represent a breakthrough in the ability to analyze and understand what the images from the new types of remote sensing satellite and their sensors can tell us – especially when combined with panchromatic or radar satellites. There is still a great deal to be done in terms of perfecting the satellite sensor capabilities, the formatting of data cubes, the efficient transmission to ground stations – and the possible future of preprocessing, and the in-depth analysis of the data on the ground.

Today there are a growing number of applications that are evolving with regard to hyperspectral imaging. The fact that hyperspectral images tell us about gases (i.e., clouds, pollution, greenhouse gases, etc.), about liquids (oceans, lakes, rivers, pollution, droughts) and about solids (minerals, ores, vegetation, forests, etc.) strongly suggest that the scientific and practical applications can continue to expand for some time to come. The rapid evolution of CCD technology for a wide range of industrial and consumer applications will doubtlessly help the evolution of this technology in terms of more effective remote sensing via satellites as well as lead to new economies.

Notes

1. Hyperspectral Remote Sensing http://www.csr.utexas.edu/projects/rs/hrs/hyper.html
2. Satellite Sends Hyperspectral Image from Space-Laser Focus World http://www.laserfocusworld.com/articles/print/volume-37/issue-5/features/hyperspectral-imaging/satellite-sends-hyperspectral-images-from-space.html
3. Going Hyperspectral, ESA Newsletter, About Proba-1 http://www.esa.int/esaMI/Proba_web_site/SEMFVBIK97G_0.html
4. Prisma – A.S.I.-Agenzia Spaziale Italiana www.asi.it/en/activity/earth_observation/**prisma**
5. D. L. Glackin, Remote sensing: an overview, *Crosslink J. Aeros. Corp*. http://www.google.com/imgres?imgurl=http://www.aero.org/publications/crosslink/summer2004/images/01_02.gif&imgrefurl=http://www.aero.org/publications/crosslink/summer2004/01.html&h=450&w=366&sz=32&tbnid=UtHoxx1qmPk-hM:&tbnh=127&tbnw=103&prev=/search%3Fq%3Delectro-optical%2Bsensors%26tbm%3Disch%26tbo%3Du&zoom=1&q=electro-optical+sensors&usg=__K476B75tvZFgP93q0Tvmr1YGz7g=&sa=X&ei=6HNCT_7_HqXu0gHtmqDJBw& ved=0CCkQ9QEwBg

Operational Applications of Radar Images 30

Vern Singhroy

Contents

Introduction 740
Agriculture 740
Geological Applications 743
SAR Data Fusion for Mineral Exploration 745
Geological Hazards 746
Flood Monitoring 748
Oil Spill Monitoring 752
Sea Ice Monitoring 752
References 755

Abstract

This chapter provides some examples of the operational uses of satellite radar images. These include the uses of polarimetric radar images for crop classification and earthquake damage assessment; radar image fusion for mineral exploration; interferometric SAR techniques for landslide and volcanic monitoring; multidate radar image enhancement techniques for oil spill monitoring and flood mapping, and sea ice mapping from enhanced scanSAR images. In the near future, new applications will be developed from current and future advanced SAR missions involving their high resolution, rapid revisits, and polarimetric capabilities.

Keywords

Radar applications • polarimetric and interferometric SAR • radar missions

V. Singhroy
Canada Centre for Remote Sensing, Natural Recources Canada, 588 Booth St, K1A0Y7 Ottawa, Canada
e-mail: vern.singhroy@ccrs.nrcan.gc.ca

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_43, © Springer Science+Business Media New York 2013

Introduction

This chapter describes the operational applications of radar images that have developed over the past 20 years. Table 30.1 provides a summary of current and future satellite missions and their technical specifications. The European and Canadian space agencies focus on launching high-resolution fully polarimetric C-band (5.7 cm) satellites. The German and Italian space agencies launch X-band (3 cm) satellites and the Japanese space agency launches L-band radar satellites. All these multifrequency satellites provide images with different resolutions, viewing geometries, and polarization. In the next few years, there will be number of high-resolution radar constellation satellites that will be launched by the European and Canadian space agencies. These new-generation smaller radar constellation satellites will provide new applications for both civilian and military uses. For instance, the RADARSAT Constellation is designed as a scalable constellation of three small satellites. With a constellation, the time between successive imaging of the same part of the Earth (revisit time) is significantly reduced. The creation of a three-satellite constellation will increase the frequency of available information, as well as the reliability of the system, making it better suited to operational requirements. The RADARSAT Constellation will provide all-weather day and night data in support of maritime surveillance, disaster management, and ecosystem monitoring. In addition, these satellites will continue to provide images for research on information extraction and image fusion techniques.

In this chapter, operational examples in agriculture, geology, geohazards, ice, oil spills, and flood monitoring are discussed. Several emerging applications such as forestry, ship detection, and others are not addressed. Table 30.2 summarizes the numerous application of radar images, some mature and others emerging. Although this table was prepared for RADARSAT applications, it represents most of the other radar satellite applications. This summary (Table 30.2, Van der Sanden and Ross 2004) also facilitates the identification of application fields and information extraction techniques that require further research and development.

Agriculture

The development of key applications for space-borne SAR data in agriculture has been hindered by the limited information content of single date, single frequency, and single polarization images such as available from first generation SAR satellites. Some successes were achieved through the application of multitemporal SAR data sets but images from optical satellites have remained the data source of choice. Optical sensors such as Landsat Thematic Mapper and SPOT are used extensively for crop inventory. However, because of cloud cover, gaps in optical image acquisition can cause inaccuracies in crop classification. In Canada, for instance, mid to late summer season optical images are essential for accurate crop

Table 30.1 Current and future radar satellite missions

Satellite	ERS-1	ERS2	Radarsat-1	JERS-1	Envisat	Radarsat-2	AIOS	TerraSAR-X	Cosmo/ SkyMEd	TanDEMX (TDX)	Sentine	RADARSAT constellation mission (RCM)
Space agency	ESA	ESA	CSA	JAXA	ESA	CSA/MDA	JAXA	DLR/Infoterra Gmbh	ASI	DLR/ Astrium	ESA	CSA
Launch	1991	1995	1995	1992	2002	2005	2004	2006	2005	2010	2013	2015
Out of service since	2000			1998			2011					
Band	C	C	C	L	C	C	L	X	X	X	C	C
Wavelength (cm)	5.7	5.7	5.7	23.5	5.7	5.7	23.5	3	3	3	5.4	5.4
Polarization	VV	VV	HH	HH	HH/ VV	QUAD-Pol	All	All	HH/VV	All	Dual (VV VH,HH HV)	HH,VV,HV, VH, compact polarimetry
Incidence angle (°)	23	23	20–50	35	15–45	10–60	8–60	15–60	Variable	20–55	20–45	10–60
Resolution range (m)	26	26	10–100	18	30–150	3–100	7–100	1–16	1–100	1–16	5–40	1–100
Resolution azimuth (m)	28	28	9–100	18	30–150	3–100	7–100	1–16	1–100	1–16	5–40	3–100
Scene width (km)	100	100	45–100	75	56–400	50–500	40–350	5–100 (up to 350)	10–200 (up to 1,300)	5–100	20–400	5–500
Repeat cycle (days)	35	35 (3)	24	44	35	24	24–6	2–11	5–16	2–11	12	5–14
Orbital elevation (km)	785	785	798	568	800	798	660	514	619	514	693	592

Table 30.2 Application potential of RADARSAT-1 and RADARSAT-2[a] (van der Sanden and Ross 2004)

	Satellite	
Application	RADARSAT-1	RADARSAT-2
Agriculture		
Crop type	–/+	++
Crop condition	–/+	++
Crop yield	–	–/+
Cartography		
DEM interferometry	+	+
DEM stereoscopy	++	++
DEM polarimetry	N.A.	–/+
Cartographic feature extraction	+	++
Disaster management		
Floods	++	++
Geological hazards	–/+	+
Hurricanes	+	+
Oil spills	+	+
Search and rescue	–/+	+
Forestry		
Forest type	–/+	–/+
Clear-cuts	–/+	+
Fire-scars	–/+	+
Biomass	–	–
Geology		
Terrain mapping	–/+	+
Structure	+	++
Lithology	–/+	–/+
Hydrology		
Soil moisture	–/+	+
Snow	–/+	+
Wetlands	–/+	+
Oceans		
Winds	+	++
Ships	+	++
Waves	–/+	–/+
Currents	–/+	+
Coastal zones	–/+	+
Sea and land ice		
Sea ice edge and ice concentration	+	++
Sea ice type	+	+
Sea ice topography and structure	–/+	++
Icebergs	–/+	+
Polar glaciology	–/+	+

Key: – minimal, –/+ limited, + moderate, ++ strong
[a]Use of single date images assumed

classification that used for crop acreage estimates. A combination of radar images acquired during the cloudy periods and optical images acquired during the cloud free periods can significantly improve crop classification accuracies (McNairn et al. 2002, 2009)

The polarization diversity offered by current radar satellites such as RADARSAT-2 can be expected to enhance the potential of space-borne SAR for application to agriculture. The introduction of systems operating with longer wavelengths will also help in mapping high biomass crops (e.g., corn) and related soil properties. Figure 30.1a shows an example of a RADARSAT 2 polarimetric classification of temperate crops near Ottawa, Canada (McNairn et al. 2009). A similar polarimetric classification was done for tropical crops in Guyana (Fig. 30.1b). Both these examples show the improvement made in recent years on using polarimetric SAR images for crop classification. In the near future, this classification will significantly improve crop inventory and land use mapping in cloudy tropical areas.

Soil moisture information also supports agriculture and hydrological management such as drought and flood prediction, crop irrigation scheduling, pest management, and others. The monitoring of soil moisture by means of space-borne SAR has been studied extensively but has to become a viable operational application. The effect of soil roughness and the limited penetration depth, in both soil and potentially overlying vegetation, of C-band SAR systems is a major issue. Systems operating at lower frequencies, for example, L-band have the capability to capture soil moisture information at larger depth in the soil profile.

Geological Applications

There are numerous studies dealing with geologic remote sensing using radar, optical, and thermal images. Many useful case studies are included in the books by Rencz (1999), Sabins (1997), Rivard (2011), and others. This chapter will focus only on new advanced SAR techniques for geological mapping, mineral exploration, and InSAR techniques for monitoring high-risk geohazards.

Recent results have shown that radar images have provided geologists with useful information for geomorphology, geological structure, and rock units (Lowman 1994; Singhroy et al. 1993; Singhroy 1992; Singhroy and Saint Jean 1999; Saint-Jean et al. 1999). Geological mapping are facilitated by the uses of high-resolution stereo SAR images which improve image interpretation techniques for terrain mapping, and that some lineament orientations are enhanced by cross-polarized images (Saint-Jean et al. 1999). Polarimetric SAR images are providing specific information on the scattering mechanism of rock units and surficial materials and are particularly useful in mapping arid areas and planetary surfaces (Singhroy and Molch 2004b).

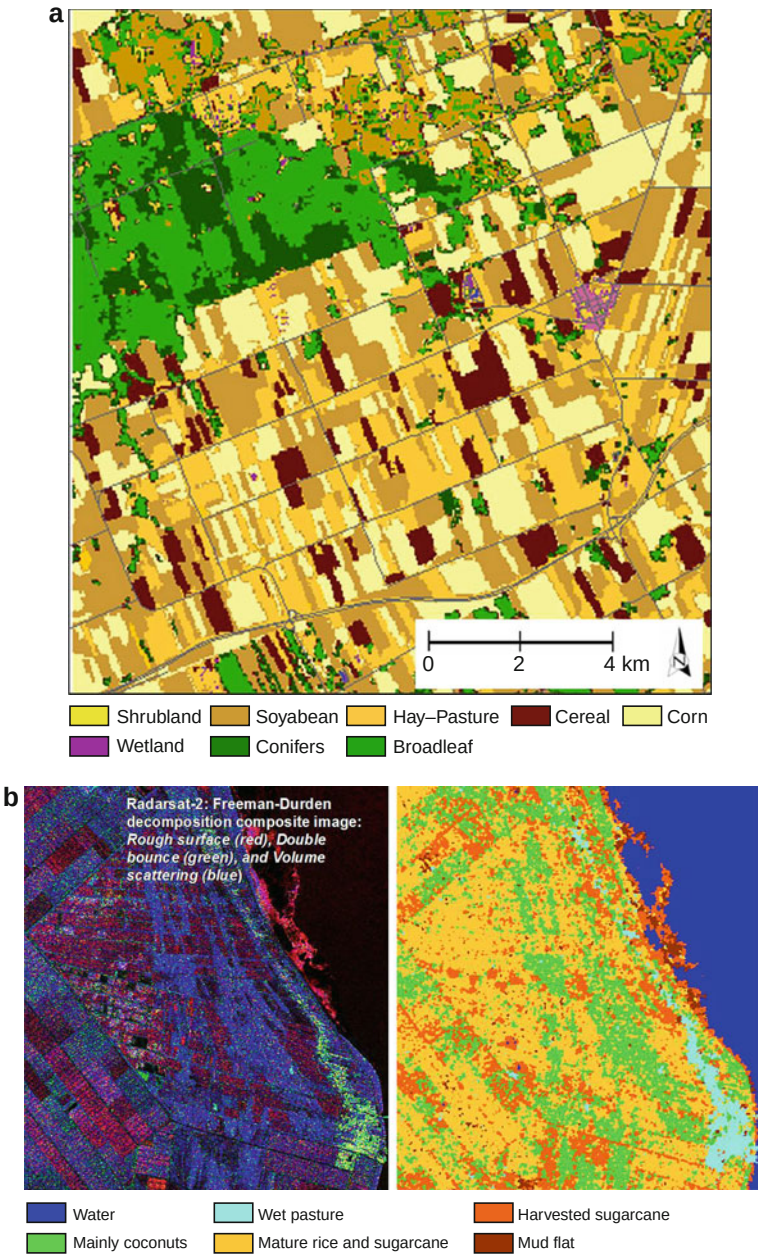


Fig. 30.1 RADARSAT 2 polarimetric classification of agricultural crops (**a**) Ottawa, Canada and (**b**) Berbice, Guyana

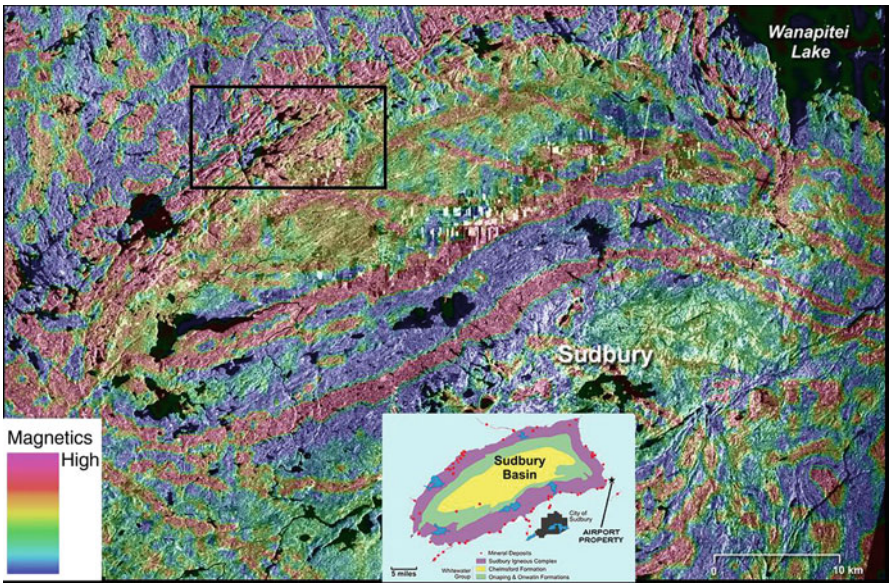


Fig. 30.2 RADARSAT image fusion map used for mineral exploration in Sudbury, Canada. This map was produced by fusing a RADARSAT standard mode image (25 m resolution) with vertical gradient magnetic image. Nickel, copper, and gold mines are found where high magnetic signatures intersect with faults shown as *linears*

SAR Data Fusion for Mineral Exploration

The fusion of RADARSAT and other geophysical images in support of geological mapping and mineral exploration is outlined in this section. Singhroy et al. (1993), Singhroy (1996), Singhroy and Molch (2004b), Rivard et al. (1999) and others have developed interpretation techniques for RADARSAT-fused images to facilitate mineral exploration programs in Canada. This technique involves fusing RADARSAT images with vertical gradient magnetic, radiometric, optical, and topographic data, to produce thematic exploration image maps for exploration and geological mapping programs.

An example of RADARSAT fused images is shown for the Sudbury Basin in Canada (Fig. 30.2). The Sudbury Basin is one of the world's richest mining areas and the oldest, largest, and best-exposed meteorite impact site. Current production of nickel, copper, and gold is about $2 billion a year. Figure 30.2 shows a standard RADARSAT–1 image (25 m resolution) integrated with magnetic data. The Standard beam mode of RADARSAT, at 20–27° incidence angles, provides an excellent view of the topography and structural features of the Sudbury area. The geological structure and rock types, which controls mineralization, are clearly shown on the fused SAR image. Areas in red outline the strong magnetic signature of the Sudbury

Igneous complex where most of the mines are found. NE-SW Magma-filled dykes also have highly magnetic signatures. The other colors represent different rock types with different magnetic signatures.

Geological Hazards

Geological hazards such as earthquakes, volcanoes, and landslides are the main natural causes of damage to human settlements and infrastructures killing thousands every year. As population increases and habitation on hazardous land areas becomes more common, the risks posed by geohazards increases. The need to observe their behavior, understand them better, and mitigate their effects becomes ever more urgent (IGOS 2004).

Current state of the art in real-time monitoring of active geohazards areas developed for early warning is very expensive. Satellite radar interferometry (InSAR) is used to complement real time monitoring such as GPS, seismometers, and in situ field measurements (Singhroy 2009), and others. Provided coherence is maintained over longer periods, it is possible to measure surface displacement of a few centimeters per year. Using InSAR data with small perpendicular baselines, short time intervals between acquisitions, and correcting the effect of topography on the interferogram, reliable measurements of surface displacement are achieved.

Landslides: Interferometric Synthetic Aperture Radar (InSAR) techniques are being used to measure small millimeter displacement on slow moving landslides. An interferogram represents the phase differences between the backscatter signals in two or more SAR images obtained from similar positions in space. In case of space-borne SAR, the images are acquired from repeat-pass orbits. The phase differences between two repeat-pass images are the result of changes in topography, changes in the line-of-sight distance (range) to the radar, due to displacement of the surface and change in the atmospheric conditions between scenes. On nonmoving target, the phase differences can be converted into a digital elevation model if very precise satellite orbit data are available. Landslide movements are measured in millimeters to centimeters per orbit cycle of the radar satellite. This orbit cycles can range from 44 days for ALOS, 24 days for RADARSAT 2 and a minimum of 2–5 days for TerraSAR X and Cosmo-SkyMed (Table 30.1). Figure 30.3 shows an example of InSAR technique for monitoring the Frank Side in Canada. The Frank Side is an active 30×106 m^3 rockslide-avalanche of Paleozoic limestone, which occurred in April 1903 from the east face of Turtle Mountain in southern Alberta, Canada. Seventy fatalities were recorded (Cruden and Hungr 1986). Detailed analysis on the use of SAR images for characterizing and monitoring the Frank slide is reported by Singhroy and Molch (2004a) and Singhroy et al. (1998). Our results show that this landslide have moved about 150 mm from 2008 to 2011, using RADARSAT 2 InSAR analysis.

Earthquakes: Plate tectonics provides a framework for understanding earthquake activity. The earthquake-prone regions of Earth are well delineated and global seismicity information is readily available. However, there is a need to

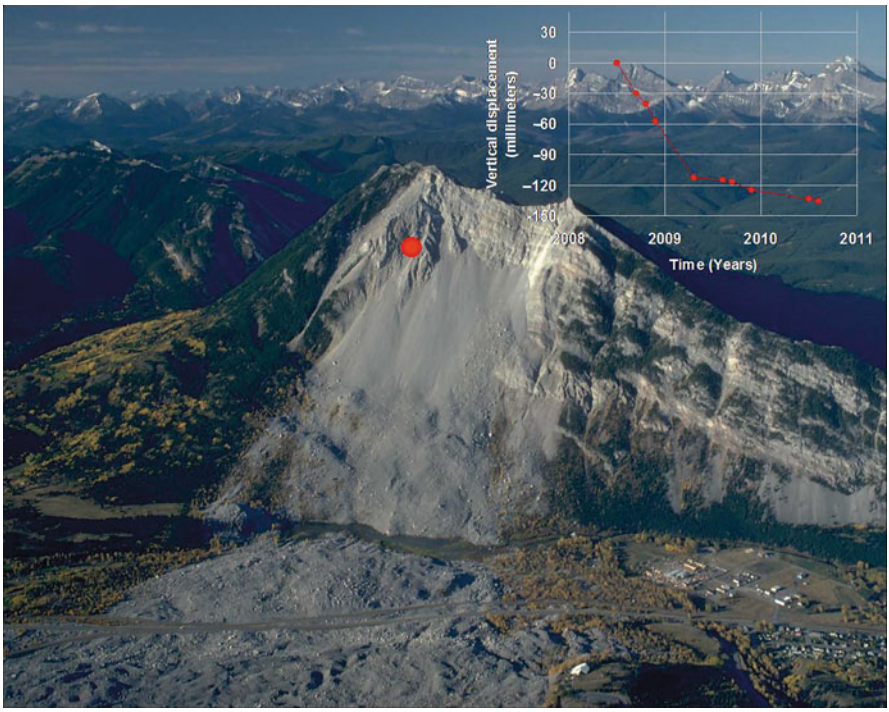


Fig. 30.3 Turtle Mountain, rock avalanche occurred in 1903, killing 70: RADARSAT 2 InSAR measurements show that the landslide is still active and has moved 150 cm since 2008

improve the identification and characterization of seismically active zones. Satellite radar interferometry is in the early stages in monitoring earthquake deformation aimed at producing more accurate hazard maps (IGOS 2004). Recently, Zhang et al. (2011), and others have used InSAR to monitor coseismic events and generated vertical displacement maps related to the 2008 Sichuan magnitude 7.5 and the 2010 Haiti, magnitude 7.0 earthquakes. In addition, high-resolution optical data are widely used for earthquake damage assessment (Ajmar et al. 2011). Figure 30.4 shows the uses of RADARSAT 2 polarimetric images to map flooded areas in the urban coastal zones and adjacent rice fields. These areas were severely damaged from the 2011 magnitude 9 earthquake triggered tsunami near Sendai, Japan.

Volcanoes: Volcanic hazards vary from one volcano to another and from one eruption to the next. The most damage is from pyroclastic flows and lahars. Civil authorities need clear information on detectable changes, such as ground cracking, associated earthquakes, and SO_2 emissions. Recently InSAR techniques have been used to monitor the deformation activity of active volcanoes for risk assessment to populated areas. Figure 30.5 shows the uses of D-InSAR techniques for post eruption shrinkage rates at the Miyake-Jima volcano, after the severe August 2000 eruption (Singhroy et al. 2004).



Fig. 30.4 Sendai, Japan, tsunami damage assessment image map showing flooded areas in urban coastal areas and adjacent rice fields. This map was produced from a polarimetric classification of RADARSAT 2 image taken soon after the 2011 magnitude 9 earthquake

Flood Monitoring

Floods are among the most devastating natural hazards in the world, claiming more lives, and causing more property damage than any other natural hazards. In 2011 alone flood devastation were experienced in Pakistan, China, Australia, Thailand, and Canada. Within the USA, an average of more than 225 people are killed and more

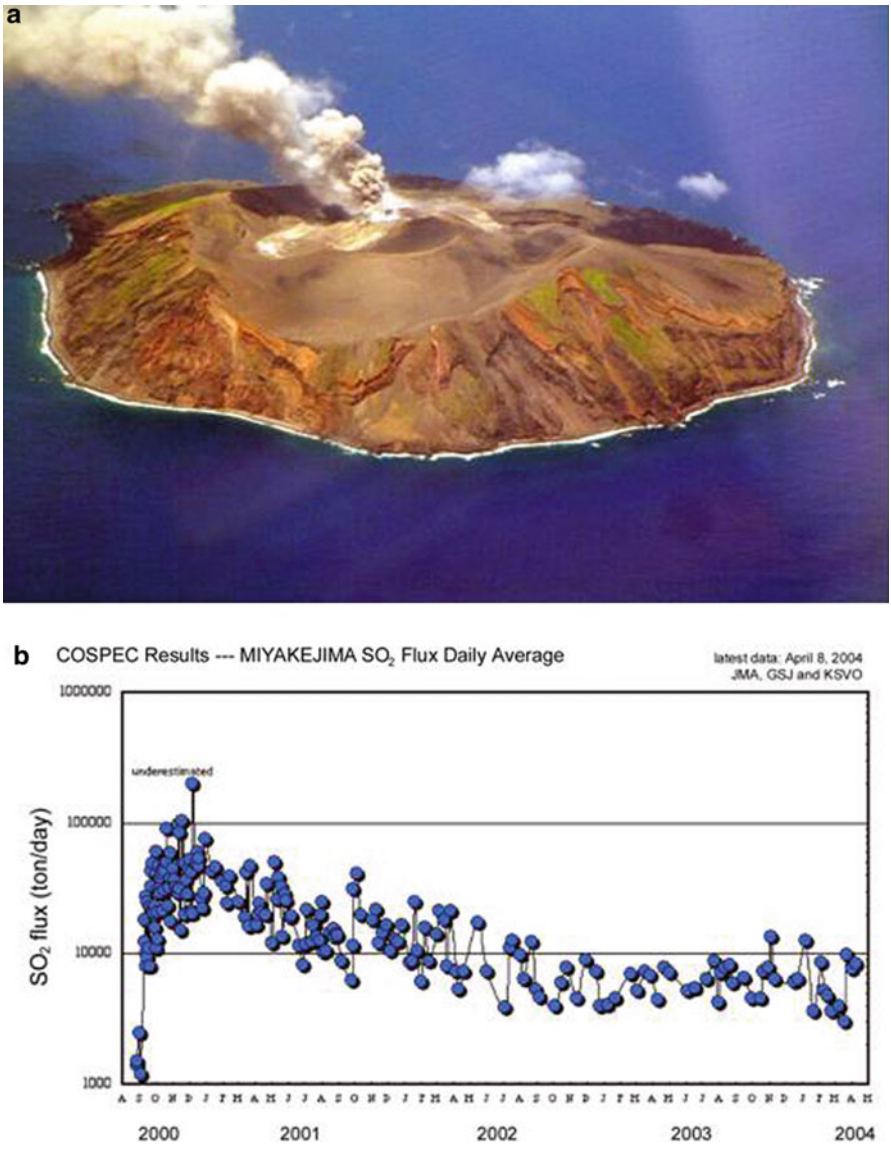


Fig. 30.5 (continued)

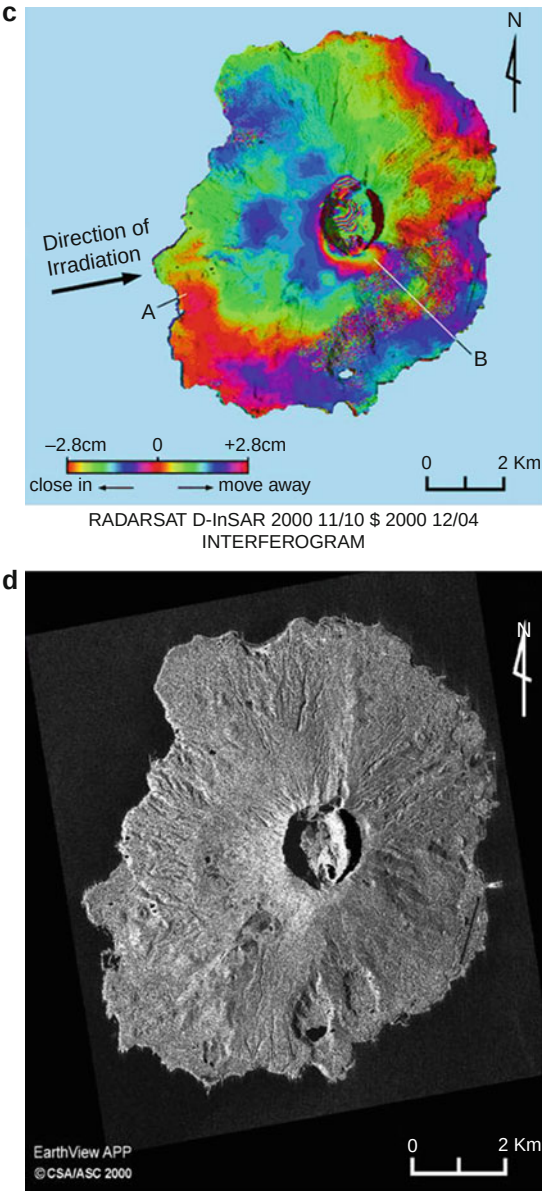


Fig. 30.5 Differential interferometric (D-InSAR) technique showing post eruption shrinkage rates at the Miyake-Jima volcano, Japan after the severe August 2000 eruption (Singhroy et al. 2004)

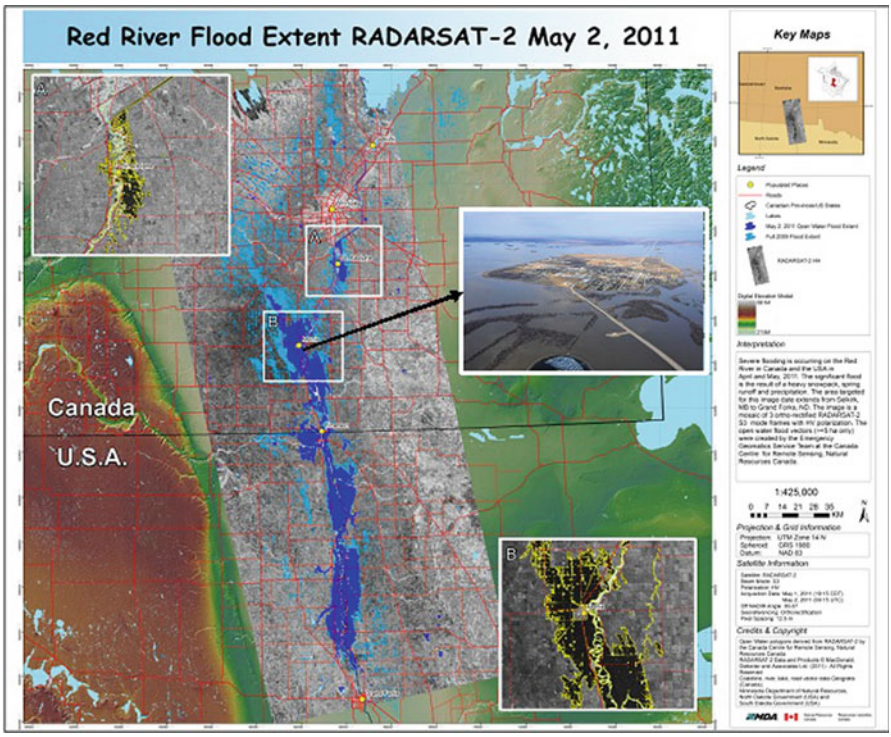


Fig. 30.6 RADARSAT 2 flood extent image map used in support of flood response programs during the 2011 Winnipeg spring floods in Canada. The radar image was combined with other GIS data such as land use, slope, aspect, and transportation and infrastructure networks

than US$3.5 billion in property is damaged by heavy rain and flooding each year (CEOS 2003) Satellite-derived flood maps produced in near-real time are therefore invaluable to local and national agencies for disaster monitoring and relief efforts.

Satellite optical observations of floods are hampered by the presence of clouds, which normally prevent near real-time data acquisitions. It is well known that SAR is an excellent tool, which supports of the monitoring of floods. The SAR data are not only restricted to flood mapping but can also be useful to the estimation of a number of hydrological parameters. For instance, radar images were used operationally to map flood extent, wet snow, to monitor wetlands, and to identify freshwater ice types (Pultz et al. 1997). Figure 30.6 shows an example of a RADARSAT 2 flood extent image map used in support of flood response programs during the 2011 Winnipeg spring floods in Canada. In this case, the RADARSAT 2 images were combined with other GIS data such as land use, slope, aspect, and transportation and infrastructure networks. The analysis was performed on the high-risk areas where decisions were made to evacuate populated areas affected by the flooding. The integration of GIS

data layers with current satellite imagery requires time and effort, but also renders this type of product shown in Fig. 30.6 as valuable for disaster mitigation and prevention and response purposes.

Oil Spill Monitoring

Radar is ideally suited for the detection of sea surface slicks. Ocean wave properties, particularly capillary waves, are significantly dampened by the presence of surfactants with higher viscosities than water. Wave roughness is reduced and radar backscatter through Bragg scattering is significantly lowered, causing a large contrast between the surrounding ocean and the area affected by oil.

Hydrocarbons, however, are main target of interest for detection and monitoring purposes as their presence may be related to natural seep, indicating a subsurface hydrocarbon source or reservoir. These locations are of interest for petroleum exploration and for understanding the hydrocarbon basins. The second important application is for man-made oil spills. These can be deliberate acts of polluting waterways by ships discharging bilge oil, or accidental leaks from ships or undersea pipeline infrastructure.

Significant benefits have been realized worldwide by transnational petroleum companies in petroleum producing basins (e.g., Gulf of Mexico, North Sea, and the NE coast of Brazil) using this technology. Figure 30.7 (a) shows an example of RADARSAT 2 images to detect the 2010 BP oil spill in the Gulf of Mexico.

To prevent illegal dumping at sea, pollution monitoring and enforcement have relied on space-borne SAR by several countries. The inset at Fig. 30.7b shows ships (white dots) discharging bilge oil near in the Indian Ocean. Currently, legislation has been set up to allow coastal states to inspect all shipping within territorial waters and also to ensure that national legislation preventing any dumping applies equally to national and foreign owned shipping.

Research into combining information from multifrequency and multitemporal SAR offers an opportunity to advance the application of SAR for oil spill detection and classification.

Sea Ice Monitoring

In the Polar Regions, sea ice varies both spatially and temporally due to high variability in the environmental processes. National ice service agencies normally produce charts and maps showing current sea ice conditions. These maps must also show the location of the ice edge, concentration and distribution, stage of development, floe size, amount of pressure ridging or topography, location and orientation of ice openings, degree of ice compaction and divergence and stage of decay during the summer melt season. Ideally, vessels at sea prefer to receive high-resolution satellite images that are less than 6 h old and have been interpreted to provide the

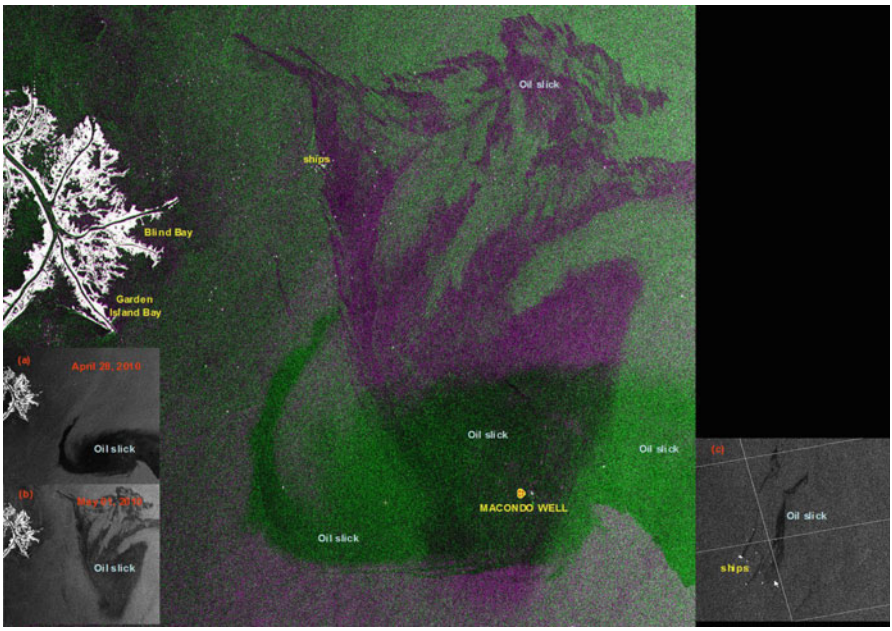


Fig. 30.7 (**a**) RADARSAT 2 image used to detect the 2010 BP oil spill in the Gulf of Mexico. The *insets* at (**a**) and (**b**) show oil slicks for April 28 and May 1, 2010. The colors represent the oil slicks during these two dates. The *white dots* show a large number of ships involved in the cleanup efforts. (**c**) Detection of illegal dumping of bilge oil from ships in the Indian Ocean

above information necessary to avoid or exploit the ice (Skriver and Pedersen 1995; CEOS 2003). It is also well known that sea ice and icebergs pose a serious hazard to shipping and other maritime activities in these regions. Because of their high resolution, all weather, wide swath, and ice detection capability, radar images are particularly useful for monitoring of sea ice conditions in the waters surrounding Canada and other polar regions.

The Canadian Ice Service (CIS) is using RADARSAT images to quantify various ice parameters. These include the mapping of ice extent (concentration) and ice type (stage of development), such as ice topography, presence of open water or thin ice openings within the sea ice pack, stages of ice decay and others (Johnson and Timco 2008). Maps of sea ice derived in real time from these data are used operationally to ensure safety of navigation by all vessels operating in the Canadian north. The data products of choice for the Canadian Ice Service, for example, are those that provide wide area coverage (i.e., ScanSAR products). Figure 30.8 is an example of a September 2011 RADARSAT 2 ScanSAR image map of the sea ice condition near Ellesmere Island in the Canadian Arctic. These image maps are not only used for navigation safety, but to monitor the ice condition that are being affected by climate change.

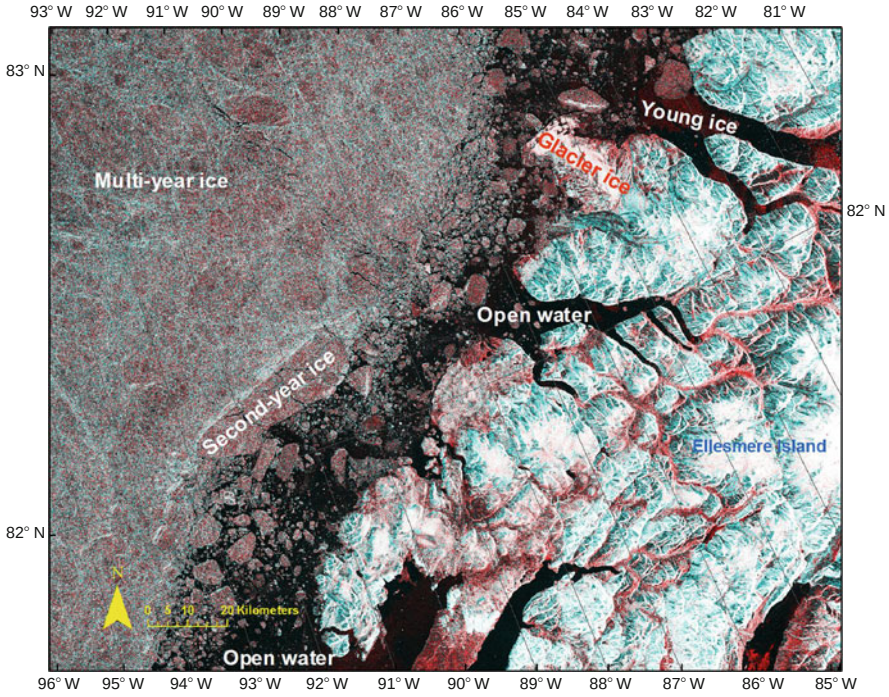


Fig. 30.8 RADARSAT 2 Scan SAR C-HH image map of the sea ice types and condition near Ellesmere Island in the Canadian Arctic. The image map is used for navigation safety and to monitor the ice condition that are being affected by climate change

Conclusion

It is clear that radar images are being used routinely for a number of operational applications. Emerging applications are being developed from current and future advanced SAR missions involving their high resolution, rapid revisits, and polarimetric capabilities. There is a lack of understanding with regard to the relationship between the image information content and the polarization of the radar signal. There have been significant developments on the uses of both dual and fully polarimetric data in some applications, for example, in forestry. However, it is clear that future research should continue to explore the additional uses of fully polarimetric and compact polarimetric data that will available from future constellations such as the CSA-RADARSAT and ESA Sentinel programs. In addition, significant developments have also been made on using InSAR techniques for geohazards and disaster mitigation. New 3D InSAR products and polarimetric InSAR techniques needed further evaluation.

References

A. Ajmar, P. Boccardo, F. Tonolo, Earthquake damage assessment based on remote sensing data. The Haiti case study. Italian J. Rem. Sens. **43**(2), 123–128 (2011)

CEOS (2003) The use of earth observing satellites for hazard support: assessment and scenarios NOAA. p147

D.M. Cruden, O. Hungr, The debris of Frank slide and theories of rockslide-avalanche mobility. Can. J. Earth Sci. **23**, 425–432 (1986)

IGOS Geohazard Theme Report (2004) IGOS Geohazard Secretariat, BRGM, France, 50 p

M.E. Johnson, G.W. Timco, Understanding and identifying old ice in summer. Publication # CHC-TR-055. Canadian Hydraulics Centre, Ottawa, 236p (2008)

P. Lowman, Radar geology of the Canadian shield – a 10 year review. Can. J. Rem. Sens. **20**(3), 198–209 (1994)

S.M. Moran, C.D. Peters-Lidard, J.M. Watts, S. McElroy, Estimating soil moisture at the watershed scale with satellite-based radar and land surface models. Can. J Rem. Sen. **30**(5), 805–826 (2004)

H. McNairn, C. Duguay, B. Brisco, T. Pultz, The effect of soil and crop residue characteristics on polarimetric radar response. Rem. Sens. Environ. **80**, 308–320 (2002)

H. McNairn, C. Champagne, J. Shang, D. Holmstrom, G. Reichert, Integration of optical and synthetic aperture radar (SAR) imagery for delivering operational annual crop inventories. J. Photogrammetry Rem. Sens. **64**(5), 434–449 (2009)

T. Pultz, Y. Crevier, R. Brown, J. Boisvert, Monitoring of local environmental conditions with SIR-C/X-SAR. Rem. Sens. Environ. **59**(4), 248–255 (1997)

A.N. Rencz, *Remote Sensing for the Earth Sciences: Manual of Remote Sensing*, 3rd edn. (Wiley, New York, 1999). 700 p

B. Rivard, L. Corriveau, L.B. Harris, Structural reconnaissance of deep crustal orogen using RADARSAT and landsat satellite imagery and airborne geophysics. Can. J. Rem. Sens. **25**(3), 258–267 (1999)

L.A. Rivard, *Satellite Geology and Photo- geomorphology* (Springer, Berlin, 2011). 270 p

F. Sabins, *Remote Sensing: Principles and Interpretations*, 3rd edn. (Freeman, New York, 1997). 432 p

R. Saint-Jean, V. Singhroy, M. Rheault (1999) Understanding multi-polarized airborne C-SAR images for geological mapping in precambrian shield terrains. in *Proceedings Applied Geologic Remote Sensing, ERIM 13th International Conference*, vol. 1 (Vancouver, British, 1999), pp. 411–418

V. Singhroy, Radar geology, techniques and results. Episodes **15**(1), 15–20 (1992)

V. Singhroy, R. Slaney, P. Lowman, J. Harris, W. Moon, RADARSAT and radar geology in Canada. Can. J. Rem. Sen. **14**(4), 329–332 (1993)

V. Singhroy, Environmental and geological site characterization in vegetated areas: Image enhancement guidelines, in *Remote Sensing and GIS: Applications and Standards*, ASTM Special Technical Publication no. 1279, ed by V. Singhroy, D. Nebert, A. Johnson (American Society for Testing and Materials, 1996), pp. 5–17

V. Singhroy, K. Mattar, L. Gray, Landslide characterization in Canada using interferometric SAR and combined SAR and TM images. Adv. Space Res. **2**(3), 465–476 (1998)

V. Singhroy, R. Saint-Jean, Effects of relief on the selection of RADARSAT-1 incidence angle for geological applications. Can. J. Rem. Sens. **25**(3), 211–217 (1999)

V. Singhroy, K. Molch, Characterizing and monitoring rockslides from SAR techniques. Adv. Space Res. **33**. 290–295 (2004a)

V. Singhroy, K. Molch, Geological applications of RADARSAT-2. Can J. Rem. Sens. **30**(6), 893–902 (2004b)

V. Singhroy, H. Okhura, K. Molch, R. Couture, Monitoring landslides and volcanic deformation from InSAR. in *Proceedings ISPRS Congress*, Istanbul, Turkey CD Paper, pp. 570–74

V. Singhroy, Remote sensing techniques for geological mapping and exploration, in *Geoinfomatics for Natural Resource Management*, ed. by P.K. Joshi, P. Pani, S.N. Mohapatra, T.P. Singh (Nova Science, New York, 2009), pp. 333–349. Chapter 15

H. Skriver, L.T. Pedersen, Polarimetric signatures of sea ice in the Greenland Sea, in *Proceedings of the International Geoscience and Remote Sensing Symposium, IGARSS '95*, Florence, Italy, 10–14 July 1995, pp 1792–1794

J. Van der Sanden, S. Ross, Applications potential of RADARSAT 2- a preview: CCRS, Earth Science Sector Contribution Series 20043000, p 86 (2004)

G. Zhang, C. Qu, X. Shan, X. Song, G. Zhang, C. Wang, J. Hu, R. Wang, Slip distribution of the 2008 Wenchuan, Mag 7.9 earthquake by joint inversion from GPS and InSAR measurements; a resolution test study. Geophys. J. Int. **186**(1), 207–220 (2011)

Lidar Remote Sensing

31

Juan Carlos Fernandez Diaz, William E. Carter, Ramesh L. Shrestha, and Craig L. Glennie

Contents

Introduction 758
Origins of LiDAR Technology 759
High-Level Technical Overview of LiDAR 762
Ranging Methods 766
  Optical Triangulation 766
  Phase Difference 766
  Time of Flight (TOF) 767
  Hybrid Systems 768
Light–Target Interaction Phenomena 768
  Scattering 768
  Reflection 769
  Absorption 769
  Fluorescence 770
  Doppler 770
  Depolarization 771

J.C. Fernandez Diaz (✉)
NSF National Center for Airborne Laser Mapping (NCALM)/Department of Civil and Environmental Engineering, University of Houston, Houston, TX, USA
and
University of Houston, 1 Hermann Museum Circle Dr. # 3006, 77004 Houston, TX, USA
e-mail: jfernandezhon@yahoo.com, jcfernan@ufl.edu

W.E. Carter • R.L. Shrestha • C.L. Glennie
NSF National Center for Airborne Laser Mapping (NCALM)/Department of Civil and Environmental Engineering, University of Houston, Houston, TX, USA
and
NSF National Center for Airborne Laser Mapping (NCALM)/Department of Civil and Environmental Engineering, University of Houston, 77204-4003 Houston, TX, USA
e-mail: carter4451@bellsouth.net, rlshrestha@uh.edu, clglenni@central.uh.edu

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_44, © Springer Science+Business Media New York 2013

Light Sources 771
High Signal-to-Noise Ratio (SNR) and Photon Counting Detectors 772
The LiDAR Equation 773
Comparison of LiDAR to Other Forms of Remote Sensing 774
Advantages and Disadvantages of Active Remote Sensing 775
Lidar Versus Radar 775
Satellite LiDAR Applications 777
Geodetic and Geodynamic Applications 777
Laser Altimetry and Topographic Mapping 783
Atmospheric Studies 792
Guidance, Navigation, Control, Inspection 801
Cross-References 804
References 804

Abstract

Light detection and ranging (LiDAR), also known as laser detection and ranging (LaDAR) or optical radar, is an active remote sensing technique which uses electromagnetic energy in the optical range to detect an object (target), determine the distance between the target and the instrument (range), and deduce physical properties of the object based on interaction of the radiation with the target through phenomena such as scattering, absorption, reflection, and fluorescence. LiDAR has many applications in the scientific, engineering, and military fields. LiDAR sensors have been deployed at fixed terrestrial stations, in mobile surface and subsurface vehicles, lighter-than-air crafts, fixed and rotary wing aircraft, satellites, interplanetary probes, and planetary landers and rovers. This chapter provides a high-level overview of the principles of operation of LiDAR technology and its main applications performed from space-based platforms such as satellite altimetry, atmospheric profiling, and on-orbit imaging and ranging.

Keywords

Active remote sensing • Atmospheric • CALIOP • CALIPSO • DIAL • Differential absorption LiDAR • Doppler LiDAR • Fluorescence LiDAR • GLAS • ICESat • International Laser Ranging Service • Ladar • Laser altimeter • Laser detection and ranging • Laser remote sensing • LiDAR • Light detection and ranging • LLR • Lunar laser ranging • OBSS • Optical radar • Raman LiDAR • Satellite laser ranging • Scattering LiDAR • SLR

Introduction

This chapter provides a description of light detection and ranging (LiDAR) as an active remote sensing technique. LiDAR has evolved over the past seven decades and as a result there are many different types of LiDAR systems in use today. Systems can be classified based on the application (atmospheric, mapping, bathymetry,

navigation), based on the ranging technique (time of flight, triangulation, phase difference), based on the target detection principle (scattering, fluorescence, reflection), or even based on the platform that the system is deployed on (ground-based, mobile terrestrial, airborne, spaceborne, marine, submarine). There are many reference works that cover LiDAR systems from alternative viewpoints. For example, Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere (Weitkamp 2005) provides an indepth review of modern atmospheric LiDAR techniques, while Topographic Laser Ranging and Scanning: Principles and Processing (Shan and Toth 2009) provides a complete review of the main terrestrial mapping LiDAR techniques. In the context of a Handbook of Satellite Applications, this chapter provides a high-level overview of LiDAR systems with a focus in those based on spaceborne platforms and their main applications. The chapter starts with a brief historical timeline of the origins of the LiDAR technique; it is followed by a high-level technical overview of the principles of operation and the hardware that constitute a generic LiDAR system; and it concludes with descriptions of the main applications of LiDAR technology to and from spaceborne platforms.

Origins of LiDAR Technology

What we know today as LiDAR is the result of the convergence of efforts by different scientific communities to use visible light sources and detectors to resolve technical or scientific issues. LiDAR was pioneered by atmospheric scientists in the 1930s for the determination of atmospheric density profiles, refined as a way to obtain precise and accurate measurements of distances by geodesists and surveyors in the 1940s and 1950s, and taken to interplanetary distances by physicists studying relativistic effects in the 1960s.

Early proposals for the use of high-power search lights to study atmospheric density and composition were developed by E. G. Synge in 1930 (Synge 1930) and M. A. Tuve et al. in 1935 (Tuve et al. 1935). Early successful measurements using bistatic systems consisting of a high-intensity searchlight and a telescopic photographic station separated by baselines of 2–18 km were conducted by J. Duclaux in 1936 (Hulburt 1937), E.O. Hulburt in 1937 (Hulburt 1937), and E.A. Johnson et al. in 1939 (Johnson et al. 1939). Using long-exposure photography, the setup by Duclaux was able to trace light scattering up to a height of 3.4 km, and the experiments by Hulburt reached heights of up to 28–30 km (Hulburt 1937). The limit of these photographic techniques was set by the saturation of the photographic film and the contrast between the beam intensity and night sky. An alternative to the saturation of the photographic film was the method proposed by Tuve et al. and implemented for the first time by Johnson et al. which consisted of modulating the intensity of the search light and using a photoelectric cell to detect the scattered radiation. The output of the photoelectric cell was amplified by an AC system tuned to the lamp modulating frequency. With this type of electric detection system Johnson et al. were able to record light scattering to heights of 34 km (Johnson et al. 1939). These early

atmospheric LiDAR experiments yielded scattering intensity information as a function of the height, but were not concerned with obtaining accurate range measurements. The need to obtain accurate range (distance) measurements using light beams came from the geodetic science community.

LiDAR as a tool to determine accurate range (distance) measurements for geodetic and surveying applications originated in the late 1930s as a technique named electronic distance measurement or EDM. The development of the first EDM instrument began in 1938 when the physicist and geodesist Erik Bergstrand, of the Swedish Geographical Survey Office, began to investigate the possibilities of using a Kerr cell as an electro-optical shutter to modulate a beam of light in an attempt to better measure the speed of light. Bergstrand's first operational instrument was reported to work in 1941 (Carter 1973). In August 1948, Bergstrand presented a paper at the meeting of the International Association of Geodesy (IAG) held in Oslo, Norway. In that paper he explained that the process could be reversed, that by measuring the light's time of flight, and using the known speed of light, it was possible to accurately compute the distance between the light source and a retroreflector. Soon after that IAG meeting, Bergstrand licensed the distance measuring concept to the Swedish AGA (Svenska Aktiebolaget Gasaccumulator) company to develop a commercial EDM instrument. AGA produced the first EDM instrument in the early 1950s, and marketed it as the Geodimeter, short for geodetic distance meter. The instrument used a Kerr cell to modulate the light and a mercury vapor lamp as the light source. Refinement of the Geodimeter by AGA continued through the 1950s and 1960s (Fernandez-Diaz 2007).

During the 1940s and 1950s while Bergstrand was developing the EDM technique, atmospheric scientists continued to build upon the early scattering measurements by using pulsed search lights. These pulsed light sources enabled the researchers to measure the range to the scattering particles using the time-of-flight principle rather than the original triangulation method. In the book *Meteorological Instruments*, published by W.E.K. Middleton and A.F. Spilhaus in 1953, the acronym LiDAR was coined for this type of time-of-flight technique (Wandinger 2005). Around the same time a group at Princeton University led by professor R.H. Dickey, working on gravitation research, investigated a concept of using a high-density and high-altitude artificial satellite to measure slow changes in the universal gravitation constant (G) by tracking the satellite orbit using retroreflectors and pulsed search lights (Bender et al. 1973). This concept incorporated elements of both the atmospheric and geodetic LiDAR research. However, the pulsed light sources and photodetectors available at that time made its implementation impractical. A breakthrough in technology was needed which increased the power and intensity of the light beams.

The breakthrough came in November 1957, when Gordon Gould, a graduate student at Columbia University, coined the acronym LASER, for Light Amplification by Stimulated Emission of Radiation, and described the principal components of the laser (Taylor 2000). The conceptual invention of the laser was followed by the first successful implementation by Theodore Maiman and his colleagues at Hughes Aircraft Company, who built the first solid-state pulsed LASER using

a ruby rod in 1960. That same year, Ali Javan and his colleagues from Bell Laboratories succeeded in building the first gas (HeNe) laser (Javan et al. 1961). Another important advancement was the development of Q-switching for ruby lasers in 1961 by F.J. McClung and R.W. Hellwarth, which enabled the generation of short (nanoseconds) laser pulses that packed relatively large amounts of energy (McClung and Hellwarth 1962). The photons produced by a laser are from a very narrow waveband, have very similar phase and polarization, and travel nearly parallel to one another. These attributes make it relatively simple to create a highly collimated beam of light (its divergence is essentially limited by the aperture of the transmitter and the atmosphere) that yield strong returns from even very distant targets.

In May 1962, L.D. Smullin and G. Fiocco were successful in obtaining ruby laser returns from the bare lunar surface (Smullin and Fiocco 1962), and between June and July 1963 obtained atmospheric returns from heights between 60 and 140 km (Fiocco and Smullin 1963). These experiments ignited an exponential development in LiDAR technology in these fields of research. Within the following decade atmospheric scientists had demonstrated all the basic atmospheric LiDAR techniques in use today (Wandinger 2005).

The physicists and geodesists working on relativity and gravitation obtained the first ruby laser returns from an artificial satellite (Beacon Explorer 22-B) equipped with corner cube reflectors (retroreflectors) on October 31, 1964 (Carter 1973; McGarry and Zagwodzki 2005). This became the origin of what is currently known as satellite laser ranging or SLR, which uses LiDAR to measure ranges from ground stations to satellite-borne retroreflectors with millimeter-level precision, and from which it is possible to obtain highly accurate orbits for critical satellites such as GPS, Glonass, Galileo, Jason, ERS, and others (The International Laser Ranging Service). However, even before the Beacon Explorer was launched, scientists realized that low orbiting satellites imposed several challenges such as very short visibility times and Earth's gravitational perturbations that would limit the quality of the relativistic experiments. To overcome these limitations they had proposed the idea of placing retroreflector arrays on the surface of the Moon, which could be used to bounce back a laser beam shot from the Earth. These lunar retroreflector arrays would allow yield better results than the ones obtained by Smullin and Fiocco in 1962 (Smullin and Fiocco 1962) and by Grasyuk et al. in 1964 (Bender et al. 1973), because they would result in "point" returns, with negligible time spread compared to returns from a patch of lunar topography.

On July 21, 1969, during the Apollo 11 mission, Neil Armstrong oriented and leveled the first lunar retroreflector array (LRRR) on the surface of the Moon. The first successful return signals from the LRRR were obtained on August 1, 1969, at Lick Observatory and on August 20, 1969 at the McDonald Observatory (Alley et al. 1969). Additional retroreflectors arrays were deployed on the Moon by the Apollo 14 and 15 missions, and French-built retroreflectors arrays were deployed by the Soviet Lunokhod 1 and 2 rovers (Dickey et al. 1994). To this date, observatories are still bouncing laser pulses from these retroreflectors in a technique called lunar laser ranging (LLR). This has provided numerous contributions to

a number of scientific fields such as gravitational physics, relativity, astronomy, lunar science, geodesy, and geodynamics (Dickey et al. 1994).

Down on Earth, during the 1960s there was also an exponential development of the EDM technique. In 1967, AGA introduced its Geodimeter Model 8, which was its first to use a helium-neon laser, and doubled the range of the lamp units from 30 to 60 km. Meanwhile, other companies were working on laser-based EDMs with the ability to determine ranges using weak return signals from natural targets rather than from retroreflectors. Examples of these reflectorless EDMs are the instruments manufactured by Spectra Physics such as Mark II and Mark III (Geodolite). These, or similar instruments, were used in the mid-1960s as the first airborne LiDAR profilers and even bathymetric LiDAR systems (Fernandez-Diaz 2007). As lasers with higher pulse rates were developed and scanners of different designs were added to distribute measurements over swaths of terrain, these laser profiling systems evolved into the high-resolution airborne mapping LiDAR systems operational today(Carter et al. 2007).

The first spaceborne LiDAR system was flown onboard the ANNA-1B (Army, Navy, NASA, and Air Force) satellite in 1962, which was a joint project between the agencies to test various satellite tracking techniques including interferometry, Doppler and strobe lights (Simons 1964). ANNA-1B was equipped with two high-intensity optical beacons, that when commanded produced a sequence of five flashes separated by 5.6 s. The flashes were recorded against star fields using stellar cameras (e.g., Wild BC-4 and PC-1000) at ground stations of the Minitrack Optical Tracking System (Harris et al. 1966).

The first spaceborne LiDAR based on a laser transmitter was flown during the Apollo 15 mission in July–August 1971. The Apollo 15 laser altimeter, based on a Q-switched ruby laser was part of the metric camera system, but was also capable of operating independently (Robertson et al. 1972). Similar laser altimeter systems were flown on the Apollo 16 and 17 missions in 1972, and their data were used for, among other things, to determine the lunar shape and infer its structure (Kaula et al. 1974). Between 1972 and the 1990s there was a hiatus in the deployment of spaceborne LiDAR systems, but since 1990 there has been a continuous progression both in terms of numbers and technological development of the deployed systems. Table 31.1 presents a summary of past, current, and future space-based LiDAR systems. Their principles of operation and applications are described in the following sections.

High-Level Technical Overview of LiDAR

In principle LiDAR consists of sending out optical energy, observing the interactions between the photons and the target, and measuring the distance between the emitter and the target. At the highest level a LiDAR system consists of three main subsystems: an optical transmitter, an optical receiver/detector, and ranging/timing/control electronics. The designs of these elements vary greatly among systems and depend upon the targeted application. To help illustrate these concepts,

Table 31.1 Spaceborne LiDAR systems

Launch date	Spacecraft	System/application
October 31, 1962	ANNA-1B	High-intensity optical beacons (Simons 1964; Harris et al. 1966)
July 26, 1971	Apollo 15, Endeavour	Apollo laser altimeter (Robertson et al. 1972; Kaula et al. 1974)
April 16, 1972	Apollo 16, Casper	
December 07, 1972	Apollo 17, America	
1982	PANTHER	LORA/laser altimeter (Werner et al. 1996; Werner et al. 1995)
September 25, 1992	Mars Observer	Mars Orbiter Laser Altimeter (MOLA-1) (Smith et al. 2001; Garvin et al. 1998)
January, 25 1994	Clementine	Clementine laser image detection and ranging system (Smith et al. 1997)
September 9, 1994	STS-64 (Discovery)	Lidar In-space Technology Experiment (LITE) (Winker et al. 1996)
May 20, 1995	Spektr/MIR	BALKAN-1 (Werner et al. 1995)
January 11, 1996	STS-72 (Endeavour)	Shuttle Laser Altimeter 1 (SLA-01) (Garvin et al. 1998)
February 17, 1996	Near Earth Asteroid Rendezvous (NEAR)	NEAR Laser Rangefinder (NLR) (Colea et al. 1996)
April 23, 1996	Priroda/MIR	l'Atmosphere par LIdar Sur SAliout (ALISSA) (Chanin et al. 1999)
November 7, 1996	Mars Global Surveyor	Mars Orbiter Laser Altimeter (MOLA-2) (Smith et al. 2001)
December 4, 1996	Mars Pathfinder/ Sojourner	Microrover Flight Experiment/rover navigation (JPL)
January 1, 1997	ALMAZ-1B	BALKAN-2 (Matvienko et al. 1994)
August 7, 1997	STS-85 (Discovery)	Shuttle Laser Altimeter 2 (SLA-02) (Carabajal et al. 1999)
August 10, 2001	STS-105 (Discovery)	Space Vision Laser Camera System (LCS) (STS-105 Shuttle Press Kit; Piedboeuf et al. 2004)
January 12, 2003	Ice, Cloud, and land Elevation Satellite (ICESat)	Geoscience Laser Altimeter System (GLAS) (Abshire et al. 2005)
August 3, 2004	Mercury Surface, Space Environment, Geochemistry and Ranging (MESSENGER)	Mercury Laser Altimeter (MLA) (Cavanaugh et al. 2007)
April 11, 2005	XSS-11	Spaceborne Scanning Lidar System (SSLS) (Nimelman et al. 2006; Dupuis et al. 2008)/ rendezvous and proximity operations.
July 26, 2005	STS-114 (Discovery)[a]	Orbiter Boom Sensor System (OBSS)[a] (NASA 2005) Laser Dynamic Range Imager (LDRI) (Smithpeter et al. 2000) Laser Camera System (LCS) (Deslauriers et al. 2005)

(*continued*)

Table 31.1 (continued)

Launch date	Spacecraft	System/application
April 28, 2006	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO)	Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) (Winker et al. 2004)
August 4, 2007	Phoenix Mars Lander	Meteorological Station (MET) Atmospheric LiDAR (Whiteway et al. 2008)
June 18, 2009	Lunar Reconnaissance Orbiter (LRO)	Lunar Orbiter Laser Altimeter (LOLA) (Ramos-Izquierdo et al. 2009)
August 28, 2009	STS-128 (Discovery)	TriDAR[b] (English et al. 2005; NEPTEC)
~2013	ADM-Aeolus	Atmospheric Laser Doppler Instrument (ALADIN) (Ansmann et al. 2007)
~2014	BepiColombo	BepiColombo Laser Altimeter (BELA) (Thomasa et al. 2007)
~2016	ICESat-II	multi-beam laser altimeter (Abdalati et al. 2010; Yua et al. 2010)
TBD	Earth Clouds, Aerosols and Radiation Explorer (EarthCARE)	Atmospheric backscattering and depolarization LiDAR (ATLID) (Le Hors et al. 2008)
TBD	Deformation, Ecosystem Structure and Dynamics of Ice (DESDynI) (Donnellan et al. 2008)	L-band polarimetric InSAR multi-beam laser altimeter
Future NASA	ASCENDS, ACE, LIST, GRACE II, 3D-Winds	These missions have been proposed by the NRC decadal survey (National Research Council 2007) and might include LiDAR instruments.
Future ESA	WALES, ASCOPE (Durand et al. 2007)	These are Earth observation missions under study by ESA that might include LiDAR.

[a]The OBSS made its first flight on STS-114 Discovery and has flown on every shuttle mission since
[b]TriDAR had its first space demonstration on STS-128 and was flown again during STS-131 (April 2010)

Figs. 31.1 and 31.2 show a 3D model and optical diagram of the atmospheric scattering LiDAR (CALIOP) onboard the NASA/CNES CALIPSO satellite.

The optical transmitter is composed of a light source, usually a laser system, and optical elements used to modify (focus, collimate, expand, split) the light beam. The optical detector is comprised of a telescopic-type instrument that collects the backscattered photons, spatial and spectral filters that discriminate the specific wavelengths intended to be detected, and an electronic photodetector that can be a simple photomultiplier or photodiode in the case of the mapping LiDAR or as elaborate as an spectrometer in the case of fluorescence or Doppler LiDAR. If the transmitter and the detector systems share the same optical elements, i.e., same optical transmit and receive paths, the system is considered to be monostatic. If the optical transmit and receive paths do not share elements, the system is defined as bistatic. From Figs. 31.1 and 31.2 it can be seen that CALIOP is a bistatic system, with a transmitter consisting

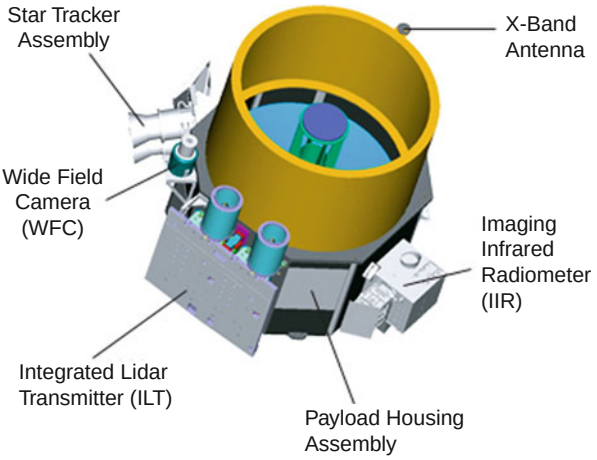


Fig. 31.1 3D optical model of the CALIOP LiDAR (Image courtesy of NASA)

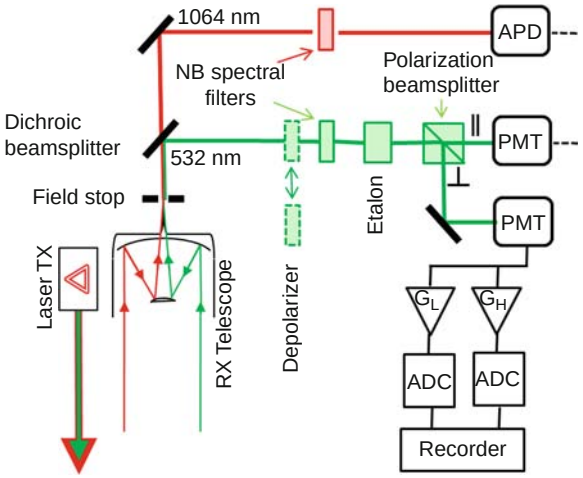


Fig. 31.2 Optical diagram of the CALIOP LiDAR

of two independent lasers located parallel to the receiving telescope. Finally, the ranging/timing electronics enable the LiDAR to determine the distance to the target. In addition, LiDAR systems very often have mechanical, optical, or electronic scanning mechanisms that allow steering the light beam.

The design of a LiDAR system starts with the definition of the purpose or application that the system will serve. The application will dictate which interaction between light and target needs to be detected (scattering, reflection, absorption, etc.), and the most suitable ranging method. The type of interaction between light and target dictates what particular wavelengths can be used and narrows down the light sources that can be selected. From this point, it remains to select the best available photodetector to sense that light–target interaction. To aid the design

process the LiDAR equation is used, which relates the expected received signal strength with sensor parameters such as transmitted optical energy and receiver telescope area, atmospheric parameters such as transmittance and scattering probability at the operating wavelength, and operating conditions such as expected range and target cross section. The following sections provide basic descriptions of the ranging methods, the light–target interaction phenomena, and the light sources and photodetectors that enable the operation of a system. These descriptions cover material that lead to different forms of the LiDAR equation.

Ranging Methods

There are three main methods that can be used to measure the distance (range) between a LiDAR instrument and the target: optical triangulation, phase difference, and time of flight (TOF). It is also possible to employ hybrid approaches combining two of these methods. Each of these ranging methods has its own set of strengths and weaknesses and range of applicability (English et al. 2005).

Optical Triangulation

Optical triangulation was the ranging method used in the early atmospheric LiDAR experiments of the 1930s. As illustrated in Fig. 31.3, it is based on the geometry principle that knowing three elements of a triangle it is possible to determine any other element of the triangle. In the case of the early atmospheric LiDARs the first element that was known was the separation between the search lights and the observing station. This leg of the triangle is known as the baseline. The other two known elements of the triangle were the horizontal angles of the search light and photographic station.

Systems based on optical triangulation are ideal for short-range measurements (few meters) yielding micrometer-level precision at high data rates. However, its accuracy depends on the relation between range and baseline distance, and it degrades rapidly with increasing range ($\sim R^2$). It is also limited due to its sensitivity to noise from exterior illumination sources (English et al. 2005).

Phase Difference

Phase difference was the ranging method used in early geodetic EDMs such as the Geodimeter, and it is currently used in some ground-based and airborne mapping systems and on one short-range spaceborne imager. This method consists on modulating the intensity of a continuous wave (CW) LASER using a superposition of sinusoidal waveforms with different spatial wavelengths. The range is determined by measuring the phase difference and the number of complete cycles between the emitted and return laser waveform. The main disadvantage with this

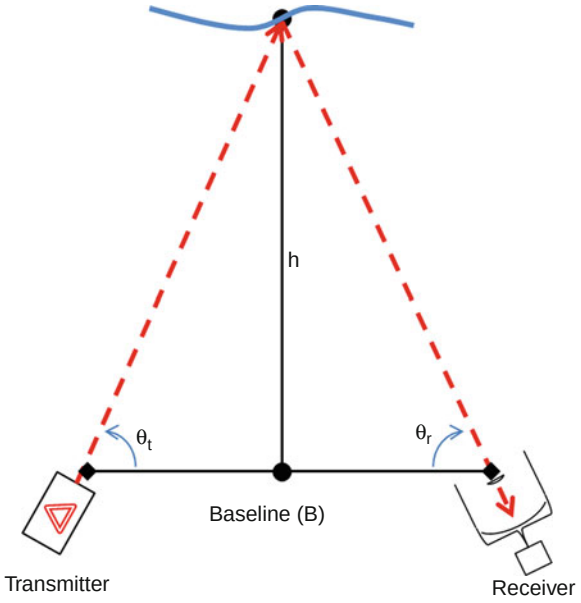


Fig. 31.3 Triangulation ranging principle

method is that phase differences are not unique, as there is always an unknown number of complete modulating wave cycles that have occurred prior to the phase difference (phase ambiguity). Compared to the time-of-flight method, the phase difference methods provides higher measurements rates. If there is no a priori knowledge of the range (for geodetic systems), the maximum range of this method is half the spatial wavelength of the carrier frequency, and the range resolution is a function of the highest modulating frequency and the phase difference resolution (English et al. 2005).

Time of Flight (TOF)

The third ranging method uses discrete pulses of light rather than continuous emitting sources. The TOF principle is the simplest and it consists of measuring the time between when the light pulse is emitted and the detection of a return signal. This two-way travel time (time of flight) is divided in half and multiplied by the speed of light in the respective medium, yielding the range between the instrument and the target. Early LiDARs that used light from lamp sources would create light pulses using optical chopper wheels or capacitive discharge devices (flash lamps). The development of Q-switching by McClung and Hellwarth in 1961 enabled the emission of very energetic LASER pulses rather than the continuous wave beams. However, even when these pulses last for a relatively short time, generally in the order of a few nanoseconds, at the high speed that light travels this translates into several centimeters in length (e.g., 1 ns = 30 cm). In order to obtain sub-centimeter

accuracy, the recording and analysis of the entire emitted and return waveform must be performed, or a specialized electronic circuit called a constant fraction discriminator (CFD) can be used on the fly to precisely time a specific point on the waveform (generally the half point of the pulse amplitude at its leading edge). Systems that range to special design retroreflectors may use mode-locked lasers which produce very narrow pulses picoseconds in width.

TOF is the most common ranging method in modern LiDAR, because it provides unambiguous range measurements of distances limited only by the dispersion of the LASER energy and the sensitivity of the detector. However, the TOF approach is limited in data collection rate by the laser repetition frequency (PRF) which is the number of laser pulses that can be emitted per second.

Hybrid Systems

Hybrid systems use two above ranging methods, combining the unique capabilities of each to overcome the limitations of a single method. For instance, a hybrid system that employs the triangulation and TOF methods can exploit the advantages of TOF for long ranges and the accuracy and speed of a triangulation system at short ranges (English et al. 2005).

Light–Target Interaction Phenomena

Recall that the "D" in LiDAR stands for detection, the detection of return optical energy backscattered from the target. Detection of a target is possible because there is an interaction between the emitted light energy and the target. There are several types of interactions, which usually depend on the relative size of the target and the wavelength of the radiation. The main interactions between light and matter employed by LiDAR technology are described next.

Scattering

Scattering is the physical phenomenon that occurs when electromagnetic radiation changes its original direction of travel due to interactions with matter in the form of atoms or molecules (Fig. 31.4). If there is only one particle, a single scattering process is produced. If the photon is scattered several times by different particles, the process is called multiple scattering. These matter and radiation interactions can occur with or without the apparent transfer of energy. In elastic scattering the photons maintain their wavelength, thus conserving energy. Examples of elastic scattering include Rayleigh and Mie scattering. Inelastic scattering occurs when part of the photon energy is transferred into the scattering particle, thus changing its wavelength. Examples of inelastic scattering include Raman and Brillouin scattering. Based on the relative size of the scattering centers with

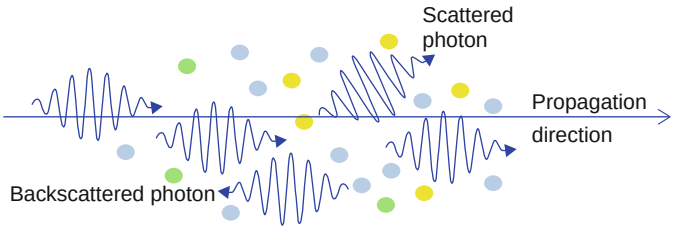


Fig. 31.4 Photon and matter interaction – scattering

respect to the wavelength of the radiation, scattering can be classified as: Rayleigh scattering when the particles are small compared to wavelength, Mie scattering when the particles size and radiation wavelength are roughly of the same order of magnitude, and geometric scattering when the particles are much larger than the wavelength.

The backscatter component is the radiation that changes direction by approximately 180°, i.e., reverses direction (Fig. 31.4). Radars and LiDARs detect the backscatter component of the radiation that was emitted. In atmospheric LiDARs, Mie scattering is used to detect aerosols in the troposphere while Rayleigh scattering is used to detect molecules in the stratosphere and mesosphere. Mapping LiDARs are based on geometric scattering as the targets are much larger than the optical wavelengths.

Reflection

Reflection is a particular type of geometric scattering following particular geometric relationships. There are two limiting theoretical models for reflective surfaces: a specular reflector is one from which incident radiation will be reflected in a single direction (like a mirror) following Snell's law, and a Lambertian reflector surface will spread the reflection over a wider pattern (Fig. 31.5). These are two limiting cases, and the actual reflection from most surfaces will be between these models. Mapping LiDAR detects reflected radiation from varied targets such as the solid rough surface of a planet (Lambertian behavior), diffuse targets like a forest canopy, or mirror-like surfaces such as a calm lake (specular behavior). An example of LiDARs that are based on specular reflections are those systems used for satellite or lunar laser ranging (SLR and LLR). To achieve extremely long ranges and millimeter-level accuracy, corner cube reflectors (retroreflectors) are used to reflect the laser beam in almost exactly the opposite direction (within a few seconds of arc) in which it was emitted.

Absorption

Absorption is another possible result of the interaction of electromagnetic radiation and matter. For a photon to be absorbed it has to be of a particular wavelength or

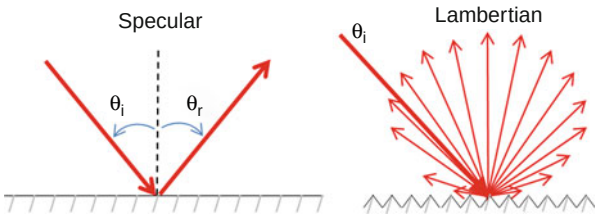


Fig. 31.5 Specular and Lambertian reflection patterns

energy, and because of the principle of conservation of energy the absorption causes a change in the energy state of the atom or molecule by either an electronic, vibrational, or rotational transition. Differential Absorption LiDAR (DIAL) systems compare the received backscattered signal for two or more different laser wavelengths to determine the differential molecular absorption coefficient. If the differential absorption cross sections for each wavelength are known, the concentration of the gas atoms or molecules can be directly deduced. Atmospheric constituents that can be detected by DIAL include ozone and water vapor. DIAL can also be used for industrial emission monitoring and forest fire detection.

Fluorescence

Fluorescence occurs when a molecule absorbs a photon and after a determined period of time emits another photon of the same or longer wavelength. It is considered resonance fluorescence when the emitted photons have the same wavelength of the absorbed photon, and normal fluorescence when the emitted photons have longer wavelengths (lower energy). The process of normal fluorescence occurs in three stages: excitation of the molecule by the incoming photon which happens on a timescale of femtoseconds (10^{-15} s), vibration relaxation which brings the molecule to a lower excited state and occurs on a timescale of picoseconds order (10^{-12} s), and the emission of a longer wavelength photon and return of the molecule to the ground state which occurs in a relatively long time period of nanoseconds (10^{-9} s). Fluorescence LiDAR usually emit ultraviolet radiation and observe the reemission of photons in the visible range with a spectrometer detector which records the relative emission at different wavelengths. Applications of fluorescence LiDAR include vegetation studies and the detection of pollutants. For instance, minute amounts of oil in water can be detected because of the UV fluorescence properties of hydrocarbons.

Doppler

The Doppler effect consists of an apparent shift in frequency or wavelength of waves (sound or electromagnetic) as a result of the relative motion between the emitter and the observer. These relative motions can be due to movement of emitter, observer, or medium (in the case of sound waves), or even the simultaneous motion of all three of

them. If the relative motion makes the emitter and observer become closer, the wavelength of the wave will appear to get shorter (blue shift), whereas if the distance becomes larger the wavelength will appear to get longer (red shift). In addition to the well-known frequency shift, the Doppler effect also causes the broadening of spectral line features in a process that is temperature dependent. Turbulence and winds are manifestations of the collective motion of the atmospheric molecules and particles. Light scattered along the line of sight (LOS) of the propagating laser beam will experience Doppler shifts and linewidth broadening due to the relative motion of the atmospheric elements with respect to the LiDAR system and due to changes in atmospheric temperature. Thus, Doppler LiDAR is applied to determine air temperature, wind speeds, and directions. The Doppler shifts are proportional to the ratio of wind speed and the speed of light as:

$\Delta\lambda = -\lambda_0 \frac{c}{v \times \cos(\theta)}$, where $\Delta\lambda$ is the wavelength shift, λ_0 is the reference or emitted wavelength, c is the speed of light, and $v \times \cos(\theta)$ is the wind speed component along the LOS. The spectral linewidth broadening is given by:

$\sigma_\lambda = \frac{1}{\lambda_0}\sqrt{\frac{k_B T}{m}}$, where k_B is the Boltzmann constant, T is the particle temperature, and m is the particle mass.

There are two main ways for measuring the Doppler shift and linewidth broadening using LiDAR: direct detection and coherent (heterodyne) detection (Wandinger 2005). In direct detection Doppler LiDAR the wavelength shift is determined by a spectrometer instrument which employs narrowband spectral filters and measures the backscattered radiation at each band. Coherent Doppler LiDAR is based on the emission of modulated pulses of single-mode single-frequency laser radiation. The detected backscattered signal is mixed with the signal of a local oscillator and by detecting the beat frequency the frequency shift is determined. To determine the sign of the shift, a frequency offset is introduced between the emitted pulse and the local oscillator.

Depolarization

Depolarization is not a LiDAR detection technique per se, however, because the polarization of the laser radiation emitted by a LiDAR is well known, it is possible to measure how much radiation is backscattered with the same polarization and at a perpendicular polarization. In atmospheric LiDARs, depolarization provides information about the nature of the scattering particles, as Mie scattering theory indicates that depolarization is caused by nonspherical scatterers. In mapping LiDAR depolarization can be used to characterize surface roughness.

Light Sources

A light source is a basic part of a LiDAR system. During the early days of LiDAR experimentation the light sources were mercury or sodium vapor lamps. Currently

the light source will most likely be a laser. Laser is an acronym for Light Amplification by Stimulated Emission of Radiation. Traditional lasers consist of an optical resonator which contains an optical gain medium. This gain medium, or lasing material, is pumped with optical or electrical energy (semiconductor lasers) causing the electrons in the lasing material to be excited to a higher nonequilibrium level, and stimulated emission occurs when an interacting photon causes an electron to drop from the higher level to its ground state releasing an additional photon at the same wavelength as the interacting photon. If this stimulated emission builds up within the optical resonator to a point where the gain of the process overcomes the cavity losses at a given resonant mode then lasing is achieved and a relatively high coherent beam of light will be emitted. Coherence refers to the laser beam's spatial and spectral characteristics; a perfectly coherent laser beam will travel in a single direction (spatial coherence), and the photons would be of a single wavelength, polarization, and phase (spectral coherence). In the real world lasers are not 100% coherent, but can emit light from several modes at different wavelengths at the same time with not necessarily the same polarization, and their beam can diverge beyond the diffraction limit. However, most lasers used in LiDARs are built to be single mode and diffraction limited. Besides the traditional electronic population inversion lasing method, it is possible to generate laser light through other processes such as relativistic free electron beams and by modifying the vibrational and rotational modes of oscillation of molecules. Lasers can produce light not only in the visible spectrum, but also in other regions of the spectrum including the infrared, the ultraviolet, and the X-ray regions.

Lasers can be classified based on the lasing medium as solid state, liquid, and gas lasers. Examples of solid state lasers include those based on crystalline paramagnetic ions, glass, solid dyes, semiconductors, polymers, and excimers. Liquid lasers can be based on organic dyes, rare earth liquids, polymers, and excimers. Gas lasers include neutral atoms, ionized gases, and molecular gases (Weber 2001). One of the most common lasers used in LiDAR technology is based on the solid state crystal: neodymium-doped yttrium aluminum garnet (Nd:YAG), which lases at 1,064 nm.

Based on their modes of operation lasers can be classified into continuous wave lasers if its output power is constant over time (although the intensity of the beam can be modulated) and pulsed lasers if the optical energy is released in sudden bursts. Laser pulses packing a relatively high amount of energy, compared to continuous operation, can be obtained through the Q-switching technique. Pulses obtained through Q-switching are typically in the range of hundreds of picoseconds to tens of nanoseconds in length. Extremely short pulses in the picosecond to the femtosecond range containing very little energy can be created using the mode-locking technique.

High Signal-to-Noise Ratio (SNR) and Photon Counting Detectors

The optical backscattered signal resulting from the interaction between the radiation and the target needs to be detected by the LiDAR system. For this purpose many

different type of photodetectors can be employed. These photodetectors include PN, PIN, and avalanche photodiodes and photomultiplier tubes. The selection of the photodetector is a crucial aspect in the design of a LiDAR system (Kaufmann 2005), and factors that must be taken into account in this process are: the wavelength and the magnitude (signal strength) and magnitude range (dynamic range) of the radiation to be detected, and the speed at which it needs to be detected. Generic characteristics of photodetectors include its wavelength band of operation (spectral response), its sensitivity (how much electric signal is produced per unit of detected radiation), its noise characteristics (how much electric signal is produced even when no radiation is incident on the detector), response speed (ability to detect distinct events separated by short times), active area, number of elements (single element vs. array of detecting elements), and its operating voltage and power consumption.

Independent of the type of photon detector used there are two main modes of operation depending on the magnitude of the detected signal: high signal-to-noise ratio (SNR) or analog detection and low SNR, also called photon counting or digital detection (Hamamatsu Corporation). In high SNR LiDAR systems, the magnitude of the detected signal is many times larger than the general background noise, including scattered solar radiation and artificial lighting, and the detector thermal noise. High SNR is typical of short-range, high-power systems such as mapping and elastic backscattering LiDARs. In the low SNR domain the magnitude of the detected signal is very close to the noise level and in some cases the detector responds to the excitation of single photon events, which is why it is also called photon counting. Photon counting is used in extremely long-range systems such as SLR and LLR, for systems where the interaction between the radiation and matter is particularly weak such as in Raman LiDAR or high atmosphere Rayleigh scattering and resonant fluorescence LiDARs (Abshire et al. 2005; Whiteway et al. 2008), water penetrating (bathymetric) LiDAR, and low-power multichannel systems (Cossio et al. 2010).

The LiDAR Equation

The LiDAR equation is a mathematical formulation that provides an estimate of the received optical signal strength by a system as a function of instrument parameters, atmospheric phenomena, and detection range. The LiDAR equation is used to design systems, to evaluate the performance of existing systems and it is inverted to determine atmospheric properties from real observations. There are many versions of the equation depending on the type of system it describes. In its most generic form it is (Wandinger 2005):

$$P_r(R) = K_s \times G(R) \times T^2(R) \times \beta(R)$$

where $P_r(R)$ is the received power as a function of the range, K_s is a constant factor dependent upon system parameters such as transmit power and optical efficiency, $G(R)$ is a factor that depends on the geometry of the observation as function of the range,

$T(R)$ is the propagation medium transmission factor, and $\beta(R)$ is a factor that describes the target backscattering properties. Each of these factors can be expanded and/or adjusted to account for the specifics of each system and application.

For instance, the LiDAR equation for elastic backscattering atmospheric LiDAR, where the targets are atmospheric constituents (atoms or molecules) can be expanded as (Wandinger 2005):

$$P_r(R) = \left[\frac{P_0 \eta c \tau A}{2}\right] \times \left[\frac{O(R)}{R^2}\right] \times \left[e^{-\int_0^r \alpha \times dR}\right]^2 \times \beta(R)$$

where P_0 is the emitted laser power (pulse energy/pulse length), η is the optical efficiency of the system, c is the speed of light in the transmission medium, τ is the laser pulse width, A is the receiving telescope area, $O(R)$ is the fractional overlap area collected by the receiver, and α is the extinction coefficient. In this case both the atmospheric transmission and scattering coefficient are the properties under study. The scattering coefficient indicates the probability that a photon will be backscattered. The atmospheric transmittance is the exponential integration of the extinction coefficient which is proportional to the amount of scattering material in the atmosphere, it can also be considered as the effective cross-sectional area of particulates per unit volume. The combined expression $c\tau A$ is considered the scattering volume, which when multiplied by the scattering coefficient $\beta(R)$ yields the scattering cross section.

For an altimetry or mapping LiDAR the equation can be expanded as (Bufton 1989):

$$P_r(R) = [P_0 \eta A] \times \left[\frac{1}{R^2}\right] \times \left[e^{-\int_0^r \alpha \times dR}\right]^2 \times \left[\frac{\rho}{\Omega}\right]$$

where ρ/Ω is the target backscatter or reflectance per solid angle.

These equations can be expanded even further to account for each interaction that affects the laser beam along its two-way travel from the transmitter to the receiver, and as stated before need to be adjusted for the particular type of LiDAR system and application.

Comparison of LiDAR to Other Forms of Remote Sensing

Having described LiDAR technology and principles of operation, it is convenient to compare this active optical detection technique against other forms of remote sensing. It is important to remember that every remote sensing technique has its strengths and limitations, and it is crucial to understand the relative advantages and intrinsic limitations of different techniques to determine which is the most appropriate for a given application. The next two sections compare active versus passive remote sensing techniques, and LiDAR versus radar.

Advantages and Disadvantages of Active Remote Sensing

Having control of the illumination source creates several advantages for active remote sensing (LiDAR and Radar) over passive techniques. The first advantage is that active systems are independent of day/night conditions. This is particularly true for Radar systems. However, certain types of LIDAR units work better under night conditions, and some can only work at night. Long-wavelength radars (>10 cm) are also independent of weather conditions and can work through clouds and rain.

With passive remote sensing techniques such as multispectral and hyperspectral imaging most of what can be inferred from the target has to do with the amplitude of the detected signal (relative or absolute reflectance). With active systems there is full knowledge and sometimes control of the parameters of the illumination signal: amplitude, frequency, phase, and polarization. This control allows researchers to study the effect that the target has on all the parameters of the emitted radiation enabling a more complete characterization of the target. The use of phase information makes it possible to accurately measure sub-wavelength scale changes in ranges, which is applied in deformation mapping using InSAR or millimeter-level ranging with LiDAR. Measuring the change in polarization (depolarization) enables the geometric characterization of the target, it is used in atmospheric LiDAR to determine if the scatterers are spherical or not, and in polarimetric SAR to determine the orientation and location of the scattering sources.

Finally, measurements of perceived changes in frequency or wavelength allow the use of Doppler techniques to determine the relative speed of the target moving along the line of sight (LoS) of the LiDAR or Radar. A parameter of the illuminating signal for which there is almost full control is the power (limited by the maximum power output of the source), which can generally be adjusted to a level that optimizes the signal-to-noise ratio (SNR) of the detected return, thereby reducing the sensitivity to background and detector noise compared to passive remote sensing techniques.

Despite the many advantages of the active remote sensing technique there are some disadvantages with respect to the passive techniques. The main disadvantage is that active sources can only sample relatively small areas at a given time, and to increase the spatial resolution it is often necessary to reduce the extent of the study area. An additional disadvantage is that active sensors provide very little spectral information, limited to a few wavelengths compared to the hundreds of channels that can be studied with a hyperspectral system.

Lidar Versus Radar

To compare LiDAR and Radar remote sensing a good starting point is their respective operational wavelengths. Most operational radars work in the wavelengths between 2 and 30 cm (10^{-2} m), while LiDARs operate between 300 and 2,000 nm (10^{-9} m). On average this is a 5 order of magnitude difference, and this has many implications for remote sensing applications. The first implication has to

do with the interaction between radiation and matter. As explained earlier, scattering is a process determined by the relative size of the particles and the wavelength. In the case of atmospheric constituents, their size is comparable to the wavelengths in the optical range, and this is why it is possible to study atmospheric scattering with LiDAR. It is also possible to measure Doppler shifts and broadening from optical radiation scattered by moving atmospheric particles, which in turns allows for the remote determination of wind velocities and temperature profiles using LiDAR. The Radar wavelengths, on the other hand, are much larger than atmospheric particles, and are not affected by atmospheric atomic and molecular constituents. However, low-wavelength ($<$10 cm) Doppler RADAR is sensitive to much larger water drops and ice crystals.

Besides the scattering interaction there is also the possibility of absorption, atmospheric extinction which is the depletion of transmitted radiation, caused by the combination of scattering and emission. Atmospheric transmission is complementary to extinction. The Earth's atmosphere is practically transparent to radio waves, but it is relatively opaque in certain optical bands. This is of crucial importance for remote sensing applications from satellite platforms for which the electromagnetic radiation to be detected needs to travel through the Earth's atmosphere. Therefore the bands of operation of spaceborne sensors are selected taking into consideration the transparency of the atmosphere. The atmosphere's transparency in the radio wavelengths allows Radar to operate under most weather conditions, which combined with its day and night operability provides a significant advantage over other forms of remote sensing. However, absorption is not entirely an undesirable phenomenon. Absorption at specific wavelengths due to atmospheric molecules is the principle used by differential absorption LiDAR (DIAL) to detect and measure the concentrations of molecules such as ozone and water vapor in the atmosphere.

A final aspect to consider in the comparison between radar and LiDAR is the divergence or spread of a Radar or laser beam. The divergence also relates to the angular resolution of a remote sensing system. Divergence is determined by diffraction at the output aperture from which optical or radio energy is emitted. The Rayleigh criterion provides an estimate of the angular resolution of optical imaging systems or the beam divergence of active systems as:

$$Sin(\theta) = 1.220\frac{\lambda}{D}$$

where θ is the angular resolution or beam divergence in radians, λ is the radiation's wavelength, and D is the diameter of the aperture (lens or antenna). Considering an average optical wavelength of 1,064 nm and a modest aperture of 1 cm, the diffraction-limited divergence of a laser beam is then 0.13 μrad. For a radio wave at an average wavelength of 10 cm and with an antenna 10 m in diameter, the divergence of the radio beam is 12.2 μrad, almost 100 times wider than the laser beam. In order to have the same divergence as the optical beam the antenna would

have to be almost 940 m in diameter. To overcome this limitation the synthetic aperture radar (SAR) technique was developed to electronically synthesize a virtual antenna many times larger than the physical antenna, based on the platform motion. Smaller divergence of laser beams implies smaller footprints and better angular and spatial resolutions for LiDARs as compared to Radar.

The contrast of higher resolution due to smaller footprints is that LiDARs generally provides smaller spatial coverage. In addition, current spaceborne LiDAR systems for atmospheric and mapping applications operate in single beam profiling mode, which means that the sampling is performed along a single line with no scanning capabilities. On the other hand, spaceborne Radar systems have multiple beams and the capability to electronically steer the beams in a direction perpendicular to the direction of flight. Larger footprint and scanning capabilities of radar systems allow for larger spatial coverage and a better temporal resolution.

Satellite LiDAR Applications

Geodetic and Geodynamic Applications

Geodesy is the study of the shape, size, orientation, motion, and gravity of the Earth; it also includes the establishment of coordinate reference systems used to uniquely describe the location of any point on the Earth. Geodesy is the discipline that enables many of current satellite applications such as satellite aided navigation (GPS, GLONASS, and Galileo) and satellite remote sensing mapping by establishing the geodetic frame of reference on which these systems operate.

The first geodetic observation is credited to Eratosthenes, a Greek philosopher who lived in the third century BC, and who was able to conclude that the Earth had a spheroid shape and was able to estimate its size. Over the centuries, geodetic instruments and techniques have evolved, but the need to measure angles, distances, and time to determine geographic coordinates and the Earth's parameters has not changed. This need for accurate distance and time measurements led geodesists to develop the electronic distance measurement (EDM) technologies, one of which evolved into modern-day ranging LiDAR. Also, for centuries, geodesists have been performing astronomical observations to derive coordinates and distances between remote stations. They realized that this could also be done by observing man-made airborne objects, and so as technology matured, they started using balloons, airplanes, rockets, and eventually satellites as targets. So it is not surprising that the first spaceborne application of LiDAR technology was developed for geodetic studies.

This was achieved by leaving the active LiDAR equipment (laser transmitter and optical detector) on the ground (Fig. 31.6) and installing passive elements (retroreflector arrays) on satellites (Fig. 31.7). This architecture has many advantages, the main one being that technology can improve continuously on the ground segment and need not stop once the satellite is integrated and launched. Also the



Fig. 31.6 NASA MOBLAS-7 mobile SLR system circa 1980 (Courtesy of NASA)



Fig. 31.7 Laser retroreflector array on COMPASS satellites (Courtesy of Shanghai Astronomical Observatory, Chinese Academy of Sciences)

spacecraft infrastructure, being passive, does not require power or maintenance and typically has extremely long lifetime. The long lifetimes and large number of satellites carrying retroreflectors have allowed the accumulation of over four decades of ranging data.

The first geodetic satellite tracked by LiDAR was the ANNA-1B launched on October 31, 1962. ANNA-1B carried equipment to test three different satellite tracking techniques; one of them was the use of high-intensity optical beacons (Simons 1964). The beacons operated on command and produced a sequence of five flashes separated by 5.6 s. These flashes were recorded using long-exposure photography, simultaneous observations from different stations allowed the determination of the satellite position (Harris and Berbert 1966). The first geodetic satellite that carried a retroreflector array was the Beacon Explorer-B (designated as the Explorer 22) (Degnan). The Explorer B was launched on October 9, 1964; it was a 116-pound satellite that in addition to the retroreflector also carried a radio beacon. The satellite was tracked from stations around the world using both radio and LiDAR technology, although, radio equipment was much cheaper than the optical Radar, and because the satellite was magnetically stabilized, the retroreflectors were oriented in such way that it was only possible to track the satellite from stations on the northern hemisphere.

The first laser tracking of the Explorer 22 was carried out on October 31, 1964, by a team from NASA's Goddard Space Flight Center (GSFC). This was the origin of a geodetic LiDAR technique named satellite laser ranging (SLR). The Explorer 22 was soon joined by more satellites carrying corner cube retroreflectors including more satellites of the Explorer series, Explorer 22 (launched on April 29, 1965), Explorer 29 also known as GEOS1 (launched in November 6, 1965), and the Explorer 36 or GEOS 2 (launched on January 11, 1968). The Centre National d'Etudes Spatiales (CNES) from France also contributed to SLR by launching a pair of geodetic satellites, Diadème-1 D1C (February 08, 1967) and the Diadème-2 D1D (February 15, 1967), equipped with dual frequency Doppler transmitters and retroreflector arrays. The first international SLR campaign occurred in the spring of 1967 with the participation of five laser stations, three operated by CNES and located in France, Algeria, and Greece, one station operated by NASA in Maryland, and one operated by the Smithsonian Astrophysics Observatory (SAO) in New Mexico. Data from this campaign was used to compare SLR to traditional optical observations and it was estimated an improvement by a factor of 4 in the accuracy of determined positions; however, most important was the development of SAO standard Earth's gravity model (Degnan).

This first international SLR campaign with stations spread across the world helps illustrate the mode of operation of this geodetic LIDAR technique. As shown in Fig. 31.8, a single satellite can be tracked simultaneously from stations separated by a few meters up to thousands of kilometer, and using triangulation it is possible to determine the baselines between the stations. Observations from SLR stations are enhanced by colocation with other global space geodetic techniques such as very long baseline interferometry (VLBI), global navigation satellite systems (GNSS), and Doppler orbitography and radiopositioning integrated by satellite (DORIS).

The early geodetic satellites were not optimal for geodesy and relativistic applications because they were launched into low orbits, and because they carried a variety of instruments which enlarged their cross section and lowered their density. The satellite's low orbit and low density limited the visibility times and



Fig. 31.8 Simultaneous SLR from three stations at the Goddard Geophysical and Astronomical Observatory (Image courtesy of NASA)

increased their susceptibility to gravitational perturbations, while the large cross section made them susceptible to atmospheric drag, radiation pressure, and other nonconservative forces. To overcome these limitations, the ultimate Earth satellite, the Moon, was equipped with retroreflectors. As early as 1962 J.E. Faller had proposed the idea of placing a retroreflector on the surface of the Moon and in 1965 the Lunar Ranging Experiment (LURE) multi-institutional team was formed. Between 1969 and 1973 a total of five retroreflectors were placed on the moon, three of them by manned Apollo missions (11, 14, and 15) and two French-built retroreflectors carried by the Russian lunar rovers Lunokhod 1 and 2 (Luna 17 and 21 mission) (Bender et al. 1973). These lunar retroreflectors made it possible to range to and track the Moon from stations around the world using a LiDAR technique called lunar laser ranging (LLR).

As a complement to the lunar retroreflectors, several satellites designed exclusively for geodesy using SLR have been launched into relatively high and very stable orbits. These "cannon ball" satellites have high densities and small surface area covered almost entirely by retroreflectors. The first was the French-built Starlette launched in 1975, followed by the American Laser Geodynamics Satellite (LAGEOS-1) launched in 1976. Other SLR-only satellites include the Japanese Ajisai (launched in 1986), the Soviet Etalon-1 and 2 (launched in 1989), Lageos-II (built by the Agenzia Spaziale Italiana and launched in 1992), and the French satellite Stella (launched in 1993).

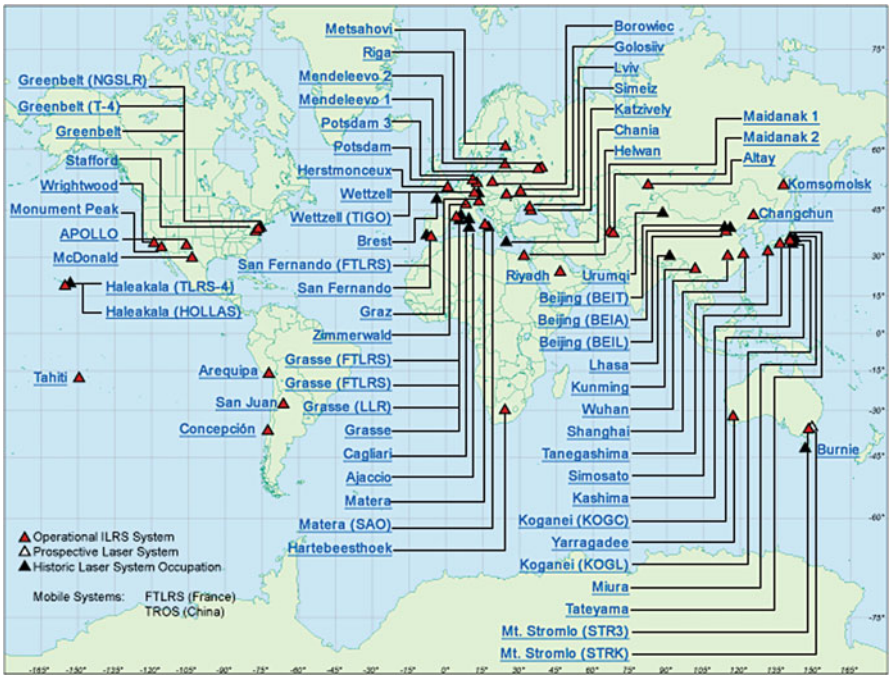


Fig. 31.9 Stations of the international laser ranging service (Courtesy of ILRS/NASA)

To this date more than 130 satellites have been tracked from more than 70 laser stations around the world (Fig. 31.9) (The International Laser Ranging Service). The massive amount of data collected for almost half a century from SLR and LLR has allowed the accurate determination of the ground station coordinates to the millimeter level and the satellite orbits to the centimeter level. These techniques combined with other space geodetic techniques such as VLBI and GNSS have been applied to scientific issues such as the modeling or establishment of the Earth's gravity field, reference frame, and orientation parameters; to prove geodynamic theories such as plate tectonics, glacial rebound, and crustal deformation, to test principles of general relativity; and to the determine Earth-lunar and solar system celestial mechanics parameters (The International Laser Ranging Service; Degnan). Also the establishment of the terrestrial reference frame (TRF) and Earth orientation parameters (EOP) along with the accurate determination of satellite orbits is crucial for satellite applications such as navigation and Earth observation. Some of these applications are described next.

Observations and Modeling of the Terrestrial Gravity Field

The Earth's gravity field is a 3D vector field that specifies the acceleration that an object will experience at a given point at or above the Earth's surface. Its main component or mean gravity, 9.8 m/s^2, is the equivalent gravity of a uniform mass

distribution and a spherical shape. The next order deviation from this simplified model is due to the Earth's rotation and oblate shape. Smaller order variations are due to mass distribution heterogeneity. In addition to spatial variations there are temporary variations due to mass redistribution through and among the atmosphere, cryosphere, hydrosphere, and solid Earth.

To study the gravity field, the gravitational potential is modeled by a spherical harmonic series of the form (Heiskanen and Moritz 1967):

$$U = \frac{GM}{r}\sum_{n=0}^{\infty}\sum_{m=0}^{n}\left(\frac{r_0}{r}\right)^n \overline{P}_{nm}(\sin\phi) \times \left[\overline{C}_{nm}\cos(m\lambda) + \overline{S}_{nm}\sin(m\lambda)\right]$$

where n is the degree and m is the order, $\overline{P}_{nm}$ are fully normalized Legendre polynomial and associated functions, r_0 is the reference radius, ϕ is the latitude and λ is the longitude, $\overline{C}_{nm}$ and $\overline{S}_{nm}$ are the series coefficients determined from observational data from a variety of sources. Similar spherical harmonics can be used to describe the shape of planetary bodies.

Before dedicated gravity satellite missions such as CHAMP (2000), GRACE (2002), and GOCE (2004), global gravity observational data were obtained by tracking satellites using SLR (Degnan). A satellite orbit is determined primarily by the Earth's gravity field and affected by nonconservative forces such as drag (atmospheric, thermal, neutral density and charged particles) and radiation pressure. If the effects of the nonconservative forces can be accounted for, then the differences between the predicted and determined orbit of a satellite can be attributed to inaccuracies in the gravity model. Data from SLR, in situ, airborne and shipborne gravimetry, and satellite altimetry have been used to produce gravity models until this last decade. However, data from SLR provide the longest baseline to study temporal variations of the low-order zonal harmonic components of the gravity field (Degnan).

Terrestrial Reference Frame (TRF) and Earth Orientation Parameters (EOP)

Satellite applications require a foundation of permanently operating reference stations to collect the observations required to provide their mapping, positioning, and timing services. This network of stations serves as a terrestrial reference frame which defines the origin (center of mass) and orientation of the Earth. Earth orientation parameters – universal time (UT1), length of day (LOD), and coordinates of the pole and celestial pole offsets – describe the irregularities of the Earth's rotation and the orientation of the axis of rotation relative to inertial space and a celestial reference system. Observations with space geodetic techniques, including SLR, LLR, GPS, and VLBI, provide the required data to define the Earth's center of mass, UT1, LOD, and polar motion. VLBI is the only technique capable of accurately determining changes in the orientations of the earth with respect to the crust and to a celestial reference frame composed of natural radio sources (quasars) – the best current approximation of a true inertial reference.

Precision Orbit Determination for Navigation and Earth Observation Missions

Precision orbit determination (POD) is an important aspect of satellite operations, and for some satellites, such as navigation and remote sensing satellites, it is of crucial importance. It is also a technique that is in cyclical improvement. In order to obtain a precise orbit an accurate gravity model is required. Over periods of years, gravity models are improved, based on observations of satellite orbits obtained from optical, radar, and SLR tracking. The improved gravity model in turn allows for better orbital determination and so the cycle continues. In the early years of the space era, satellites were tracked from the ground using optical photographic cameras and basic Doppler radar techniques with accuracies of approximately 10 m for satellites in a 1,000 km altitude orbit (Vetter et al. 2007). The introduction of SLR in 1964 provided and alternate method for satellite tracking with an improved accuracy of a few meters. The ability to track satellites has continued to improve over the years to the millimeter-level accuracy obtainable today (McGarry et al. 2005).

SLR is a more precise technique than radar because it can obtain accurate ranges to retroreflector arrays, whose position with respect to the satellite center of mass is well known, whereas radar obtains a range to the center of the satellite radar cross section, whose position relative to the center of mass is known to a lower level of accuracy. Currently, satellites with orbital altitudes below 20,000 km can be continuously tracked using GNSS (or other systems such as NASA's TDRS) with centimeter-level precision or better. However, for GNSS satellites to provide positioning, timing, and navigation accurately is necessary to have accurate knowledge of their own orbits. GNSS satellites are tracked by a variety of means including optical and radar. Most of GLONASS satellites, the two current Galileo spacecrafts (GIOVE-A and -B), one of the Chinese COMPASS, the Japanese QZS-1, and one GPS satellite (GPS-36) carry retroreflector arrays to be tracked by SLR (GPS-35 decommissioned in April 2009 also carried an array) (The International Laser Ranging Service). Other satellites whose orbit needs to be accurately determined for the fulfillment of their scientific objectives, and are therefore tracked by SLR include gravity mappers GOCE and GRACE; radar and LiDAR altimeters Cryosat, Jason 1 and 2, ICESat (decommissioned); and remote sensing satellites Envisat, ERS-2, TerraSAR-X, and TanDEM-X (The International Laser Ranging Service).

Laser Altimetry and Topographic Mapping

Laser altimetry was the first application of spaceborne LiDAR on which the active equipment was carried by the spacecraft. Laser altimetry originated as an alternative to more traditional Radar altimeter. This was because the large divergence of radio beams makes its footprint on the surface of the planet many times larger than the footprint of a narrower laser beam. In altimetry, a smaller footprint results in a more accurate and representative estimate of height (Bufton 1989). As illustrated

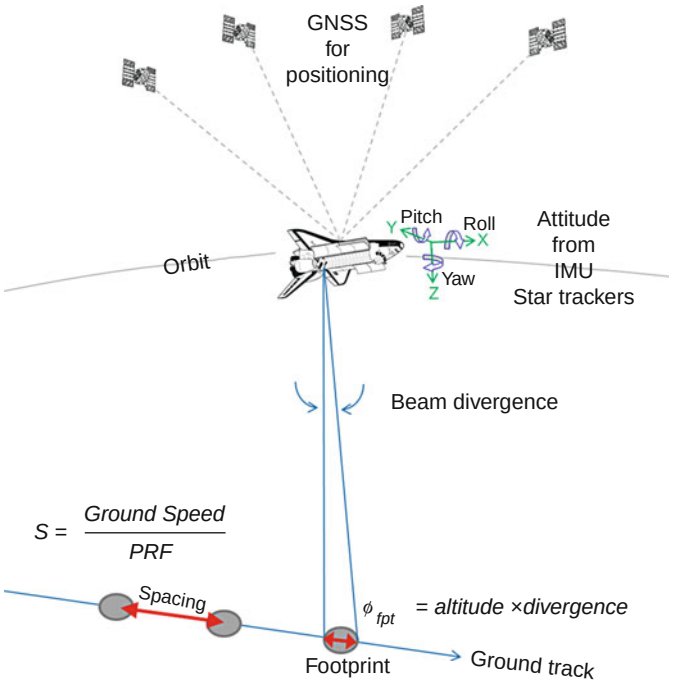


Fig. 31.10 Principles of operation of satellite LiDAR altimetry

in Fig. 31.10, in nadir-looking satellite LiDAR altimetry, the laser footprint is dependent on the satellite orbital altitude and laser beam divergence, while the spacing between footprints (spatial resolution) depends on the orbital velocity and the laser pulse repetition rate (PRF). The accuracy of the derived elevation depends on the precise determination of the spacecraft orbit and attitude.

The first spaceborne altimetry systems were not deployed on Earth observation missions, but rather on missions to the Moon and Mars. This was because the Earth's atmosphere presented a huge challenge as most of the laser energy is scattered by atmospheric constituents on a two-way trip from outside the atmosphere to the ground and back. Table 31.2 presents a historical evolution of spaceborne LiDAR altimeters and their main technical characteristics. The first laser altimeter system was deployed with the Apollo 15 mission to the Moon in 1971. The altimeter was part of the orbital science investigation, and was designed to take an altitude reading for each photograph taken with a mapping metric camera (every 20–28 s), although the altimeter was also able to range independently of the camera (at least every 20 s) (Alley et al. 1969). The metric camera, the altimeter, and two other cameras (panoramic and stellar) were located in the scientific instrument module (SIM) within the Apollo service module. The Apollo laser altimeter was based on a Q-switched ruby laser and a photomultiplier tube detector,

Table 31.2 Evolution of technical characteristics of spaceborne laser altimeters

Year	System	λ (nm)	PRF (Hz)	Pulse width (ns)	Pulse energy (mJ)	Beam diver. (μrad)	Range (km)	Detector
1971	Apollo 15 laser altimeter	694	<0.05	10	200	300	110	PMT
1992	Mars Orbiter Laser Altimeter 1	1,064	10	8	48	420	780	Si APD
1994	Clementine LiDAR	1,064	0.6	10	171	500	640	Si APD
1994	Lidar In-space Technology Experiment[a]	1, 064, 532, 335	10	27	470	1, 800	260	APD PMT PMT
1996	Shuttle Laser Altimeter 1	1,064	10	8	40	350	305	Si APD
1996	NEAR Laser Rangefinder	1,064	8	12	15	235	50	Si APD
2003	Geoscience Laser Altimeter System	1,064, 532	40	5	75,32	110	600	Si APD
2004	Mercury Laser Altimeter	1,064	8	6	20	80	800	Si APD
2009	Lunar Orbiter Laser Altimeter	1, 064, 532	28	6	2.7	100	50	Si APD

[a]LITE was designed primarily as an atmospheric LiDAR although it performed ranging to land and ocean surface. Difference in the design of atmospheric LiDARs and altimeters can be observed in terms of the pulse widths and beam divergence which tend to be larger in atmospheric LiDARs

the system was also deployed on the Apollo 16 and 17 missions in 1972. At its highest sampling rate of 0.05 Hz, the altimeter sampled the lunar surface height every 30–43 km with a footprint of roughly 30 m in diameter. The main problem with this instrument was its short lifetime; during the Apollo 15 mission the altimeter showed anomalous operation and stopped working in lunar orbit #38. As a result, only two complete and two partial surface profiles had useful data (Robertson et al. 1972). For the Apollo 16 mission, the sampling rate was reduced and the instrument lifetime was extended to lunar orbit #63, some 2,372 laser pulses, of which 69% had valid data, yielding five complete lunar surface profiles (Wollenhaupt et al. 1972). For the last Apollo lunar mission the laser was modified to increase its lifetime, and the altimeter lasted during the entire mission. The laser fired 4,026 pulses, and yielded 16 complete lunar surface profiles (Wollenhaupt et al. 1973). Data from all the missions combined yielded 7,080 height points, and from these a lunar mean radius was determined and a spherical harmonic representation of the lunar shape was produced complete to the 12th order and degree. However, coverage was limited to $\pm 26^{\circ}$ lunar latitude.

LiDAR altimetry returned to the Moon in 1994 onboard the Clementine mission. This instrument had a mass of only 2.4 kg (Smith et al. 1997) (compared to the 22.5 kg of the Apollo Altimeter (Robertson et al. 1972)), yet it fired around 650,000 laser pulses. Because the system was designed as a military ranging system and not

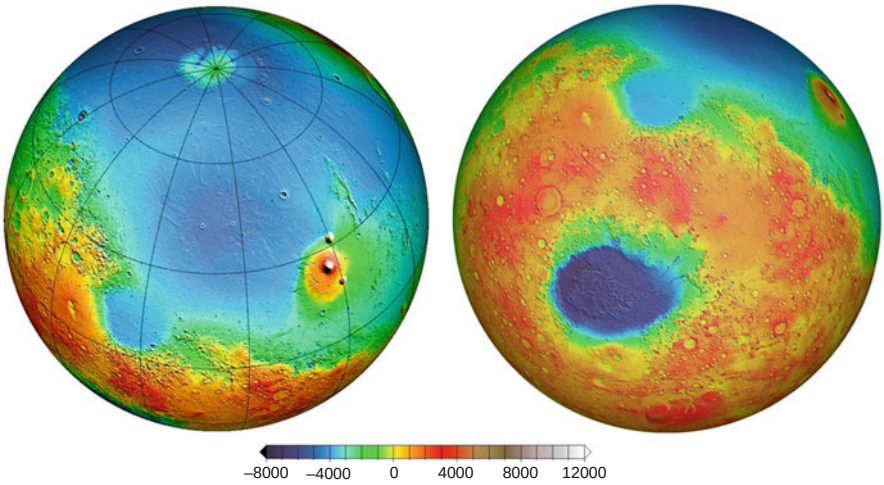


Fig. 31.11 Mars topography from MGS – MOLA-2 (Image courtesy of NASA)

an altimeter, only 19% of the fired pulses caused reflections that were detected, and of these only 72,548 where filtered out as valid surface returns (Smith et al. 1997). These data covered the lunar surface between 79°S and 81°N latitude, with a minimum along-track resolution of 20 km and an across-track resolution of roughly 60 km. From these data a spherical harmonic representation of the lunar shape complete to the 72th order and degree was produced (Smith et al. 1997). Most recently, the Lunar Reconnaissance Orbit (LRO) carrying the Lunar Orbiter Laser Altimeter (LOLA) (Ramos-Izquierdo et al. 2009) has been mapping the moon since September 2009, and as of June 19, 2010, LOLA had collected over two billion elevation measurements using its multichannel technology (Smith et al. 2010).

Besides the Moon, the shape and topography of three other extraterrestrial solar system bodies have been mapped: Mars, Mercury, and the asteroid 433 Eros. The first attempt to use LiDAR to map the Martian topography was the Mars Orbiter Laser Altimeter 1 (MOLA-1) launched onboard the Mars Observer launched in 1992 (Smith et al. 2001; Garvin et al. 1998). Unfortunately, the Mars Observer was lost on August 21, 1993, a few days before the orbit insertion maneuver. The second attempt was by MOLA-2 onboard the Mars Global Surveyor, MOLA-2 performed regular mapping operations between February 28, 1999, and June 30, 2001, and within this time frame approximately 640 million points were collected of the Martian surface (Smith et al. 2001; NASA). Figure 31.11 shows some samples of Mars topography from MOLA-2 data.

Eros was mapped by the laser rangefinder (Colea et al. 1996) onboard the NEAR-Shoemaker spacecraft launched in 1996, which entered orbit around Eros on February 14, 2000, landed on the surface of the asteroid on February 12, 2001, and the mission was terminated on February 28, 2001. During its mapping mission the laser range finder collected around 11 million measurements and allowed the best determination of shape, gravity, and rotational state of any asteroid to date

(Zuber et al. 2000; Miller et al. 2002). The most recent extraterrestrial body whose surface has been studied with LiDAR is Mercury. The Mercury surface, space environment, geochemistry and ranging (MESSENGER) mission was launched on August 3, 2004, carrying the Mercury Laser Altimeter (MLA) (Cavanaugh et al. 2007). After launching from Earth, MESSENGER has to perform six reversed gravity assists to obtain an orbital orientation and velocity suitable for its orbital insertion around Mercury in March 2011. These gravity assists are the result of flybys of planetary bodies, one with Earth (2005), two with Venus (2006 and 2007), and three with Mercury (January and October 2008, September 2009). MLA has been activated on the three Mercury flybys, and results have been reported for second flyby. During the flyby a 3,200-km-long profile along the equatorial region was collected (Zuber et al. 2008). The laser footprint at the surface ranged between 23 and 134 m while the spacing between the footprints varied from 725 to 888 m. Even this modest data profile has improved our knowledge of the shape and topography of the planet and has provided a preview of Mercurian crater morphology.

With regard to planet Earth, there are a few reports that indicate the existence of an altimetry LiDAR system named LORA, which was used to obtain precise altitude of photographs taken from a large format camera onboard a Soviet satellite (Werner et al. 1996; Werner et al. 1995). This LIDAR was reported to be operational as early as 1984; however, it has been hard to obtain independent confirmation of these reports. The first confirmed LiDAR returns from the surface of the Earth were obtained in September 1994 during the STS-64 mission. The LiDAR In-space Technology Experiment (LITE) was flown into space in the cargo bay of the space shuttle Discovery (Winker et al. 1996). However, LITE was designed primarily as an experimental atmospheric LiDAR and will be discussed in greater length in the next section.

The first LiDAR altimeter designed for Earth observation was the Shuttle Laser Altimeter (SLA) (Garvin et al. 1998). SLA was designed to fit in two hitchhiker canisters mounted on as special bridge structure carried in the Shuttle cargo bay as part of the small self-contained payload program (SSCP) most commonly known as the Getaway Special (GAS). This compact design allowed the SLA to be carried on any shuttle mission on which there was room for the GAS bridge. The SLA design was based on MOLA-1 and was constructed using MOLA spares. One of the GAS canisters housed the optical receiver that consisted of a 38 cm Cassegrain telescope and at its prime focus a silicon avalanche photodiode detector (Si APD). It also contained a coaxial transmitter based on a diode-pumped, Q-switched, Nd:YAG laser. The second canister contained the flight computer, power electronics, temperature sensors, and ancillary equipment. An upgrade from the MOLA architecture was the inclusion of a waveform recorder which digitized each received pulse in 4 ns samples quantized at 8 bits. The digitizer allows the determination of a redundant time of flight obtained from the time interval meter (TIM) to characterize the structure of the surface that caused the backscattering. SLA was flown twice: the first time was during the Endeavour STS-72 mission in January 1996 (Garvin et al. 1998), and the second during the Discovery STS-85 mission in August 1997 (Carabajal et al. 1999). During the STS-72 mission, SLA-01 collected about

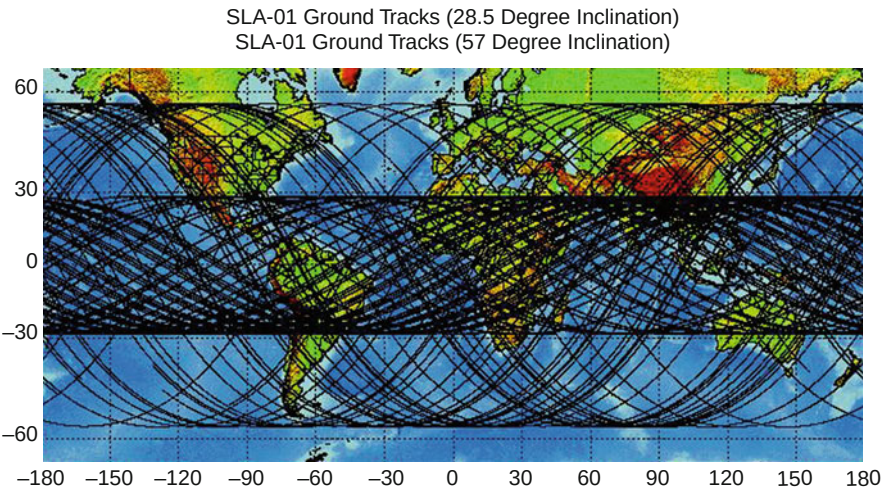


Fig. 31.12 Ground tracks for the SLA-01 and SLA-02 collections

82 h of nadir-looking altimetry data, roughly totaling three million observations. The Endeavour orbit for STS-72 had an altitude of 300 km, an inclination of 28.45°, and an average orbital velocity of 7 km/s. The orbit inclination and the nadir-looking orientation of SLA constrained the ranging acquisition in the mid latitudes between 28.45°N and 28.45°S, the laser footprint size determined from the altitude and beam divergence was ∼100 m, and the spacing between footprints determined by the combination of the velocity and PRF was ∼700 m. After preprocessing and filtering, roughly 475,000 valid returns were obtained from land and 1.1 million from the ocean surface (Garvin et al. 1998).

For the second flight of SLA onboard the Discovery STS-85 mission the hardware was upgraded to include a variable gain amplifier (VGA) that allowed the detector to adjust to the high dynamic range of the laser returns observed during SLA-01 that had caused the saturation of the waveform recorder. Similar to SLA-01, SLA-02 collected almost 83 h of data, firing close to three million points (Carabajal et al. 1999). The main difference was that the orbital inclination of STS-85 was 57° which allowed altimetry sampling up to high latitudes. After preprocessing and filtering, roughly 590,000 valid returns were obtained from land and 1.5 million from the ocean surface. There were plans for two more flights of SLA to keep improving the system by reducing the beam footprint and increasing the PRF. A third flight was planned for late 1998 in partial support of the Shuttle radar topography mission (SRTM). However, no additional flights of SLA past SLA-02 were executed. Figure 31.12 shows the ground tracks of collected data from the SLA-01 and -02 experiments. Data from the SLA mission were compared against other ground and sea surface elevation databases (Behn and Zuber 2000; Harding et al. 1999) and were also used to perform accuracy assessments of the later collected STRM dataset (Suna et al. 2003).

The lessons learned from the two SLA missions were incorporated into the most recent and advanced spaceborne LiDAR altimeter for Earth observation to date: the geoscience laser altimeter system (GLAS). GLAS was deployed on a dedicated platform: the Ice, Cloud, and land Elevation Satellite, ICESat (Abshire et al. 2005; Schutz et al. 2005). ICESat was launched on January 13, 2002, into a 600 km altitude orbit with a 94° inclination. This orbit has a nadir repetition cycle (within 1 km) of 183 days (or 2,753 revolutions), and ground track spacing within the cycle is 15 km at the equator and 2.5 km at ±80° latitude.

The GLAS transmitter was powered by three diode-pumped, Q-switched Nd: YAG lasers which operate one at a time (Abshire et al. 2005; Schutz et al. 2005). The lasers produced 5 ns pulses at 40 Hz and 1,064 nm, part of the 1,064 nm pulse was passed through a nonlinear frequency-doubler crystal to obtain a 532 nm pulse. The transmitted pulse energy was 75 mJ at the infrared wavelength and 35 mJ at the green wavelength with a beam divergence of 110 μrad. The orbital and laser characteristics yielded a footprint of 65 m on the surface with successive spots spacing of 172 m. The backscattered radiation was collected by a 100 cm diameter beryllium telescope, the 1,064 nm component is used to detect strong backscattering in analog mode from clouds, water, ice, and land surfaces, while the 532 nm component was use in photon-counting mode to detect scattering from thin high-altitude clouds. The 1,064 nm signal was filtered through an 800 pm spectral filter and detected by a Si APD (there were actually two APDs for redundancy). The APD output was digitized separately at 1 GHz and 2 MHz rates, the 1 GHz rate enables a range resolution of 15 cm for accurate surface determination while the 2 MHz yield a 77 m resolution for the detection of thick clouds and aerosols. The 532 nm component was filtered twice through 370 and 30 pm spectral filters to limit background light, and the resultant beam was split into eight beamlets that were individually detected by eight Si APD detectors operating in Geiger mode (Abshire et al. 2005; Schutz et al. 2005).

To obtain an accurate geolocation of the laser returns, besides the accurate determination of the two way time of flight, it is necessary to determine the position and attitude of the instrument and the orientation of the fired laser shot. Precise orbit determination (POD) was performed via GPS tracking using two redundant dual-frequency blackjack receivers connected to two separate antennas on the zenith deck of the spacecraft (Schutz et al. 2005). On the nadir deck, a corner cube reflector array allowed the satellite to be tracked using SLR for an accuracy assessment of the GPS-derived orbit (Schutz et al. 2005). There were two attitude determination systems onboard the spacecraft, one for the satellite and one for the sensor optical bench. GLAS's optical bench attitude was determined to better than 10 μrad with reference to inertial space through a stellar reference system (SRS) based on data acquired from a 10 Hz zenith looking star camera, and a precision gyroscope (Schutz et al. 2005). In addition, the far-field pattern of the laser beam for each laser pulse was imaged and its orientation determined with respect to the optical bench and inertial space (Schutz et al. 2005). GLAS was designed to perform nadir pointing ranging; however, the spacecraft could be commanded so GLAS could point ±5° off-nadir to acquire targets of opportunity. Figures 31.13

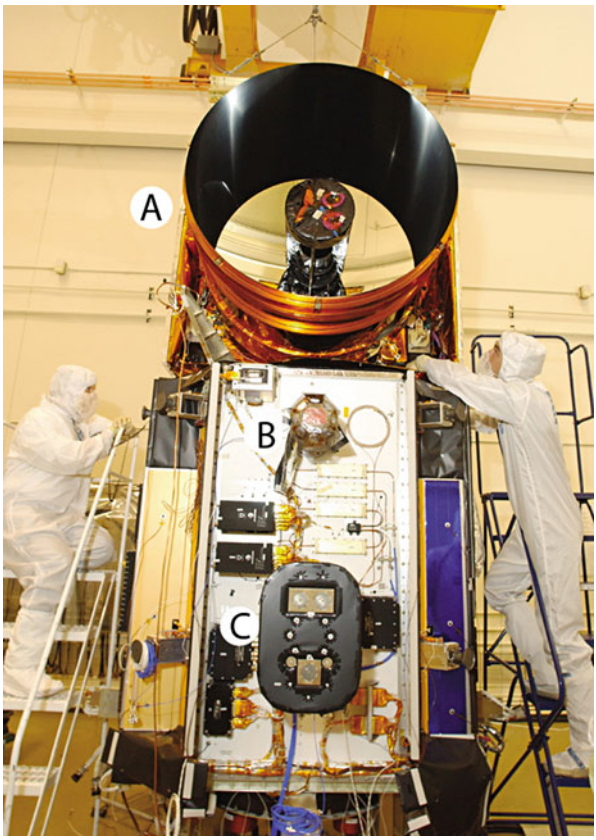


Fig. 31.13 ICESat's nadir deck showing (**a**) receiving telescope, (**b**) retroreflector array, and (**c**) telemetry antenna (Image courtesy NASA)

and 31.14 show photos of the ICESat satellite integration, which highlight crucial elements of GLAS and the subsystems that enabled precise orbit and attitude determination.

The GLAS lasers were expected to last for 3 continual years of operation. Unfortunately, laser 1 failed prematurely after 37 days. This failure prompted a change in the collection strategy for the mission from a continual collection with an 8 day repeat cycle to a campaign collection mode with a 33 day repeat cycle, resulting in less temporal and spatial resolution but allowing the measurement of polar ice height over the extended 7 year period. The last GLAS laser ceased operation on October 11, 2009, and was decommissioned in August 14, 2010 (Abshire et al. 2005). In its almost 8 years in space GLAS fired almost two billion laser pulses (Abdalati et al. May 2010). The primary objective of the ICESat mission was the accurate determination of interannual and long-term changes of polar ice volume and mass balance; however, additional applications included the monitoring of land topography, hydrology, vegetation canopy height, cloud heights, and atmospheric aerosol distributions (Abshire et al. 2005). Figure 31.15 illustrates

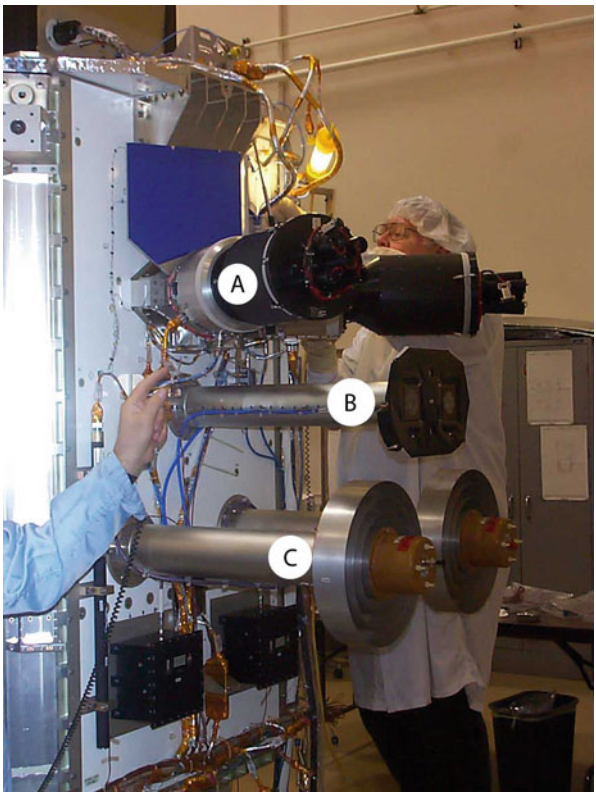


Fig. 31.14 ICESat's Zenith deck showing (**a**) the satellite star trackers, (**b**) telemetry antenna, and (**c**) GPS antennas (Image courtesy NASA)

the use of GLAS data collected between 2003 and 2007 to generate maps of Antarctic and Greenland's ice sheets elevation changes rates. The images indicate the dynamic thinning of ice sheets in certain areas and the accumulation of ice and snow in others.

To continue the critical measurement of the polar ice sheets an improved ICESat-2 mission is currently being developed and scheduled for launch in 2016 (Abdalati et al. 2010). To obtain a denser spatial sampling than that of ICESat-1, a multi-beam approach (Figs. 31.16 and 31.17) combined with a higher PRF is under study (Yua et al. 2010). The baseline design consists of a micropulse laser with a PRF of a 10 kHz, 0.1 mJ of energy per pulse and a pulse width of ~1 ns. A diffractive optical element (DOE) splits the beam into nine beamlets with different energy and arranged in a 3 × 3 slanted array (Fig. 31.17). The footprint of each beam is expected to be 10 m in diameter, and because the array is slanted with respect to the flight line, the projection on the ground produces nine parallel tracks grouped in threes. Each group will be spaced 3 km apart and within the group, the spot separation in the across track direction will be 50 m (Fig. 31.17).

Besides ICESat-2, the NRC decadal survey recommends two additional LiDAR altimetry missions to be launched before 2020. The most immediate mission is the

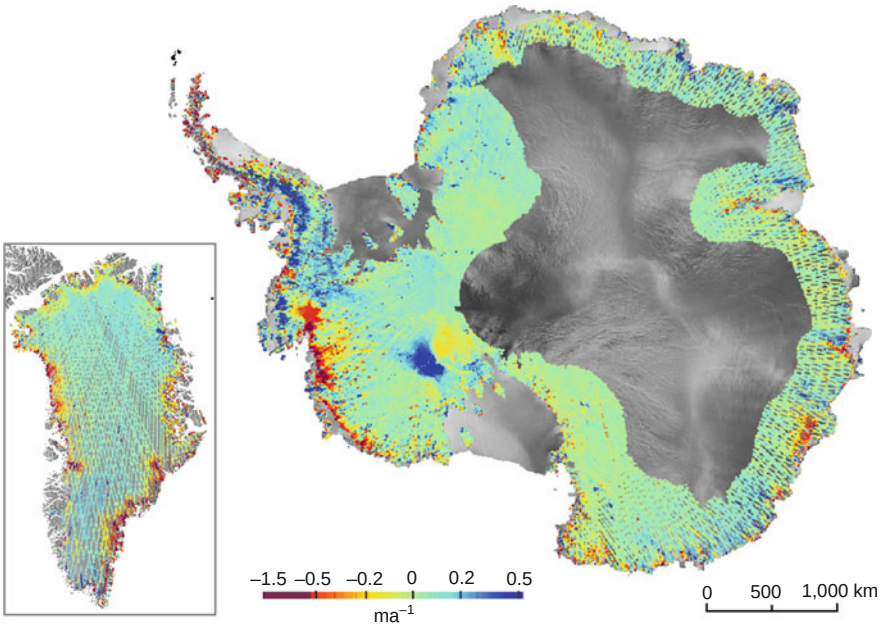


Fig. 31.15 ICESat data showing changes in elevation (m/year) in the Greenland and Antarctica ice sheets (Image courtesy of NASA)

Deformation, Ecosystem Structure and Dynamics of Ice (DESDynI) (National Research Council 2007) which will attempt to exploit the synergy between L-band polarimetric InSAR and multi-beam LiDAR altimeter. The second mission recommended for the last half of the decade is LiDAR surface topography (LIST). The objective of LIST will be to produce a global elevation dataset with a horizontal resolution of 5 m with at least 10 cm vertical precision (National Research Council 2007).

Atmospheric Studies

Obtaining global datasets on atmospheric composition, structure, and circulation is of crucial importance for the development of global climatic models. These datasets are obtained with a myriad of instruments using both direct detection and remote sensing. It is also important to use both a bottom-to-top and top-to-bottom approach. In situ Radar and LiDAR provide the bottom-to-top measurements which are able to detect phenomena in the lower denser layers of the atmosphere, but because of the higher density of the lower layers are not able to obtain measurements of the thinner upper layers. Spaceborne sensors provide the top-to-bottom view detecting phenomena in the higher and thinner layers of the atmosphere and generating a much needed global coverage not attainable any other way.

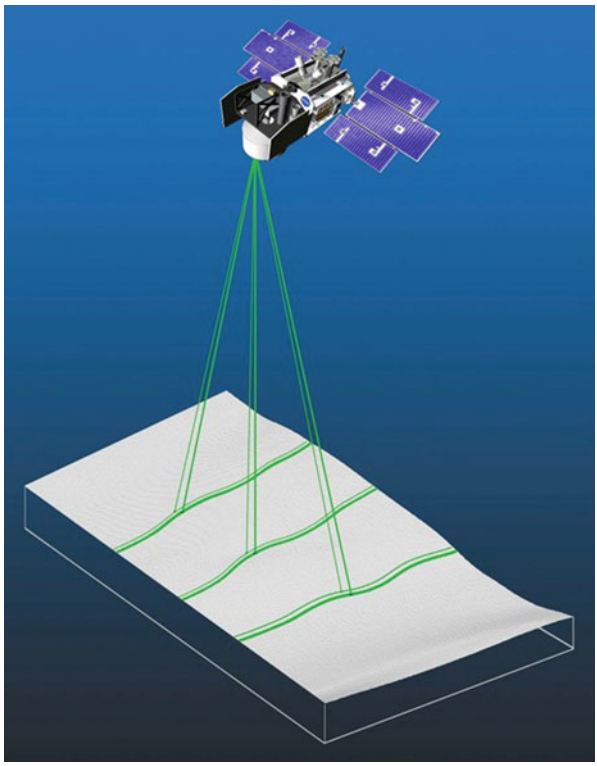


Fig. 31.16 Multi-beam LiDAR transmitter concept for the ICESat 2 mission (Image courtesy of NASA)

Spaceborne atmospheric LiDARs have been employed mainly to study the Earth's atmosphere and in particular cases the Martian atmosphere. All other planetary atmospheres in the solar system are too dense to be probed in the optical wavelengths.

Although there are some unconfirmed reports that as early as 1984 a Soviet reconnaissance satellite carried a LiDAR to obtain precise altitude of photographs taken from a large format camera, and was used for early atmospheric observations (Werner et al. 1996; Werner et al. 1995), the first confirmed spaceborne LiDAR built primarily for atmospheric studies flew into space in September 1994. The LiDAR In-space Technology Experiment (LITE) was flown into space in the cargo bay of the space shuttle Discovery during the STS-64 mission (Winker et al. 1996). LITE was designed and built based on the experience accumulated over two decades by NASA's Langley Research Center designing, building, and operating ground-based and airborne atmospheric LiDARs. LITE was designed mainly to detect and measure clouds and aerosols in the troposphere and stratosphere, determine the height of the planetary boundary layer (PBL), and to derive temperature and density profiles in the stratosphere at heights between 25 and 40 km. It was also capable of detecting returns from land and sea surfaces, however, without the precision of an altimetry system.

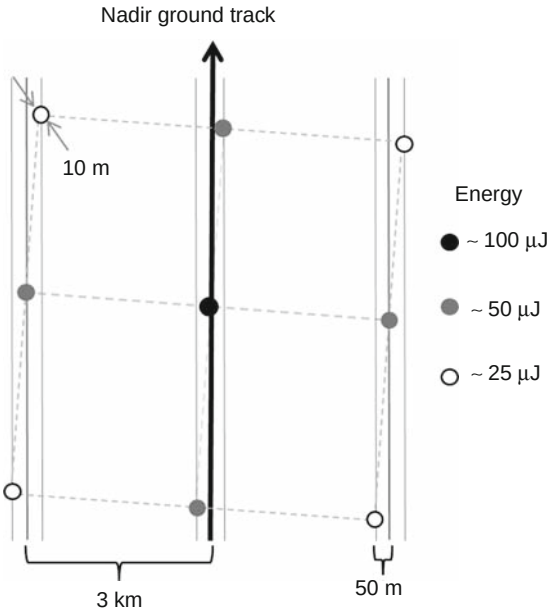


Fig. 31.17 Layout of the multi-footprint concept for ICESat 2 obtained from the DOE

As shown in Fig. 31.18, LITE was designed to fly in the cargo bay of the space shuttle integrated into a Spacelab 3-m pallet. The laser transmitter system was based on two redundant flashlamp-pumped, Q-switched, Nd:YAG lasers (Winker et al. 1996). Part of the energy of the 1,064 nm fundamental wavelength was passed through nonlinear frequency-doubling crystals to obtain 532 and 355 nm beams. The pulse repetition frequency (PRF) was 10 Hz and each pulse had a width of 27 ns, energy per pulse was 470, 530, and 170 mJoules, with a divergence of 1.8, 1.1, and 0.9 mrad for the 1,064, 532, and 355 nm wavelengths, respectively. The laser beams were steered through a two-axis gimbaled prism to maintain optical alignment with the field of view of the receiver. The receiver was based on a 1-m diameter Ritchey–Chrétien telescope, with a rotating wheel with multiple aperture stop settings to configure the instrument for day or night collections. Dichroic beamsplitters separate the return signal into the three spectral components, and part of the 532 return signal was used to determine and control the boresight alignment between the transmitter and receiver. The three beams were directed through narrowband spectral filters before their respective detectors, photomultiplier tubes (PMT) for the 355 and 532 nm components and an avalanche photodiode (APD) for the 1,064 nm component. The output from the detectors was digitized with 12 bit amplitude at 10 MHz (550 µs).

The STS-64 carrying LITE was launched into a 260 km altitude orbit with a 57° inclination and 7.4 km/s orbital velocity. These orbital characteristics combined with optical transmitter specifications yielded footprints 470 and 290 m in diameter for the 1,064 and 532 nm beams, respectively, and footprints were spaced



Fig. 31.18 The LITE LiDAR onboard the space shuttle Discovery (Image courtesy of NASA)

every 740 m. During the 11-day mission, LITE was operated roughly 5° off-nadir to avoid saturation from high specular reflections, and acquired a total of 53.6 h of quick view data (43.5 high rate profiles) collection, almost two million laser pulses were fired (1.16 from the first laser, 0.77 from the second) (Winker et al. 1996). These collections provided the first ever high-resolution transects of the atmospheric constituents and cloud structures. LITE data was validated against ground-based and airborne measurements.

After LITE there were several short-lived spaceborne atmospheric LiDAR experiments. Including the Balkan-1 onboard the Spektr module of the Russian MIR space station (launched on May 20, 1995) (Werner et al. 1995), the French designed and built l'Atmosphere par LIdar Sur SAliout (ALISSA) onboard the Priroda module (launched April 23, 1996) also of the MIR space station (Chanin et al. 1999), and the Balkan-2 onboard the ALMAZ-1B Earth observation satellite (launched on January 1, 1997) (Matvienko et al. 1994). Currently, the joint NASA and CNES Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite carries the only operational spaceborne terrestrial atmospheric LiDAR: the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP). Launched on April 28, 2006, CALIPSO is managed by NASA's Langley Research Center due to the center's overall expertise on atmospheric LIDAR systems. An interesting aspect of the CALIPSO mission is that it is part of the afternoon or "A-Train" satellite constellation which also includes the Aqua, Aura, PARASOL,

and CloudSat satellites. All the satellites follow the same Sunsynchronous orbit (705 km altitude, 98° inclination), and are separated from each other by a few seconds to minutes. The sensor suite carried by the satellites in the constellation enables the first global, near simultaneous measurements of aerosols, clouds, temperature, relative humidity, and radiative fluxes. CloudSat with its cloud profiling radar (CPR) leads CALIPSO by 10–15 s, which allows for the simultaneously profiling of the same cloud systems with both the radar and LiDAR. CALIPSO attitude is controlled such that the LIDAR points 0.3° ahead of nadir in the along-track direction, to avoid saturating the detector with strong specular returns from calm water bodies. Based on the spacecraft orbital parameters and the transmitter characteristics, the footprint on the ground is 70 m in diameter, and adjacent spots are separated 333 m in the along-track direction.

Figures 31.1 and 31.2 (see the technical Overview of LiDAR section) illustrate CALIOP's system design; its transmitter is based on two redundant, diode-pumped, Q-switched Nd:YAG lasers with a PRF of 20.16 Hz, with a nominal energy per pulse of 220 mJ (Winker et al. 2004). Part of the energy of the 1,064 nm pulses is passed through a frequency-doubling crystal to produce a 532 nm component. The energy of the transmitted pulses at both wavelengths is nominally 110 mJ and is measured before passing through a beam expander that limits the divergence to 100 μrad. The laser polarization is also controlled to be linearly polarized with a purity greater than 99%. Backscattered photons are collected by a 1-m beryllium mirror telescope, a field stop at the telescope focus limits the receiver field of view and provides a spatial filter limiting background noise. A dichroic beamsplitter separates the 1,064 and the 532 nm components, the 1,064 nm stream is filtered through a narrow band spectral filter and then is directly detected by an avalanche photodiode (APD). The 532 nm stream is passed by a double spectral and etalon filter to limit the background noise. The pure 532 nm component is then passed through a polarization beam splitter to separate the perpendicular and parallel polarization components, and from there directed to separate photomultiplier tubes (PMT). For each channel, the output of the detector is amplified by two parallel amplifiers and 14bit digitizers which provides an effective 22-bit dynamic range. This dynamic range covers the expected magnitude range of the backscattering signals from molecules, aerosols, and cloud surfaces encountered in the atmosphere. Data acquisition starts when the laser pulses are estimated to be 115 km above sea level and finish at above 18.5 km below MSL, the output from the digitizers is sampled and recorded at 10 MHz (15 range bin).

CALIOP's data provides thin transects of the Earth's atmosphere that characterize the vertical distribution of atmospheric aerosols and molecules. It is a valuable complement to other type of meteorological sensors that provide information on the horizontal distribution of clouds and other atmospheric features. Figure 31.19 shows an example of one of such atmospheric transect as CALIPSO was on an ascending pass from South America up to the north Atlantic (Fig. 31.20). Figure 31.19 illustrate the difference in detected scattering at the 532 and 1,064 nm channels. The 532 nm is the most sensitive channel due to its shorter wavelength. Figure 31.21 illustrates the complementary value of atmospheric LiDAR to other

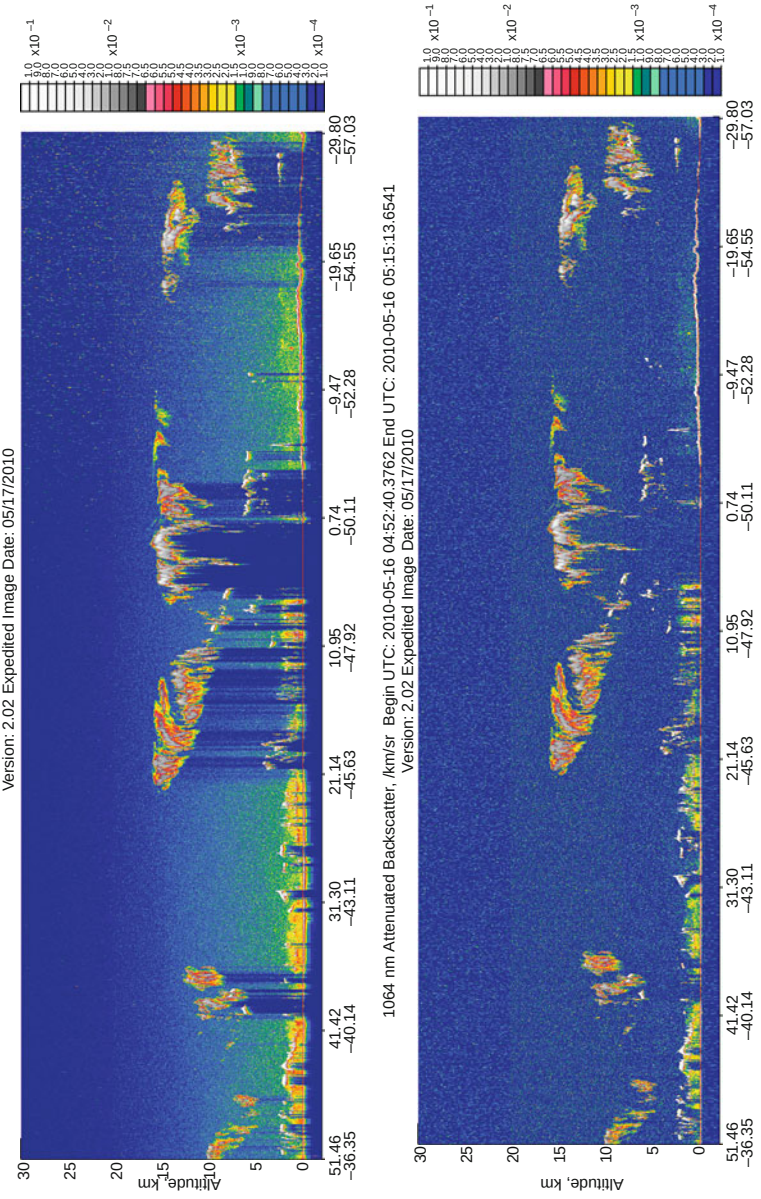


Fig. 31.19 Example of an atmospheric backscattering profile detected by CALIOP's 532 nm parallel + perpendicular and 1,064 nm channels (Image courtesy of NASA)

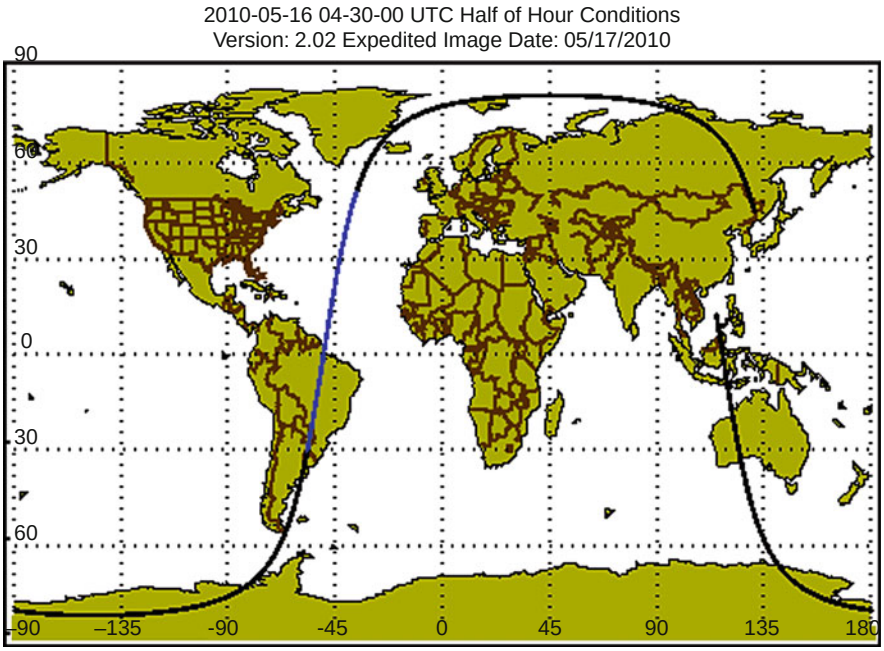


Fig. 31.20 Ground track for the atmospheric scattering profiles shown in Fig. 31.19 (Image courtesy of NASA)

forms of passive remote sensing; overlaid over an AQUA MODIS image is the ground track of the CALIOP profile. Complementing the horizontal cloud structure from the MODIS image, CALIOP data shows the vertical cloud structure, including an ash plume produced by the eruption of the Eyjafjallajökull volcano in May 2010.

An interesting implementation of atmospheric LiDAR occurred in 2008, when a ground-based atmospheric LiDAR was deployed and made successful measurements of the Martian atmosphere for 152 days. The LiDAR was part of the meteorological station (MET) onboard the Phoenix Mars Lander that was launched from Earth on August 4, 2007, landed on Mars on May 25, 2008, and collected and transmitted scientific information until October 29, 2008 (total of 152 Martian days) (Whiteway et al. 2011). What is outstanding about this LiDAR system is the degree of miniaturization that was achieved. The entire unit had a total mass of 6 kg. The transmitter was based on a single diode pumped, Q-switched Nd:YAG laser with a PRF of 100 Hz and a pulse width of 10 ns (Whiteway et al. 2008). Part of the energy of the 1,064 nm pulses was passed through a frequency-doubling crystal to produce a 532 nm component. The pulse energy was 0.3 mJ at 1064 nm and 0.4 mJ at the 532 nm. The divergence of the laser beams was 250 μrad. The backscattered photons were collected by a 10-cm diameter reflective telescope, separated into the two spectral components by a dichroic mirror. The 1,064 beam was filtered through a 2 nm interference filter and detected by a Si APD working in analog mode.

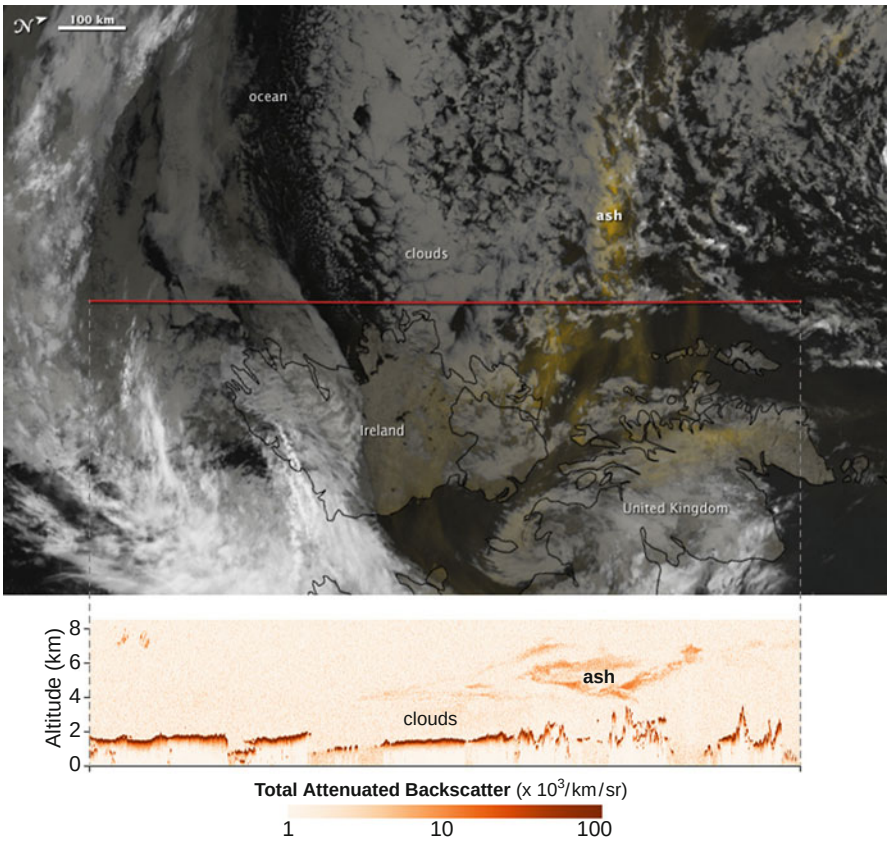


Fig. 31.21 Horizontal cloud structure from AQUA MODIS and vertical profile from CALIOP showing the ash plume from the Eyjafjallajökull Volcano (Image courtesy of NASA)

The 532 beam was passed through a 1 nm interference filter, limited by a field stop and detected by a PMT, whose signal was collected in both analog and photon-counting modes. The analog output was recorded with a 14-bit amplitude at a 30 MHz sampling frequency (333 μs per bin). Analog detection was used for backscattering below 10 km, while photon counting was used to detect weak signals from backscattering up to 20 km.

A future satellite, ADM Aeolus scheduled for launch in 2013, will carry the first atmospheric LiDAR to be used for the remote determination of global wind speed profiles. Along with temperature, pressure, and humidity, wind velocities are the basic variables used to describe the state of the atmosphere, and the knowledge of global circulation is crucial for the improvement of global climate models. The Atmospheric Laser Doppler Instrument (ALADIN) onboard Aeolus is designed and constructed as a direct detection Doppler LiDAR. The operation principle of the instrument, illustrated in Fig. 31.22, consists of detecting Mie and Rayleigh

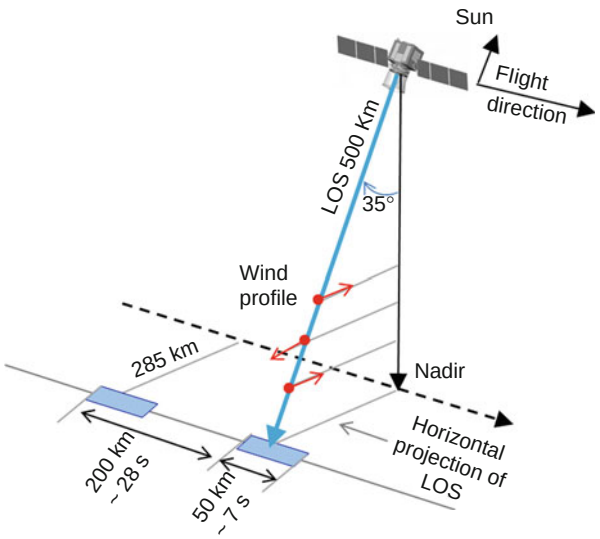


Fig. 31.22 Concept for the future ALADIN Doppler wind LiDAR

scattering by aerosols and atmospheric molecules and using a high-resolution spectrometer to measure the wavelength shift of the backscattered radiation with respect to that emitted by the laser transmitter (Ansmann et al. 2007). The wavelength shift is proportional to the relative velocity along the line-of sight (LOS) between the satellite and the scattering particles. By taking into account the spacecraft motion and the Earth's rotation it is possible to isolate the wind velocity. The satellite is planned to orbit at a 400 km altitude (7.21 km/s ground speed), and the ALADIN will point 35° off-nadir in the across-track direction. The Doppler shift and thus the wind speed is to be determined at different range (heights) along the LOS, and the wind horizontal component perpendicular to the satellite ground track will be projected from the slanted vector. Mission requirements call for the wind measurements to be averaged across 50 km cells, and average measurements are to be obtained about 200 km apart. To achieve these, the LiDAR will operate in burst modes, transmitting continuous burst for 7 s every 28 s (Ansmann et al. 2007).

ALADIN has a monostatic design, i.e., the transmit and receive paths go through the same telescope (Ansmann et al. 2007). The telescope is an afocal Cassegrain design with a diameter of 1.5 m and its field of view of only 12 μrad, which produce a footprint of 12–15 m at the end of the 500 km LOS. The optical transmitter is based on a single-mode, diode-pumped, Q-switched, frequency-tripled Nd:YAG laser. The output laser pulse in the ultraviolet range has a wavelength of 355 nm with a pulse width of 30 ns, energy per pulse of 120 mJ, and a planned PRF of 100 Hz. The spectrally pure laser pulses are passed through linear and circular polarizers before being expanded through the telescope. The backscattered photons are collected by the telescope and passed through the polarizers; only the parallel polarized components are accepted and passed through a field stop and 1 nm spectral filter to limit the effect of background illumination. Once the return beam is spatially and spectrally filtered it

is directed to the spectrometer system which is comprised of a Fizeau interferometer, which detects the spectrally narrow Mie backscattered peak (Channel 1), and two Fabry Perot etalons (Channels 2 and 3), to detect the wide Rayleigh–Brillouin backscatter spectrum. The output of the spectrometers is detected by two accumulation charged coupled devices (ACCD).

Besides the described spaceborne atmospheric LiDARs, both NASA and ESA are currently contemplating future mission that would incorporate atmospheric LiDAR instruments. Currently in the design phase, the joint ESA JAXA Earth Clouds, Aerosols and Radiation Explorer (EarthCARE) is aimed at improving our understanding of the interactions between cloud, radiative, and aerosol processes. EarthCARE proposes a suite of atmospheric instruments which includes a cloud profiling radar (CPR), multispectral imager (MSI), a broadband radiometer (BBR), and an atmospheric backscattering and depolarization LiDAR (ATLID). ALTLID is envisioned to be an ultraviolet high spectral resolution backscattering LiDAR (Le Hors et al. 2008), much like an upgraded version of CALIOP. On the Decadal Survey, the NRC recommended to NASA the design and implementation of three missions that incorporate atmospheric LiDARs (National Research Council 2007). The most immediate is the Active Sensing of CO_2 Emissions over Nights, Days, and Seasons (ASCENDS), which is envisioned to incorporate a multiwavelength LiDAR system. The Aerosol-Cloud-Ecosystems (ACE) mission, with a primary goal to reduce uncertainty about climate forcing in aerosol–cloud interactions and ocean ecosystem carbon dioxide (CO_2) uptake, will incorporate an atmospheric backscattering LiDAR, a multiangle polarimeter, and a Doppler radar. Finally a demonstration mission is recommended, the 3D-Winds, which should incorporate a Doppler LiDAR to map tropospheric winds for weather forecasting and pollution transport modeling.

Guidance, Navigation, Control, Inspection

The most recent application of LiDAR in spaceborne platforms is for on-orbit operations such as ranging for rendezvous and docking, active imaging for inspection and servicing, and robot vision for autonomous operation. The first use of LiDAR technology for semiautonomous/autonomous vehicle operation was for the Mars Pathfinder microrover: "Sojourner," which landed on Mars on July 4, 1997, and operated until September 27, 1997, when communication with the Lander was suddenly lost. The Sojourner microrover was equipped with a stereo-pair imaging system for rover navigation. To aid the camera in proximity operations a laser triangulation system was included which consisted of five semiconductor diode laser stripe projectors (JPL). Using preflight calibration tables, the system was able to determine distances from the rover to the projected laser stripes based on the pixel position on which the laser spots were detected.

The increasing need for active imaging for on-orbit inspection and servicing has created another application of LiDAR technology. This need was extremely evident after the tragic loss of the Columbia Shuttle during atmospheric reentry on February

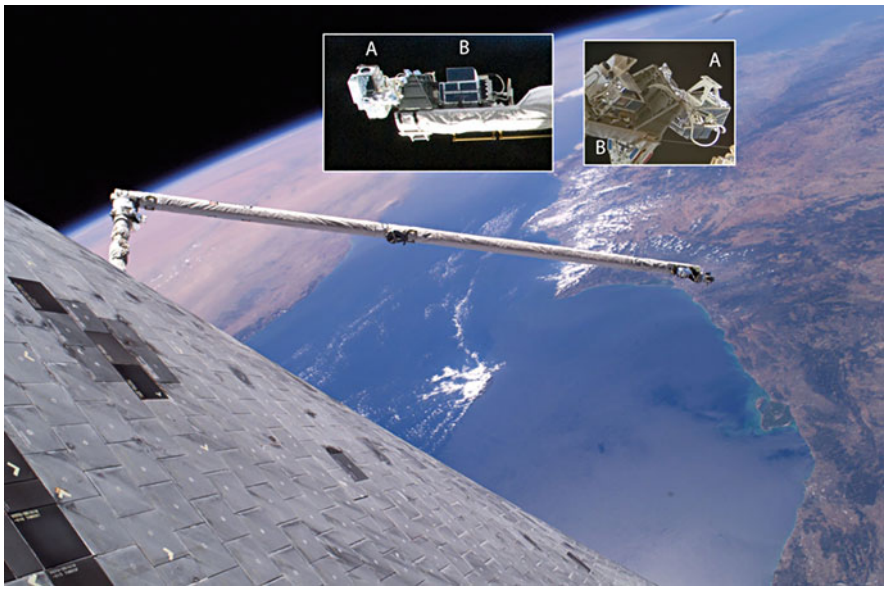


Fig. 31.23 The shuttle orbiter boom and sensor system (OBSS), inset images show close-ups of the laser dynamic range imager (LDRI) (**a**) and the laser camera system (LCS) (**b**) (Images courtesy of NASA)

1, 2003. Among the many improvements to the systems and procedures of the space shuttle program during the Return to Flight effort was to include on every future mission tools for on-orbit inspection. The inspection of the critical shuttle areas such as the wings leading edge and thermal protection tiling is now performed with a suite of passive and active imaging systems mounted on the end of 15-m long boom that serves as an extension to the shuttle remote manipulator system (SRMS). The orbiter boom and sensor system (OBSS) includes three sensors, two of them, the laser dynamic range imager (LDRI) and the intensified television camera (ITVC) are mounted on a pan and tilt platform, while the third sensor, the laser camera system (LCS) is rigidly mounted on the side of the boom (Fig. 31.23) (NASA 2005). LDRI and LCS are active imaging sensors that use nonconventional LiDAR technology for the collection of 3D data.

LDRI was developed by Sandia National Laboratories and uses a combination of phase difference ranging and video to derive 3D information. LDRI has a laser transmitter based on a continuous wave (CW) diode laser emitting light at 805 nm with a maximum power of 12 W (Smithpeter et al. 2000). The CW amplitude (intensity) is modulated at 3.125 or 140 MHz. In contrast to most LiDAR systems, where the divergence of a laser beam is restricted, the light from the LDRI is expanded and then passed through a diffuser plate to produce a floodlight effect, different plates can yield beam spreads of 10–60°, with a normal used value of 40°. This expanded beam is used to illuminate the target, the backscattered photons are collected by a refractive lens and passed through a narrow 30 nm spectral filter to

limit the contribution from external illumination and then focused on the cathode of an image intensifier tube. The optical gain of the intensifier tube is modulated with the same signal used to modulate the laser output. The output from the image intensifier is coupled by a fiber optic taper to a CDD detector which is read by a conventional 640 × 480 analog video recorder operating at 30 frames per second. To perform ranging of a given target area, the area is illuminated by the variable intensity laser for a given time on which several video frames are recorded. Assuming that each pixel of the frame is imaging the same target area and that the range remains constant throughout the different collected frames, it is possible to derive the phase difference for each pixel between the emitted and backscattered radiation by comparing the changes in intensity from several frame captures. Knowing the phase difference it is possible to determine the range between the sensor and the target on a pixel by pixel basis, thus generating intensity and spatial datasets of the illuminated area.

LCS was developed by the Neptec Design Group of Canada and first flew into space on August 10, 2001, as a detailed test objective (DTO) during the STS-105 mission of the Discovery space shuttle (STS-105 Shuttle Press Kit). LCS is a hybrid video and imaging LiDAR sensor, the LiDAR sensor is based on the triangulation ranging principle capable of imaging a 30° × 30° field of regard (FOR) from a range between 1 and 10 m (Dupuis et al. 2008; Deslauriers et al. 2005). The transmitter of LCS is based on a continuous wavelength-shifted Nd:YAG laser emitting at 1,500 nm. Scanning mirror/galvanometers are used to steer the laser beam in two dimensions over the FOR to illuminate the target. The reflected photons are captured by a refractive lens, filtered by a narrow bandpass spectral filter, and detected by a linear detector array (LDA). By determining the array coordinates of the pixel that detects the highest intensity signal, and knowing the baseline distance and the galvanometer angles is possible to determine the range to the target using the triangulation principle with high precision (3 mm @ 5 m) and at fast acquisition rates. LCS has been upgraded by Neptec to have a hybrid LiDAR design which combines a triangulation LiDAR operating at 1,400 nm with a time of flight (TOF) LiDAR operating at 1,540 nm which share the same scanning mechanisms (Dupuis et al. 2008; English et al. 2005). This upgraded sensor also includes a thermal imager and it is designated as TriDAR. TriDAR exploits in a synergistic approach the advantages of the TOF and triangulation ranging mechanism, combining the long-range capabilities with coarse precision of the TOF (range <3 km, <25 mm precision) with the sub-cm accuracy in the short range of the triangulation units. TriDAR first flew into space as a DTO onboard Discovery during STS-128 in August–September 2009, to demonstrate its capabilities to perform autonomous acquisition and tracking of the ISS. It also performed real time docking measurements during the STS-31 mission in April 2010.

An additional space-based ranging and imaging LiDAR was carried by the Air Force XSS-11 satellite which operated between 2005 and 2007. The rendezvous laser system (RLS) sensor, also referred to as spaceborne scanning lidar system (SSLS), was designed and manufactured by Optech and MDA as a system to allow XSS-11 to perform autonomous rendezvous and proximity maneuvers

(Nimelman et al. 2006; Dupuis et al. 2008). RLS was a time-of-flight scanning LiDAR with a 20° × 20° field of view, a laser beam divergence of 500 μrad, a maximum range of 5 km with a resolution of 1 cm, and an accuracy of 5 cm. During its 22 month operations, XSS-11 used RLS to perform rendezvous and proximity operation around its expended Minotaur launch vehicle and with several US-owned dead or inactive resident space objects.

Conclusion

Entire books have been written on the subject of LiDAR remote sensing from specific points of view. This chapter is meant to provide a broad overview of LiDAR technology, highlighting the most common applications from spaceborne platforms. It describes the versatility of LiDAR, not only as a remote sensing technique, but also as a method of enabling and supporting other remote sensing techniques and satellite applications. LiDAR, despite originating roughly at the same time as Radar, is not yet as mature as Radar or other forms of remote sensing. However, exponential development of its enabling technologies (lasers, photodetectors, positioning and attitude sensor) as well as LiDAR data processing algorithms over the last two decades is speeding its maturation process. As is the case with any other technology, further technical developments will enable new applications, and even when there is a much room for the development of LiDAR on its own, and a great deal of progress is also expected from a synergistic approach of combining it with other forms of active and passive remote sensing techniques.

Cross-References

- ▶ Astronaut Photography: Handheld Camera Imagery from Low Earth Orbit
- ▶ Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing
- ▶ Electro-optical and Hyper-spectral Remote Sensing
- ▶ Introduction and History of Space Remote Sensing
- ▶ Operational Applications of Radar Images

References

W. Abdalati, H.J. Zwally, R. Bindschadler, B. Csatho, S.L. Farrell, H.A. Fricker, D. Harding, R. Kwok, M. Lefsky, T. Markus, A. Marshak, T. Neumann, S. Palm, B. Schutz, B. Smith, J. Spinhirne, C. Webb, The ICESat-2 laser altimetry mission. Proc. IEEE **98**(5), 735–751 (2010)

J.B. Abshire, X. Sun, H. Riris, J. M. Sirota, J. F. McGarry, S. Palm, D. Yi, P. Liiva, Geoscience laser altimeter system (GLAS) on the ICESat mission: on-orbit measurement performance. Geophys. Res. Lett. **32**, L21S02 (2005)

C.O. Alley, P.L. Bender, R.F. Chang, D.G. Currie, R.H. Dicke, J.E. Faller, W.M. Kaula, G.J.F. MacDonald, J.D. Mulholland, H.H. Plotkin, S.K. Poultney, D.T. Wilkingson, Irwin Winer,

Walter Carrion, Tom Johnson, Paul Spadin, Lloyd Robinson, E. Joseph Wampler, Donald Wiebrr, E. Silverberg, C. Steggerda, J. Mullendore, J. Bayner, W. Williams, Brian Warner, Harvey Richardson, and B. Bopp, Laser ranging retroreflector. Section 7, of Apollo 11 Preliminary Science Report. NASA SP 214, 1969

A. Ansmann, U. Wandinger, O. Le Rille, D. Lajas, A.G. Straume, Particle backscatter and extinction profiling with the spaceborne high-spectral-resolution Doppler lidar ALADIN: methodology and simulations. Appl. Opt. **46, is. 26**, 6606–6622 (2007)

M.D. Behn, M.T. Zuber, A comparison of ocean topography derived from the Shuttle Laser Altimeter-01 and TOPEX/POSEIDON. IEEE Trans. Geosci. Rem. Sens. **38**(3), 1425–1438 (2000)

P.L. Bender, D.G. Currie, R.H. Dickey, D.H. Eckhardt, J.E. Faller, W.M. Kaula, J.D. Mulholland, H.H. Plotkin, S.K. Poultney, E.C. Silverberg, D.T. Wilkinson, J.G. Williams, C.O. Alley, The lunar ranging experiment. Science **182**(4109), 229–238 (1973). New Series

J.L. Bufton, Laser altimetry measurements from aircraft and spacecrat. Proc. IEEE **77**(3), 463–477 (1989)

C.C. Carabajal, D.J. Hardin, S. B. Luthcke, W. Fong, S. C. Rowton, J.J. Frawley, Processing of shuttle laser altimeter range and return pulse data in support of SLA-02, in *Proceedings of the ISPRS Workshop Mapping Surface Structure and Topography by Airborne and Spaceborne Lasers*, Portland, 1999

W.E. Carter, The lunar laser ranging pointing problem. Unpublished doctoral dissertation, University of Arizona, Tucson, 1973

W.E. Carter, R. L. Shrestha, K.C. Slatton, Geodetic laser scanning. Phys Today. **60**(12) (2007)

J.F. Cavanaugh, J.C. Smith, X. Sun, A.E. Bartels, L. Ramos-Izquierdo, D.J. Krebs, J.F. McGarry, R. Trunzo, A.M. Novo-Gradac, J.L. Britt, J. Karsh, R.B. Katz, A.T. Lukermire, R. Szymkiewicz, D.L. Berry, J.P. Swinski, G.A. Neumann, M.T. Zuber, D. Smith, The Mercury laser altimeter instrument for the MESSENGER mission. Space Sci Rev **131**(1–4), 451–479 (2007)

M.L. Chanin, A. Hauchecorne, C. Malique, D. Nedeljkovic, J.E. Blamont, M. Desbois, G. Tulinov and V. Melnikov, "Premiers résultats du lidar Alissa embarqué à bord de la station Mir," Comptes Rendus de l'Académie des Sciences – Series IIA – Earth and Planetary Science. **328**(6), 359–366 (1999)

T.D. Colea, A.F. Chenga, M. Zuberb and D. Smith, "The Laser Rangefinder on the near Earth Asteroid Rendezvous spacecraft," Acta Astronautica, in *Second IAA International Conference on Low-Cost Planetary Missions*, Laurel, vol. 39, Issue no 1–4, pp. 303–313, July-August 1996

T.K. Cossio, K.C. Slatton, W.E. Carter, K.Y. Shrestha, D. Harding, Predicting small target detection performance of Low-SNR airborne LiDAR. IEEE J. Select. Topics Appl. Earth Observ. Rem. Sens. **3**(4), 672–688 (2010)

A. Deslauriers, I. Showalter, A. Montpool, R. Taylor, I. Christie, "Shuttle TPS inspection using triangulation scanning technology," Spaceborne sensors II. Proc. SPIE **5798**, 26–33 (2005)

J.O. Dickey, P.L. Bender, J.E. Faller, X.X. Newhall, R.L. Ricklefs, J.G. Ries, P.J. Shelus, C. Veillet, A.L. Whipple, J.R. Wiant, J.G. Williams, C.F. Yoder, Lunar laser ranging: a continuing legacy of the apollo program. Science **265**(5171), 482–490 (1994). New Series

A. Donnellan, P. Rosen, J. Ranson, H. Zebker, Deformation, ecosystem structure, and dynamics of ice (DESDynI), in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, IGARSS*, Honolulu, 2008

E. Dupuis, J.C. Piedboeuf and E. Martin, Canadian activities in intelligent robotic systems: an overview, in Proceedings of International Symposium on Artificial Intelligence, Robotics and Automation in Space, Hollywood, CA, USA, Feb 2008

J.J. Degnan, 30 Years of SLR (invited paper), Proc. of the 9th International Workshop on Laser Ranging Instrumentation, Australian Government Publishing Service, Canberra, p. 8, 1994, http://ilrs.gsfc.nasa.gov/docs/ThirtyYearsOfSatelliteLaserRanging.pdf

Y. Durand, A. Hélière, P. Bensi, J.-L. Bézy, and R. Meynart, Lidars in ESA's earth explorer missions, in *14th Coherent Laser Radar Conference*, Snowmass, 2007

C. English, S. Zhu, C. Smith, S. Ruel, I. Christie, Tridar: a hybrid sensor for exploiting the complimentary nature of triangulation and LIDAR technologies, in *The 8th International Symposium on Artificial Intelligence, Robotics and Automation in Space*. ed. by B. Battrick. ESA SP-603. European Space Agency, München, 2005

J.C. Fernandez-Diaz, Scientific applications of the mobile terrestrial laser scanner (M-TLS) system, M.S. thesis, Department of Civil Engineering, University of Florida, Gainesville, Florida, 2007, http://purl.fcla.edu/fcla/etd/UFE0021101. Accessed Feb 2011

G. Fiocco, L.D. Smullin, Detection of scattering layers in the upper atmosphere (60–140 km) by optical radar. Nature **199**, 1275–1276 (1963)

J. Garvin, J. Bufton, J. Blair, D. Harding, S. Luthcke, J. Frawley, D. Rowlands, Observations of the Earth's topography from the shuttle laser altimeter (SLA): laser-pulse echo-recovery measurements of terrestrial surfaces. Phys. Chem. Earth **23**(9–10), 1053–1068 (1998)

Hamamatsu Corporation, Photon counting using photomultiplier tubes, 2005, http://sales.hamamatsu.com/assets/applications/ETD/PhotonCounting_TPHO9001E04.pdf

D.J. Harding, D.B. Gesch, C.C. Carabajal, and S.B. Luthcke, Application of the shuttle laser altimeter in an accuracy assessment of GTOP30, a global 1-kilometer digital elevation model, in *Proceedings of the ISPRS Workshop Mapping Surface Structure and Topography by Airborne and Spaceborne Lasers*, Portland, Nov 1999

D.W. Harris, J.H. Berbert, NASA/MOTS optical observations of the ANNA 1B satellite, NASA Technical Note D-3174, Jan 1966

W.A. Heiskanen, H. Moritz, *Physical Geodesy* (Freeman, San Francisco, 1967)

E.O. Hulburt, Observations of a searchlight beam to an altitude of 28 kilometers. J. Opt. Soc. Am. **27**, 377–382 (1937)

A. Javan, W.R. Bennett, D.R. Herrott, Population inversion and continuous optical maser oscillation in a gas discharge containing a He–Ne mixture. Phys. Rev. Lett. **6**, 106–110 (1961)

E.A. Johnson, R.C. Meyer, R.E. Hopkins, W.H. Mock, The measurement of light scattered by the upper atmosphere from a search-light beam. J. Opt. Soc. Am. **29**, 512–517 (1939)

JPL, Rover camera instrument description, 1997, http://starbase.jpl.nasa.gov/mpfr-m-rvreng-2_3-edr_rdr-v1.0/mprv_0001/document/rcinst.htm. Accessed Feb 2011

K. Kaufmann, Choosing your Detector, OE Magazine, March 2005

W.M. Kaula, G. Schubert, R.E. Lingenfelter, W.L. Sjogren, W.R. Wollenhaupt, Apollo laser altimetry and inferences as to lunar structure, *in Lunar Science Conference, 5th, Houston, Tex., 18 Mar 1974, Proceedings*, vol. 3, (A75-39540 19-91) Pergamon Press, New York, pp. 3049–3058, 1974

L. Le Hors, Y. Toulemont, A. Hélière, *"Design and Development of the Backscatter Lidar Atlid for Earthcare", proceedings of the International Conference on Space Optics* (Toulouse, France, 2008)

G.G. Matvienko, V.E. Zuev, V.S. Shamanaev, G.P. Kokhanenko, A.M. Sutormin, A.I. Buranskii, S.E. Belousov, A.A. Tikhomirov, Lidar BALKAN-2 for the space platform ALMAZ-1B, Lidar Techniques for Remote Sensing, in *Proceedings of SPIE*, vol. 2310, 1994, http://spie.org/x648.html?product_id=195859

F.J. McClung, R.W. Hellwarth, Giant optical pulsations from Ruby. J. Appl. Phys. **33**(3), 828–829 (1962)

J. McGarry, T. Zagwodzki, A brief history of satellite laser ranging: 1964 – present. Published by the Crustal Dynamics Data Information System (CDDIS), NASA Goddard Space Flight Center, Greenbelt, Maryland (2005) http://cddis.nasa.gov/slr2000/docs/gsfcslr_jm0504.pdf. Accessed Feb 2011

J.K. Miller, A.S. Konopliv, P.G. Antreasian, J.J. Bordi, S. Chesley, C.E. Helfrich, W.M. Owen, T.C. Wang, B.G. Williams, D.K. Yeomans, D.J. Scheeres, Determiantion of shape, gravity and rotational state of Asteroid 433 Eros. Icarus **155**, 3–17 (2002)

NASA, Mars topography, http://mola.gsfc.nasa.gov/topography.html

NASA, The space shuttle's return to flight, mission STS-114 press kit (2005)

National Research Council, *"Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond," Committee on Earth Science and Applications from Space: A Community Assessment and Strategy for the Future* (The National Academies Press, Washington, DC, 2007). ISBN 0-309-10387-8

NEPTEC, Tridar, http://www.neptec.com/media/brochures/Canadian/Space-TriDAR.pdf

M. Nimelman, J. Tripp, A. Allen, D.M. Hiemstra, S.A. McDonald, Spaceborne scanning lidar system (SSLS) upgrade path. Proc. SPIE **6201**, 62011V-1–62U11V-10 (2006)

J.C. Piedboeuf, E. Martin, M. Doyon, On-orbit servicing in Canada: advanced developments and demonstrations, in *Proceedings of the 8th ESA Workshop on Advanced Space Technologies for Robotics and Automation ASTRA*, Noordwick, 2004

L. Ramos-Izquierdo, V.S. Scott III, J. Connelly, S. Schmidt, W. Mamakos, J. Guzek, C. Peters, P. Liiva, M. Rodriguez, J. Cavanaugh, H. Riris, Optical system design and integration of the Lunar orbiter laser altimeter. Appl. Opt. **48**, 3035–3049 (2009)

F.I. Robertson, W.M. Kaula, Apollo 15 laser altimeter, Part D, Section 25, Apollo 15 Preliminary Science Report. NASA SP-289, 1972

B.E. Schutz, H.J. Zwally, C.A. Shuman, D. Hancock, J.P. DiMarzio, Overview of the ICESat mission. Geophys. Res. Lett. **32**, L21S01 (2005)

J. Shan, C.K. Toth (eds.), *Topographic Laser Ranging and Scanning: Principles and Processing* (CRC Press, Boca Raton, 2009)

H. Simons, Secret mapping by satellite, New Sci. **21**(381) (1964)

D.E. Smith, M.T. Zuber, G.A. Neumann, F.G. Lemoine, Topography of the Moon from the Clementine lidar. J Geophys. Res. **102**(E1), 1591–1611 (1997)

D.E. Smith, M.T. Zuber, H.V. Frey, J.B. Garvin, J.W. Head, D.O. Muhleman, G.H. Pettengill, R.J. Phillips, S.C. Solomon, H.J. Zwally, W.B. Banerdt, T.C. Duxbury, M.P. Golombek, F.G. Lemoine, G.A. Neumann, D.D. Rowlands, O. Aharonson, P.G. Ford, A.B. Ivanov, C.L. Johnson, P.J. McGovern, J.B. Abshire, R.S. Afzal, X. Sun, Mars orbiter laser altimeter: experiment summary after the first year of global mapping of Mars. J. Geophys. Res **106**(E10), 23689–23722 (2001)

D.E. Smith, M.T. Zuber, G.A. Neumann, F.G. Lemoine, E. Mazarico, M.H. Torrence, J.F. McGarry, D.D. Rowlands, J.W. Head III, T.H. Duxbury, O. Aharonson, P.G. Lucey, M.S. Robinson, O.S. Bamouin, J.F. Cavanaugh, X. Sun, P. Liiva, D. Mao, K.C. Smith, A.E. Bartels, Initial observations from the Lunar Orbiter Laser Altimeter (LOLA). Geophys. Res. Lett. **37**, L18204 (2010)

C.L. Smithpeter, R.O. Nellums, S.M. Lebien, G. Studor, Miniature high-resolution laser radar operating at video rates. Proc. SPIE **4035**, 279–286 (2000)

L.D. Smullin, G. Fiocco, Optical echoes from the Moon. Nature **194**, 1267 (1962)

STS-105 Shuttle press kit, 2001, http://www.shuttlepresskit.com/sts-105/index.htm

G. Suna, K.J. Ranson, V.I. Kharuk, K. Kovacs, Validation of surface height from shuttle radar topography mission using shuttle laser altimeter. Rem. Sens. Environ. **88**(4), 401–411 (2003)

E.H. Synge, Phil. Mag. **9**, 1014–20 (1930)

N. Taylor, *LASER: The Inventor, the Nobel Laureate, and the Thirty-Year Patent War* (Simon & Schuster, New York, 2000)

The International Laser Ranging Service, http://ilrs.gsfc.nasa.gov/. Accessed February 2011

N. Thomasa, T. Spohnb, J.-P. Barriotc, W. Benza, G. Beutlerd, U. Christensene, V. Dehantf, C. Fallnichg, D. Giardinih, O. Groussini, K. Gundersona, E. Hauberb, M. Hilchenbache, L. Iessj, P. Lamyk, L.-M. Laral, P. Lognonnem, J.J. Lopez-Morenol, H. Michaelisb, J. Oberstb, D. Resendesn, J.-L. Reynaudk, R. Rodrigol, S. Sasakio, K. Seiferlina, M. Wieczorekm, J. Whitbya, The BepiColombo laser altimeter (BELA): concept and baseline design. Planet. Space Sci. **55**(10), 1398–1413 (2007)

M.A. Tuve, E.A. Johnson, O.R. Wulf, A new experimental method for study of the upper atmosphere. J. Terrest. Magnet. **40**, 452–454 (1935)

J.R. Vetter, "Fifty years of orbit determination, development of modern astrodynamic methods, Johns Hopkins APL Tech. Dig. **27**, 3 239–252 (2007)

U. Wandinger, Introduction to Lidar, in *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere*, ed. by C. Weitkamp (Springer, New York, 2005), pp. 1–18

M.J. Weber (ed.), *Handbook of LASERS* (CRC Press, Baco Raton, 2001). ISBN 978-1-4200-5017-2

C. Weitkamp (ed.), *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere* (New York, Springer, 2005)

C. Werner, G. Kokhanenko, G. Matvienko, V. Shamanaev, Y. Grachjov, I. Znamenskii, U.G. Oppel, Spaceborne laser rangefinder "LORA" used as a cloud lidar. Opt. Rev. **2**(3), 221–224 (1995)

C. Werner, Spaceborne lidar mission, past and future, in *Proceedings Conference on Lasers and Electro-optics Europe, CLEO/Europe*, Hamburg, pp. 212, Sep 1996

J. Whiteway, J.M. Daly, A. Carswell, T. Duck, C. Dickinson, L. Komguem, C. Cook, Lidar on the Phoenix mission to Mars. J. Geophys. Res. **113**, E00A08, 2008

J. Whiteway, L. Komguem, C. Dickinson, Observations of Mars atmospheric dust and clouds with the Lidar instrument on the phoenix mission, in *Abstract on the Forth International Workshop on the Mars Atmosphere: Modeling and Observations*, Paris, Feb 2011

D.M. Winker, R.H. Couch, M.P. McCormick, An overview of LITE: NASA's Lidar in-space technology experiment. Proc. IEEE **84**(2), 164–180 (1996)

D.M. Winker, W.H. Hunt, C.A. Hostetler, Status and performance of the CALIOP lidar, in *Laser Radar Techniques for Atmospheric Sensing* (*Proceedings of the SPIE*), vol. 5575, ed. by U.N. Singh, pp. 8–15, Maspalomas, Gran Canaria, Spain, 2004

W.R. Wollenhaupt, W.L. Sjogren, Apollo 16 laser altimeter, Chapter 30, Part A, Apollo 16 Preliminary Science Report SP-315, 1972

W.R. Wollenhaupt, W.L. Sjogren, R.E. Lingenfelter, G. Schubert, and W.M. Kaula, Apollo 17 laser altimeter, Chapter 33, Part E, Apollo 17 Preliminary Science Report SP-330, 1973

A.W. Yua, M.A. Stephen, S.X. Li, G.B. Shawa, A. Seasa, E. Dowdyea, E. Troupakib, P. Liivab, D. Pouliosc, K. Mascetti, Space laser transmitter development for ICESat-2 mission. Proc. SPIE **7578**, 757–809 (2010)

M.T. Zuber, D.E. Smith, A.F. Cheng, J.B. Garvin, O. Aharonson, T.D. Cole, P.J. Dunn, Y. Guo, F.G. Lemoine, G.A. Neumann, D.D. Rowlands, M.H. Torrance, The shape of 433 Eros from the NEAR-shoemaker laser rangefinder. Science **289**, 2097 (2000)

M.T. Zuber, D.E. Smith, S.C. Solomon, R.J. Phillips, S.J. Peale, J.W. Head III, S.A. Hauck II, R.L. McNutt Jr., J. Oberst, G.A. Neumann, F.G. Lemoine, X. Sun, O. Barnouin-Jha, J.K. Harmon, Laser altimeter observations from MESSENGER's first Mercury Flyby. Science **321**, 77 (2008)

Digital Image Acquisition: Preprocessing and Data Reduction

32

Siamak Khorram, Stacy A. C. Nelson, Halil Cakir, and Cynthia F. van der Wiele

Contents

Introduction 810
Data Acquisition and Policies 812
Satellite and Airborne Data Resolution and Formats 812
Image Processing Tools 822
Image Display 822
Image Preprocessing 823
Registration and Coordinate Systems 826
Data Reduction and Fusion Techniques 828
International Policies Governing Remotely Sensed Data 833
Notes 835
References 835

S. Khorram (✉)
Department of Environmental Science, Policy, and Management, University of California, Berkeley, 94720, CA, USA
and
North Carolina State University, Raleigh, 27695-7106, NC, USA
e-mail: khorram@berkeley.edu

S.A.C. Nelson
Center for Earth Observation, North Carolina State University Campus Box 7106; 5123 Jordan Hall, 27695, Raleigh, NC, USA
e-mail: Stacy_Nelson@ncsu.edu

H. Cakir
Air Quality Analysis Group/AQAD/OAQPS, US Environmental Protection Agency, 109 T. W. Alexander Drive, Research Triangle Park, NC, 27709
e-mail: Cakir.Halil@epa.gov

C.F. van der Wiele
Cynthia Van Der Wiele and Associates, LLC, 715 Shepherd Street, 27701, Durham, NC, USA
e-mail: cynthia.vanderwiele@alumni.duke.edu

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_46, © Springer Science+Business Media New York 2013

Abstract

The main objective of this chapter is to focus on the digital preprocessing and data reduction techniques as applied to remotely sensed data for the purpose of extracting useful Earth resources information. The image processing and post-processing tools are described in the next chapter. The concepts discussed in this chapter include:

- Image acquisition considerations including currently available remotely sensed data
- Image characteristics in terms of spatial, spectral, radiometric, and temporal resolutions
- Preprocessing techniques such as geometric distortion removals, atmospheric correction algorithms, image registration, enhancement, masking, and data transformations
- Data reduction, fusion, and integration techniques
- International policies governing acquisition and distribution of remotely sensed data

Keywords

Data fusion • Digital image processing • Digital integration • Electromagnetic spectrum • Hyperspectral imaging • Light detection and ranging (LiDAR) • Multispectral imaging • Pixel • Preprocessing • Radio detection and ranging (RADAR) • Radiometric resolution • Satellite remote sensing • Spatial resolution • Spectral resolution • Temporal resolution

Introduction

Remote sensing can be defined as the acquisition and measurement of data/information of one or more properties of a phenomenon, object, or material by a recording device not in physical contact with the feature(s) under surveillance.

Remote sensing involves the collection of data in digital or analog forms (e.g., aerial photographs and videos) by space-based instruments or sensors without any physical contact. In general, remotely sensed data are collected from airborne platforms such as satellites, with brightness values at various wavelength regions of the electromagnetic spectrum in a variety of spectral, spatial, radiometric, and temporal resolutions. It should be mentioned that most analog products can be digitized for digital processing with certain limitations.

Satellites collect sensed data in raw or unprocessed form. The raw spectral data or radio signals must be processed or enhanced in order to produce images or other products. This typically involves computer-aided digital processing, although visual image interpretation is still used for some applications. The advantages of digital processing include: (1) the ability to go well beyond the limits of the human eye, which is up to 32 gray levels; most data are in 1,024 gray levels or higher; (2) the speed of processing and extracting information over large areas; (3) the lower cost for processing remotely sensed data; (4) the repeatability of producing



Fig. 32.1 Digitally processed image of the world (Figure courtesy of NASA)

the same results when using the same algorithms; and (5) the nature of having all inputs, outputs, and intermediate results in digital form, which makes it possible to combine a variety of data types for solving complex problems and performing geospatial modeling.

The spectacular portrait in Fig. 32.1, reminiscent of the famous Apollo-era images of Earth taken by lunar astronauts, is of the Western Hemisphere during one of the strongest hurricanes ever recorded in the Eastern Pacific. It was produced through the utilization of digital image processing techniques. This combination of science, engineering, and artistry was generated by researchers in the Laboratory for Atmospheres at NASA's Goddard Space Flight Center, Greenbelt, Maryland, by digital processing of data from three different Earth observing satellite instruments.

Many scientists and engineers have contributed significantly to advancing image processing techniques as applied to Earth resources, planetary explorations and studies, and medical imaging and to their scientific definition. These include the various individuals and their relevant books and articles as referenced in end notes (Swain and Davis 1978; Dobson et al. 1995; Jähne 1991; Dai and Khorram 1997; Lunetta and Elvidge 2000; Russ 2002; Herold et al. 2003; Jenser 2005; Lillesand et al. 2008; Yuan et al. 2009).

Thematic maps, color composites, and color-coded classified images are some of the most characteristic output products from digital formats. A variety of image

enhancement and visualization techniques can be applied to these output products in order to extract useful information. Fortunately, many commercial software packages have been developed specifically for processing satellite imagery.

Data Acquisition and Policies

Satellite and Airborne Data Resolution and Formats

Over the last decade, the number of airborne and space-borne remote sensing platforms and sensors has increased dramatically. While airborne platforms have emerged to become much more customizable and application-specific tools, remote sensing satellites (space borne) offer the distinct advantage of synoptic acquisition of large and/or remote areas with repetitive revisit cycles.

Airborne platforms normally involve the capture of features on the Earth's surface by means of photographic cameras (air photos), video, spectro-radiometers, and Radio Detection and Ranging (RADAR) apparatus. Film and video-based air photos normally fall into two categories that are based on the orientation of the camera's optical axis with respect to the Earth's surface (Avery 1992). The first category is known as vertical air photos. Vertical air photos are photographic images captured from a camera angle that is either vertical (90°) or near vertical (90° ± 3°) to the Earth's surface. Oblique air photos are photographic images captured from camera angles that exceed an angular amount of 20° from being vertical to the Earth's surface. Both vertical and oblique air photos may incorporate photographic images created from color-, black and white-, color inferred-, and video-based film.

The variety of the space-borne platforms includes two main categories: optical/infrared satellite platforms and microwave satellite platforms. Optical/infrared satellites operate with the visible, near infrared and short-wave infrared portions of the electromagnetic spectrum (EMS). These sensors form images of the Earth's surface by capturing varied amounts of the Sun's radiation that is reflected from objects on the ground. Microwave satellites, either passive or active, record longer EMS wavelengths (radio waves) compared to those of optical sensors. Figure 32.2 illustrates the electromagnetic spectrum.

Specialized airborne platforms include spectro-radiometer applications developed and employed by Federal, private, and commercial enterprises. These sensors take advantage of capturing narrow bands of light reflectance within specified portions of the electromagnetic spectrum (EMS). The resultant image scale or spatial resolution can be controlled by the aircraft's altitude. Two examples of specialized airborne platforms include the NASA Stennis Space Center's Airborne Terrestrial Applications Sensor (ATLAS) and the Real-time Data Acquisition Camera System (RDACS) which are both aircraft-deployable remote sensing systems. The ATLAS system is a multispectral instrument, flown onboard a NASA Lear Jet, that is capable of collecting data on 12 channels ranging from 0.45 to 12.2 μm. The NASA RDACS sensor is a 120-band multispectral airborne

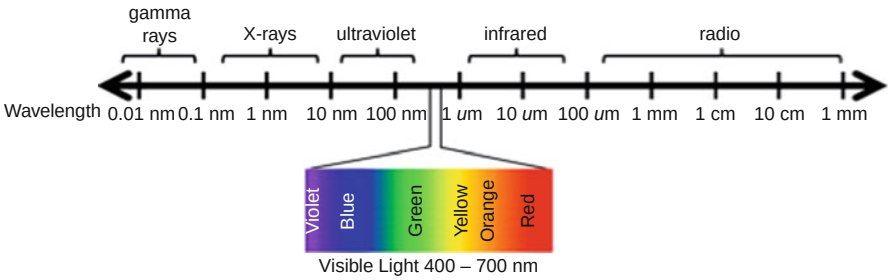


Fig. 32.2 The electromagnetic spectrum

system made up of an array of three digital cameras mounted underneath a single engine airplane. Each camera of the RDACS sensor is fitted with a filter that determines the camera wavelength sensitivity.

Airborne as well as space-borne platforms may also include sensors that operate within the microwave portion of the EMS. This portion of the spectrum incorporates wavelengths from approximately 1 mm to 1 m (Lillesand et al. 2008).

RAdio Detection And Ranging (RADAR) sensors are active microwave sensors that utilize emitted radio waves which give day or night, all weather (through cloud cover, smoke, and haze) imaging capability. This is an extremely important capability during disasters. Thus, RADAR sensors find wide applications in disaster management.

RADAR and optical sensors are complimentary to each other. In other words, they utilize the electromagnetic spectrum in a way to provide different but not competing information about the feature or the phenomenon observed. For example, in mineral deposit exploration, while hyperspectral imagery exploits the different spectral reflectance properties of the minerals for explorations, RADAR data exploits the structural geology by highlighting faults, folds, synclines and anticlines, and lineaments.

General application areas of RADAR data are summarized below:

- Geology: topography and topographic changes, mapping surface expressions such as fault lines and movements
- Hazards and disaster management: flooding, volcanoes, earthquakes, marine oil spill, river ice buildup
- Oceanography: ocean currents, winds, and surface features, sea ice thickness and coastal processes (usually these have important implications for naval navigation and off-shore oil explorations)
- Hydrology: soil moisture, drainage network, land-water delineation
- Ecology: land cover classification, biomass measurements, monitoring deforestation (especially in rainforests since optical sensors are limited because of the cloud cover)
- Ice sheets: extent, icebergs, surface morphology, ice velocity, and surface topography

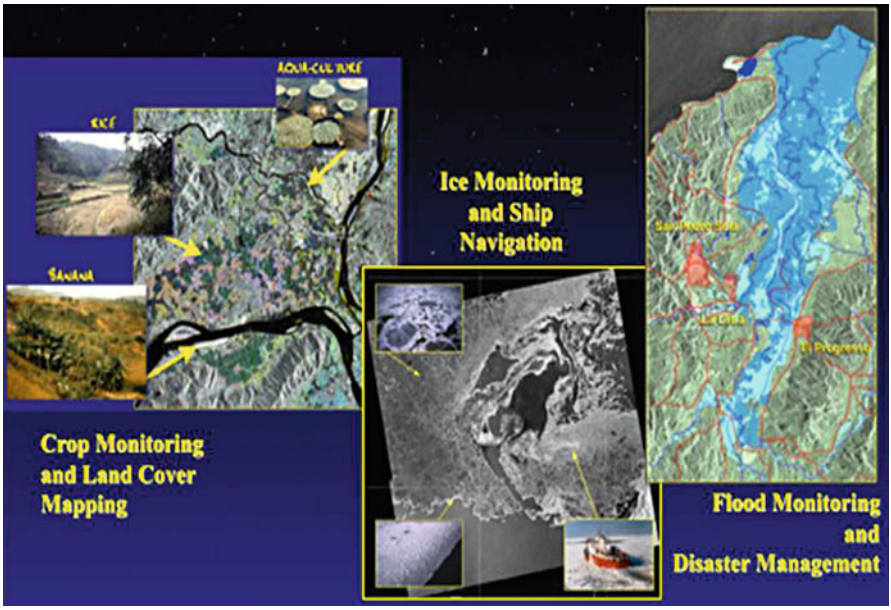


Fig. 32.3 Examples of some of the common applications of RADAR data from RADARSAT, Canada (Graphics courtesy of the Canadian Space Agency)

Figure 32.3 illustrates some examples of passive RADAR data applications (courtesy of Shannon Ross, Canadian Center for Remote Sensing). Light Detection And Ranging (or LiDAR) is also an active senor that is usually operated from an airborne platform. LiDAR is similar to RADAR in that they both are considered active remote sensing technologies. That is, the images are created from returning energy originally emitted from the senor. However, unlike RADAR, LiDAR takes advantage of ultraviolet, visible, or near infrared light to image objects.

Light energy (photons) is emitted from the LiDAR sensor in pluses; images of the ground are created by calculating the time of the pulse return, usually in the form of an elevation dataset (Fig. 32.4).

This 3-m digital elevation map is created by a perennial stream channel derived from LiDAR data. The black line in Fig. 32.4 depicts the channel's center line modeled from LiDAR and field-based applications.

Space-borne remote sensing platforms, given the altitude advantages over airborne platforms, are often geared toward applications of mapping, monitoring, and analyzing features or changing conditions occurring on (or just below) the Earth's surface. A global inventory of the current major Earth observing satellite systems is provided in Table 32.1.

Airborne platforms provide a diversity of image formats (i.e., air photos, video, spectral images, radar images, etc.). In fact, air photos represent the longest historic record of monitoring and capturing images of the Earth's surface from distance.

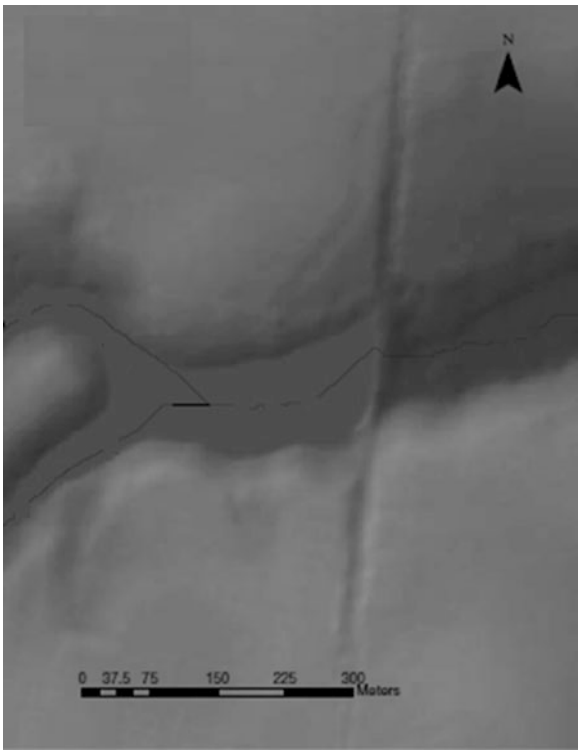


Fig. 32.4 Illustration of a 3-m resolution digital elevation map (DEM) (Graphic courtesy of University of North Carolina State)

However, the majority of spectral remote sensing images acquired from space-borne platforms are not photographs but rather digital images that are recorded within very specific ranges (or bands) of the electromagnetic spectrum. The EMS is made up of wavelengths of light or natural energy that extends from very short wavelengths (gamma rays and X-rays), through the section of visible light that humans use to see (blue, green, red), to the much longer wavelengths of infrared and radio waves (see Fig. 32.2).

Spectral images that are captured by the sensor are made up of picture elements known as *pixels*. Within an image, pixels that make up images are oriented in horizontal rows and vertical columns. Each pixel contains the reflectance data of specific ranges (or bands) of the electromagnetic spectrum that the sensor was designed to capture. Each pixel corresponds to the particular location on the Earth's surface from where the image was collected. Additionally, the ground spatial resolution of the sensor corresponds to the individual pixel size that each image is comprised of. For example, the spatial resolution of the multispectral scanner (ETM+) aboard the LANDSAT-7 satellite is 30 m. This means that of the thousands of pixels that make up a 185 km (across track) by 180 km (along track) LANDSAT-7 scene, each pixel will have a spatial resolution corresponding to a 30 m by 30 m area on the ground.

Table 32.1 Global summary of major Earth observing satellites

Satellite platform/ program	Launch date	Spatial resolution (m)	Spectral bands (μm)	Revisit cycles (days)
Landsat 7 (EMT+)	April 15, 1999	15 m	Pan: 0.48–0.68	16
		30 m	1: 0.45–0.52	
		60 m	2: 0.53–0.61	
			3: 0.63–0.69	
			4: 0.75–0.90	
			5: 1.55–1.75	
			6: 10.40–12.50 (thermal)	
			7: 2.09–2.35	
SPOT-5	May 3, 2002	2.5–5	Pan: 0.48–0.71	26/2–3
		10–20	1: 0.50–0.59	
			2: 0.61–0.68	
			3: 0.78–0.89	
			4: 1.58–1.75	
ALOS	January 24, 2006	PRISM	Pan	46/2
		2.5 m	1: 0.42–0.50	
		AVNIR-2	2: 0.52–0.60	
		PALSAR	3: 0.61–0.69	
		10 m	4: 0.76–0.89	
			L-Band	
CBERS-3	2011	AWFI	1: 0.45–0.52	5
		60 m	2: 0.52–0.59	26
		IRMSS	3: 0.63–0.69	
		40 m	4: 0.77–0.89	
		80 m	Pan: 0.50–1.10	
			2: 1.55–1.75	
			3: 2.08–2.35	
			4: 10.40–12.50	
IKONOS	September 24, 1999	1	Pan: 0.45–0.90	3
		4	1: 0.45–0.52	
			2: 0.52–0.60	
			3: 0.63–0.69	
			4: 0.76–0.90	
QuickBird	December 5, 2000	0.61	Pan: 0.45–0.90	
		2.40	1: 0.45–0.52	
			2: 0.52–0.60	
			3: 0.63–0.69	
			4: 0.76–0.90	

(*continued*)

Table 32.1 (continued)

Satellite platform/ program	Launch date	Spatial resolution (m)	Spectral bands (μm)	Revisit cycles (days)
GeoEye-1 (formerly Orbview-5)	September 6, 2008	0.41	Pan: 0.45–0.90	2.1–8.3
		1.65	1: 0.45–0.80	
			2: 0.45–0.51	
			3: 0.51–0.58	
			4: 0.66–0.69	
			5: 0.78–0.92	
SeaWiFS (Orbview-2)	August 1, 1997	1,100	1: 0.40–0.42	1
			2: 0.43–0.45	
			3: 0.48–0.50	
			4: 0.50–0.52	
			5: 0.56–0.57	
			6: 0.66–0.68	
			7: 0.75–0.79	
			8: 0.85–0.89	
WorldView-2	October 8, 2009	0.46	Pan: 0.45–0.80	1.1
		1.84	1: 0.40–0.45	
			2: 0.45–0.51	
			3: 0.51–0.58	
			4: 0.59–0.63	
			5: 0.63–0.69	
			6: 0.71–0.75	
			7: 0.77–0.90	
			8: 0.86–1.04	
AVHRR	February 6, 2009 (latest)	1,090	1: 0.58–0.68	<1
			2: 0.73–1.00	
			3: 1.58–1.64	
			4: 3.55–3.93	
			5: 10.30–11.30	
			6: 11.50–12.50	
Meteosat-9	December 21, 2005	1,000–3,000	1: 0.40	15 min
			2: 0.60	
			3: 0.80	
			4: 1.60	
			5: 3.00	
			6: 6.20	
			7: 7.30	

(*continued*)

Table 32.1 (continued)

Satellite platform/ program	Launch date	Spatial resolution (m)	Spectral bands (μm)	Revisit cycles (days)
			8: 8.70	
			9: 9.70	
			10: 10.8	
			11: 12.0	
			12: 13.4	
RADARsat-2 (Synthetic Aperture RADAR satellite)	December 14, 2007	3–100	5.4 GHz frequency band	24
GOES-13 (Imager)	May 24, 2006 (active)	1,000–8,000	1: 0.55–075	Full-disc view of Earth, less than or equal to 26 min
			2: 3.80–4.00	
			3: 6.50–7.00	
			4: 10.20–11.20	
			5: 11.50–12.50	
Terra Earth Observing System – ASTER	December 18, 1999	15–90	1: 0.52–0.60	16
			2: 0.63–0.69	
			3: 0.76–0.86	
			4: 1.60–1.70	
			5: 2.15–2.19	
			6: 2.19–2.22	
			7: 2.24–2.29	
			8: 2.30–2.37	
			9: 2.37–2.43	
			10: 8.13–8.48	
			11: 8.48–8.83	
			12: 8.93–9.28	
			13: 10.3–11.0	
			14: 11.0–11.7	
Terra Earth Observing System – MODIS	May 4, 2002 (Aqua EOS PM-1)	250	1: 0.62–0.67	1–2
			2: 0.84–0.88	
			3: 0.46–0.48	
			4: 0.55–0.57	
			5: 1.23–1.25	
			6: 1.63–1.65	
			7: 2.11–2.16	

(*continued*)

Table 32.1 (continued)

Satellite platform/ program	Launch date	Spatial resolution (m)	Spectral bands (μm)	Revisit cycles (days)
			8: 0.41–0.42	
			9: 0.44–0.45	
			10: 0.48–0.50	
			11: 0.53–0.54	
			12: 0.55–0.56	
			13: 0.66–0.67	
			14: 0.67–0.68	
			15: 0.74–0.75	
			16: 0.86–0.88	
			17: 0.89–0.92	
			18: 0.93–0.94	
			19: 0.92–0.97	
			20–36: 3.7–14.4	
Envisat (AATSR) European Space Agency	March 1, 2002	1,000	1: 0.56	35
			2: 0.66	
			3: 0.87	
			4: 1.60	
			5: 3.70	
			6: 10.9	
			7: 12.0	
CARTOSAT 2B Indian Space Research Organization	July 12, 2010	0.80	Spectral band range: 0.50–0.85	4–5

Hyperspectral sensors also referred as "imaging spectroscopy" are passive sensors that take images of the spectral radiance using a wider range of the electromagnetic spectrum than the multispectral sensors.

Hyperspectral images may contain hundreds of bands as illustrated in Fig. 32.5. For example, while the MODIS (or Moderate Resolution Imaging Spectroradiometer) sensor, aboard the Terra (EOS AM) and Aqua (EOS PM) satellites, is acquiring data in 36 spectral bands,[1] the AVIRIS (Airborne Visible Infrared Imaging Spectrometer) instrument collects data from 224 continuous spectral channels (bands) ranging from ultraviolet to near infrared.[2]

Hyperspectral imagery finds a broad array of applications ranging from mineralogy to air quality. Some of these applications are as follows:

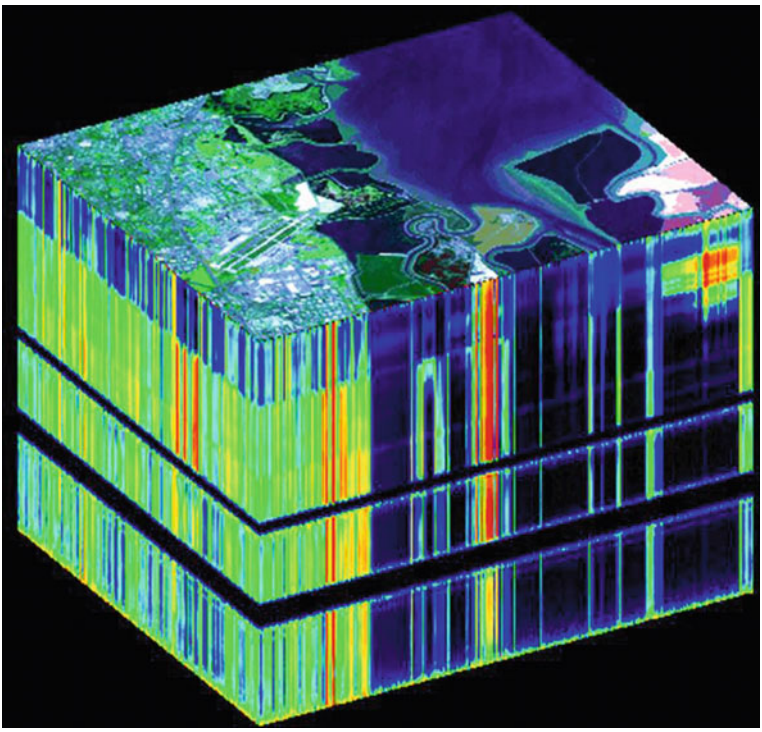


Fig. 32.5 An example of AVIRIS data that was acquired on August 20, 1992, over Moffett Field, California, at the southern end of the San Francisco Bay[3] (Image courtesy of NASA JPL and NASA Ames)

- Atmosphere: cloud properties, air quality, aerosols
- Terrestrial Systems: mineralogy, geology, ecology, agriculture
- Aquatic Systems: monitoring coral reefs, mapping chlorophyll, phytoplankton, dissolved organic materials, and suspended sediment concentrations
- Hazards: Biomass burning, flooding, hurricanes, oil spills as illustrated in Figure 32.7
- Snow/Ice: monitoring snow/ice extent, climate change

The small increments of wavelengths between each band allow detection and mapping of features, objects, and phenomena, which often is not possible by conventional airborne or satellite operational scanners. Figure 32.6 depicts an example of hyperspectral data over a terrestrial system.

Figure 32.7 illustrates and example of MODIS application for detecting the oil slick on July 14, 2010, from BP's Deep Horizon well. The largest oil slick is located in the center of the image, but a few isolated ribbons of oil are visible to the east. The lighter-colored waters around the river delta are full of sediment and do not reflect oil slicks.

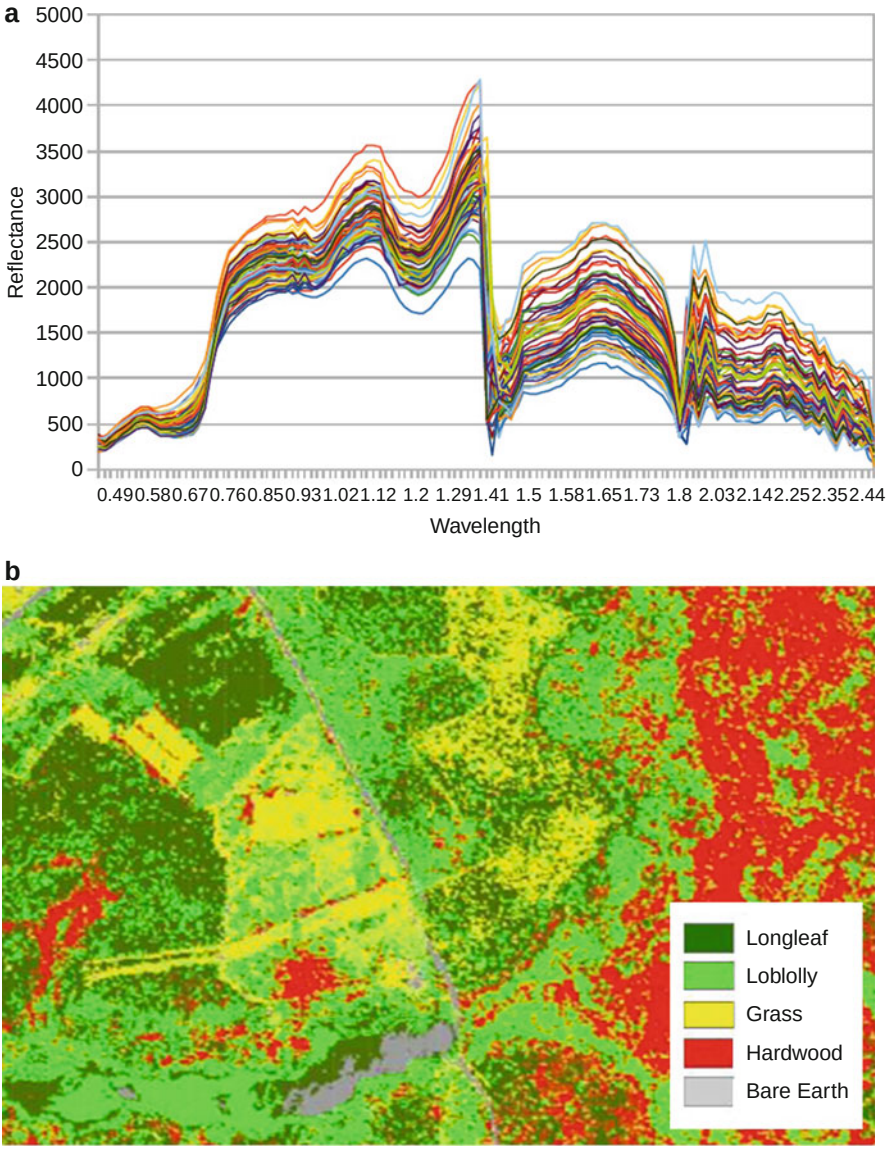


Fig. 32.6 (**a**) Illustrates the stacked spectral reflectance data collected from a hyperspectral scanner. (**b**) Illustrates the results of a land cover classification performed on the hyperspectral data resulting in five categories of interest (Image credit: Jeff Schmaltz, MODIS Rapid Response Team, NASA Goddard Space Flight Center. Text credit: Rebecca Lindsey, NASA Earth Observatory)

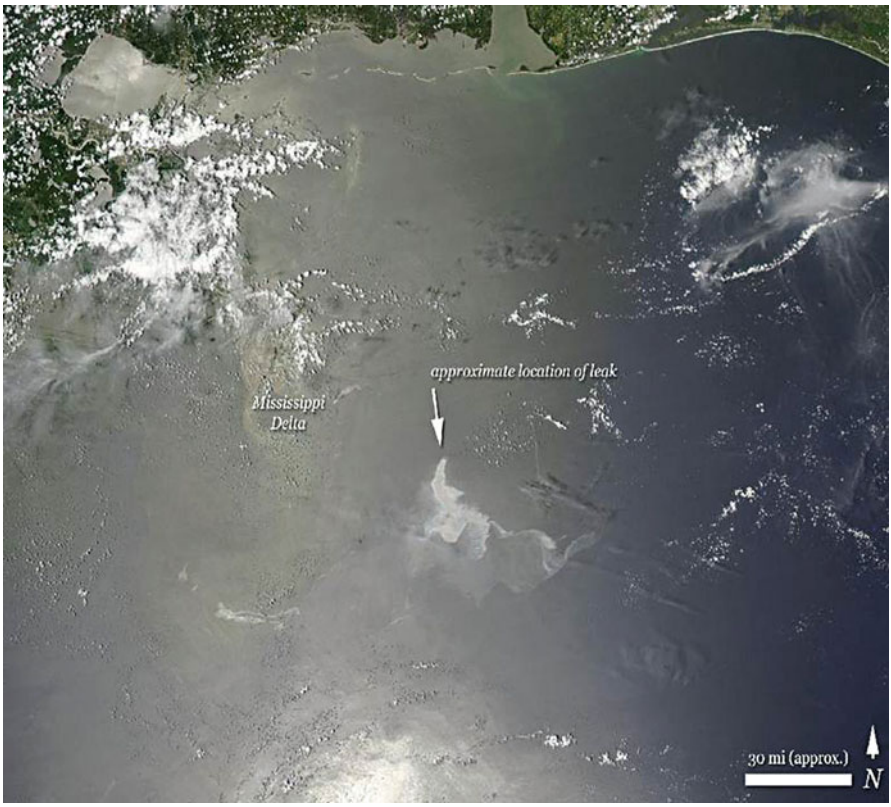


Fig. 32.7 Illustrating the BP 2010 oil spill in the Gulf of Mexico from MODIS Data (Image courtesy of NASA)

Image Processing Tools

Image Display

Images are typically displayed in True Color Composites, False Color Composites, or classified form. Raw data displays are most frequently provided as True Color Composite (TCC) and the False Color Composite (FCC). True Color Composite displays the Blue band in blue color, the Green band in green color, and the Red band in red color while False Color Composite displays any three bands in blue, green, and red colors. The standard FCC provides an Infrared band displayed in red color, the Red band displayed in green color, and the Green band displayed in blue color.

Raw data displays are most frequently provided as True Color Composite (TCC) and the False Color Composite (FCC). True Color Composite displays the Blue band in blue color, the Green band in green color, and the Red band in

red color while False Color Composite displays any three bands in blue, green, and red colors. The standard FCC provides an Infrared band displayed in red color, the Red band displayed in green color, and the Green band displayed in blue color.

Image Preprocessing

Remotely sensed data often contain various types of distortions due to less than optimal atmospheric conditions, rotation of the Earth, satellite or aircraft motion, curvature of the Earth, and the exact location of a given point within an image. The remotely sensed data, particularly in higher spatial resolutions, involve enormous amounts of data that often needs to be reduced before image processing. Various techniques designed to remove systematic and nonsystematic effects and distortions as well as most commonly accepted data reduction methodologies are discussed in this section. Image processing and post-processing are described in the following chapter.

Atmospheric Correction

Clouds, suspended particles, or other materials in the atmosphere at the time of data acquisition change the digital numbers on an image. While most terrestrial applications of remotely sensed data have looked for virtually cloud-free days and/or multiple scenes, and thus have essentially avoided or ignored this issue, haze, clouds, and other atmospheric effects over coastal and near-shore ocean areas are pervasive. Four primary methods have been developed to remove or minimize the atmospheric effects on an image.

- *Absolute radiometric correction* of atmospheric attenuation takes into account the solar zenith angle at the time of satellite overpass, the total optical thickness caused by molecular scattering, the atmospheric transmittance for a given angle of incidence, the spectral irradiance at the top of the atmosphere, and the Rayleigh and Mie scattering laws (Turner and Spencer 1972; And also Forster 1984).
- *Relative radiometric correction* of atmospheric attenuation normalizes the intensities among different bands within a scene to remove detector-related problems, and then corrects the intensities through a comparison with a standard reference surface on the same date and same scene.
- *Single-image normalization* uses a histogram adjustment for any shift in the histogram that may have been caused by atmospheric effects. This method is based on the fact that infrared data are largely free of atmospheric scattering effects as compared to the visible region. Thus, the histogram shifts due to haze can be used to adjust for the atmospheric effects. This method involves a subtractive bias established for each band and is very simple to use.
- *Multiple image normalization* uses regression analysis for a number of dates. This method is primarily used to make sure the spectral values from one date are

comparable to other dates, which implicitly takes the atmospheric corrections into account. This method is primarily used for change detection purposes and is also fairly simple to use.

Geometric Corrections

Sources of errors causing geometric distortions in remotely sensed data are divided into systematic and nonsystematic. Systematic distortions are due to image motion caused by forward movement of the spacecraft, variations in the mirror scanning rate, panoramic distortions, variations in platform velocity, and distortions due to the curvature of the Earth. Nonsystematic distortions are due to variations in satellite altitude and attitude. Much of the systematic errors are removed in commercially available remotely sensed data. The most common techniques for removing the remaining systematic and nonsystematic distortions are image-to-map rectification and image-to-image registration through the selection of a large number of ground control points.

Radiometric Corrections

The type of radiometric distortions varies among the different sensors. Typically, the Sun elevation corrections and Earth–Sun distance corrections are applied to satellite data for removing the effects of the seasonal position of the Sun relative to the Earth, and to normalize for the seasonal variations in the distance between the Earth and the Sun. Noise removal algorithms can be applied to remove unwanted disturbance due to sensing, signal quantization, and recording. Several de-striping algorithms are available to remove the striping and banding effects in satellite data. The line drops can be corrected by replacing the spectral values in the missing band with the average of the line(s) above and below them. The nonsystematic variations in gray levels from pixel to pixel (i.e., bit errors) can be corrected by replacing these values with neighboring values that exceed the threshold values established by analysts. Figures 32.8 and 32.9 depict the removal of radiometric differences before and after normalization.

Occasionally, due to a malfunction in the scanners or other issues, images contain a large number of line drops that need to be corrected. Often the digital values for these line drops are zero. A common way of correcting this issue is to replace the zero values with the average of the values for the scan lines above and below the bad scan line. A linear averaging is often sufficient for drop lines that are not more than a few pixels up or down. When the number of drop lines are more than a few pixels (usually 3–5 depending on the local scene dynamic), the linear averaging does not seem to be appropriate. In this case, either a more complicated modeling approach is considered, or these large drop lines should be left uncorrected. Figure 32.10a depicts the uncorrected image and Fig. 32.10b depicts the image after radiometric correction with one large line which was left uncorrected based on airborne Ocean Color Scanner data (Khorram 1982).

Spectral distortions may also be evident in the raw imagery that may limit the ability to develop accurate classifications. For example, the left side of Fig. 32.6

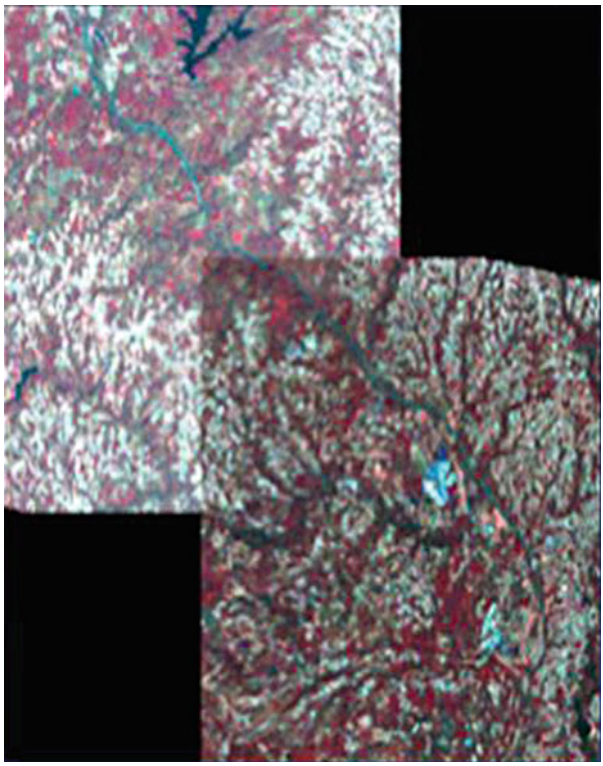


Fig. 32.8 Before normalization of radiometric differences (Graphic courtesy of Siamak Khorram)

shows imagery of two individual multispectral scanner (MSS) bands that were acquired from LANDSAT-4. Linear distortion features are noticeable across each image, while for general land use/land cover classification purposes these distortions may not have a large impact on developing overall general classification categories for a single image. However, to ensure the best possible use of training sites between multiple images, including images that may not have these types of distortions, it is generally good practice to remove or reduce these types of distortions.

In the example depicted in Fig. 32.11, the linear distortions occur in a uniformed orientation pattern, most likely due to the deteriorating age of the sensor. Uniformed distortions are generally less of a challenge to remove than distortions that occur randomly across an image. By dividing band 1 by band 2 to create a new ratioed image, the combined band data was able to reduce the original distortions apparent in the individual scenes. The newly created "ratioed image" may then be substituted in the band combination used for the classification procedure (Nelson et al. 2002). This type of radiometric correction, as well as any corrections that improve the interpretability of an image, is also known as a type of image enhancement procedure.

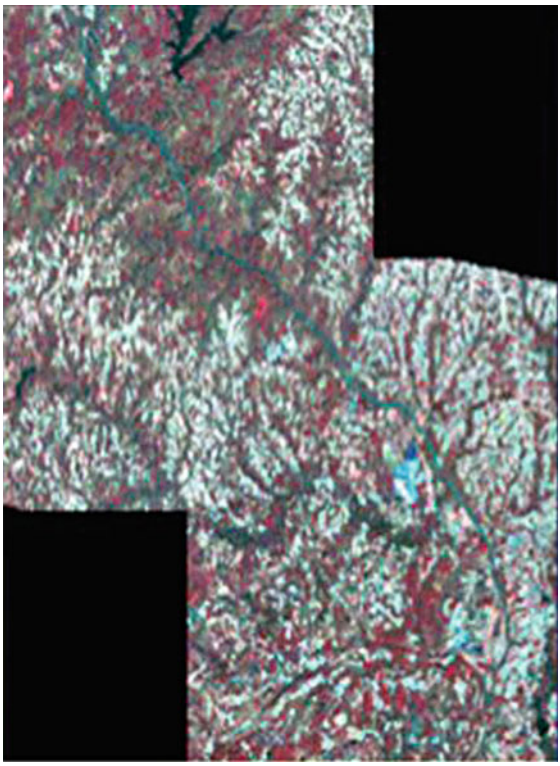


Fig. 32.9 After normalization of the radiometric differences (Graphic courtesy of Siamak Khorram)

Registration and Coordinate Systems

Image registration or image alignment is the process of transforming multiple sets of images, or other spatially referenced data that will be used together with the image data, into a single coordinate system of interest. A coordinate system (i.e., angular or geographical) is a system which uses a series of numbers, or coordinates, to uniquely determine locations on the Earth's surface. Coordinate systems also provide a reference system used to measure horizontal and vertical distances on a planimetric map (i.e., flat map). The coordinate system is usually defined by a map projection, a spheroid of reference, a datum, one or more standard parallels, a central meridian, and possible shifts in the *x*- and *y*-directions to locate X, Y positions. The map projections are used to transform a position on a curved Earth's surface (identified by latitude and longitude) into Cartesian Coordinates (X, Y).

Image data may be acquired at different times or from different data sources or sensors. All of these data may be registered in different coordinate systems that prevent corresponding locations on each image to overlay on top of each other. For this data useful in subsequent analyses involving multiple images or spatial datasets, the multiple images must be co-registered in a process that brings all data

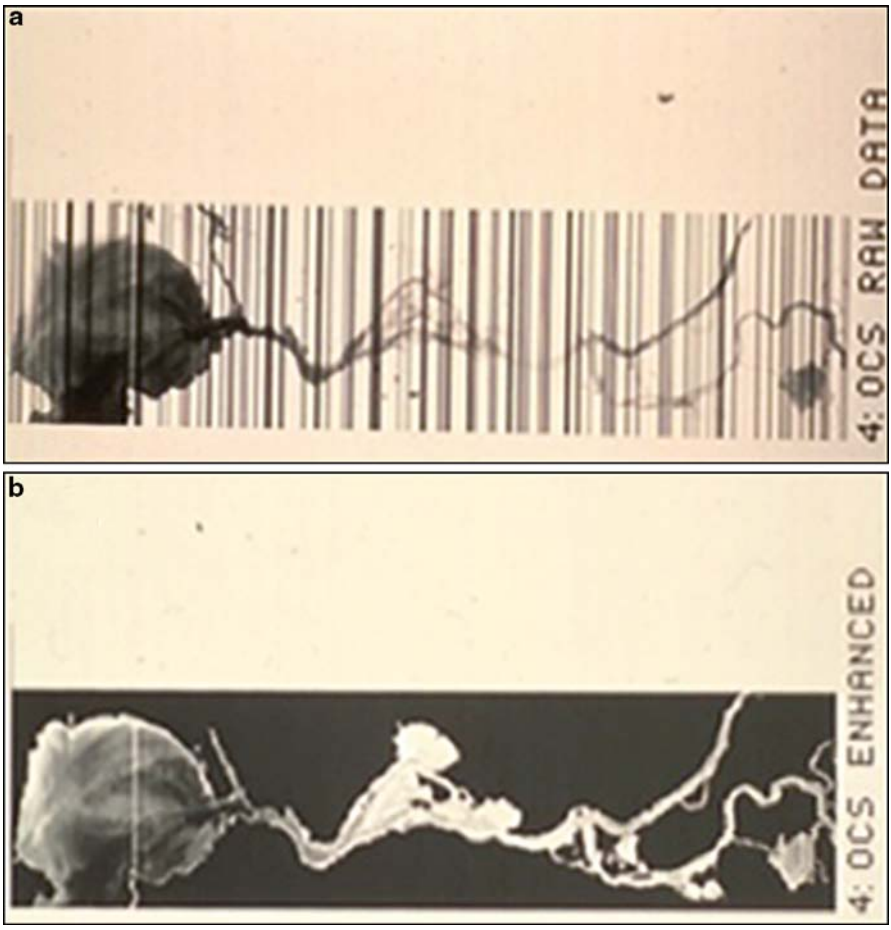


Fig. 32.10 (**a**) Remotely sensed image data containing a large number of dropped lines. (**b**) Remotely sensed image data where *dropped lines* (zero values) have been replaced with the average of the values of the *scan lines* above and below the *bad scan lines* (Graphic Courtesy of Siamak Khorram)

into a single coordinate system. For example, positional image registration errors (misregistration errors) have been shown to significantly limit change detection accuracies (Dai and Khorram 1999).

The majority of current change detection techniques depend on the accuracy of geometric registration of the two images being used as the resultant change analysis is generally performed on a pixel-by-pixel basis (Townshend et al. 1992). If accurate registration between images is not achieved, false differences will be detected due to the misalignment of features on the ground that would be evaluated as real changes at the same location from one time and another. To ensure maximum spatial fidelity between various types of image data it is often recommended to co-register all data, as well as geo-reference the data using

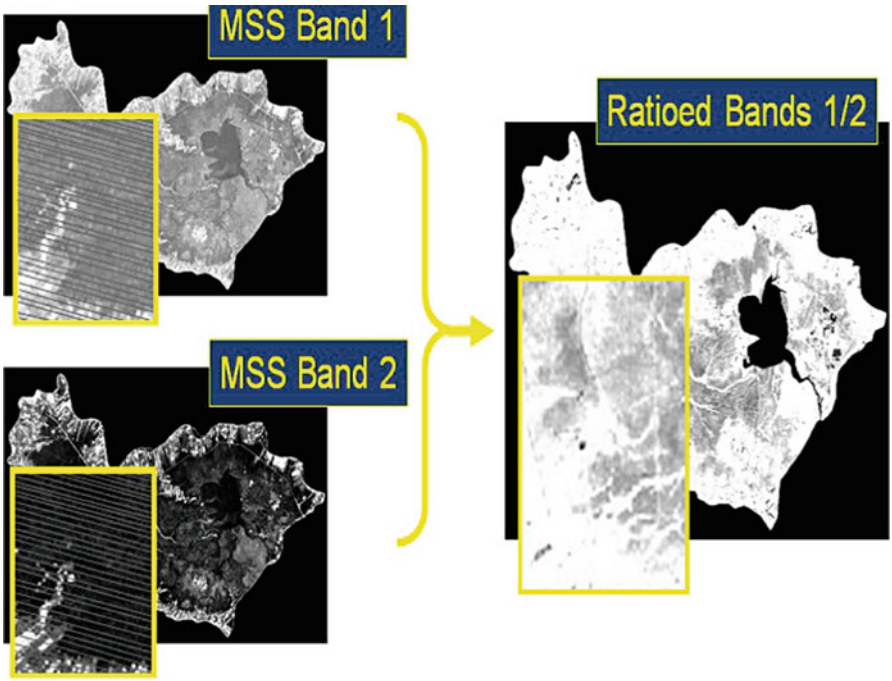


Fig. 32.11 Images to the left represent two individual raw Landsat MSS band scenes that contain linear distortions across the imagery, most likely due to the age of the sensor. The right image represents a new scene created from the "ratioed bands" to reduce visible distortions (Images Courtesy of NOAA)

ground control points (GCPs). A collection of a suitable number of GCPs (generally 50 points per class) will allow for an adequate feature matching and boundary-matching approach that will allow for a statistically valid computation of the root mean square error (RMSE). The RMSE is the absolute fit of the model to the data or the difference between values predicted by the transformation model and the actual values observed.

Data Reduction and Fusion Techniques

Data reduction and fusion are important components of image processing. Data reduction aims to reduce redundancy in digital images thus saving crucial storage space and processing time. Usually dubbed as image compression, data reduction techniques find applications in multimedia, communications, remote sensing, and digital image databases. One of the most widely known image compression method is the JPEG 2000 standard which is based on wavelet analysis techniques. Image compression can be lossless or loss-y meaning that some information might be lost depending upon the technique used.

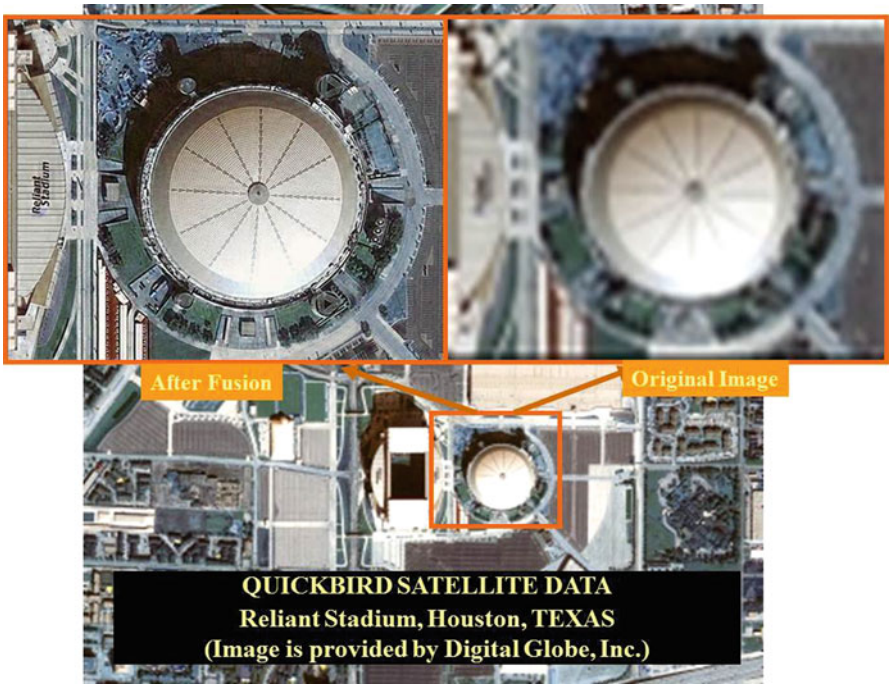


Fig. 32.12 An example of Data Fusion illustrating "before" (*on the right*) and "after" (*on the left*) (Graphic courtesy of Digital Globe, Inc.)

Data fusion techniques, on the other hand, try to assimilate information from multiple sources into a more useful form depending upon the objectives and the need of the user. If the need is, for example, to have high-resolution color images, colored (multispectral) but low spatial resolution satellite imagery can be fused with high-spatial-resolution panchromatic image to have sharp multispectral images. This fusion process is known as pan sharpening.

The most widely used tools for data reduction and fusion are described in this section, with Figs. 32.12 and 32.13 to illustrate their application.

Principal Components Analysis

Principal Component Analysis (PCA) is a multivariate statistical technique that originated from Pearson's early work (Pearson 1901). PCA is a mathematical way of determining the linear transformation of a sample of points in N-dimensional space that exhibits the properties of the sample most clearly along the coordinate axes (Gauch 1982).

Along the new axes, the sample variances are extremes (maxima and minima), that are uncorrelated. An analysis in terms of principal components can show (linear) interdependence in data (Legendre and Legendre 1998).

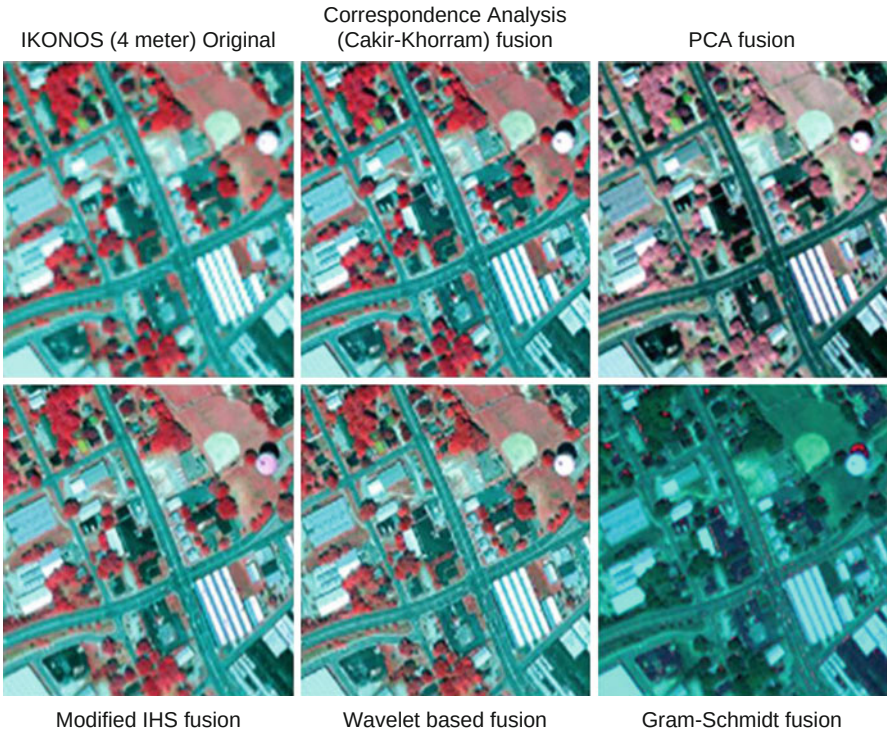


Fig. 32.13 An illustration of various data fusion techniques to a false color composite (FCC) of IKONOS satellite data (Graphic courtesy of IKONOS)

PCA is calculated either using variance-covariances or correlations. (If a correlation matrix is used, it is also known as "factor analysis" or "standardized principal components.") Given an image with n-number of bands, n-number of PCs can be calculated. The first PC explains most of the variability of the original data set. The second, third, fourth, and so on, components contain decreasing amounts of variance found in the data set. Thus, eliminating the components that have less information, associated with smallest eigenvalues, reduces the abundance of data to be analyzed. This is extremely helpful when classifying correlated sets of multi-spectral data.

The success of PCA depends on the data analyzed. By definition, PCA can be applied best if the data are Gaussian and grouped into a single cluster. Also, the success of PCA depends on the extent of multi-co-linearity within the data to be analyzed.

In remote sensing, PCA is mostly applied to correlated data sets to improve interpretability of the images (Short 2003). PCA is useful for image

encoding, image data compressing, image enhancement, digital change detection, multi-temporal dimensionality, and image fusion (Pohl and Genderen 1998).

Correspondence Analysis

A more recent multivariate method, correspondence analysis (CA) was developed independently by several authors. An algebraic derivation of CA is often accredited to Hirschfield (1935) who developed a formulation of the correlation between the rows and columns of a two-way contingency table (Beh 2004). For the development, history, and more information of the technique, readers can refer to the various citations in the end notes (Greenacre 1984; Benzécri 1992; Beh 2004). The term "correspondence analysis" is a translation of the French "analyse factorielle des correspondances."

Correspondence analysis can be applied to data tables other than contingency tables as long as the elements of a table to be analyzed are dimensionally homogenous (i.e., same physical units, so that they can be added) and nonnegative (so that they can be transformed into probabilities or proportions). The difference between PCA and the CA is that CA preserves the chi-square (χ^2) distance when computing the association between bands or variables (Carr and Matanawi 1999). In PCA, the distance among objects, in both the multidimensional space of original descriptors and the reduced space, are calculated using Euclidean distances. Most of the time, the last CA component is omitted from the analyzing procedure because the last eigenvalue is insignificant or small.

Although it is very well known to ecologists correspondence analysis is rarely explored in the remote sensing community. Carr and Matanawi (1999) introduced CA for principal component transformations of multispectral and hyperspectral imagery. Later, it was successfully applied to detection of change and data fusion (Cakir et al. 2006).

In image processing, CA, similar to PCA, can be used for image encoding, dimensionality reduction and image compression, data fusion, multi-temporal analysis and change detection, image classification, etc. While PCA normally highlights similarities in image contents, CA highlights differences.

Intensity, Hue, and Saturation Analysis

Intensity, hue, and saturation refer to the parameters of human color perception. Intensity refers to the total brightness of a color. Hue refers to the dominant or average wavelength of light contributing to a color. Saturation specifies the purity of a color relative to gray. Vivid colors are highly saturated, while pale pastel colors have low saturation. IHS transformation has been mostly used to fuse panchromatic data with three-band multispectral data in order to enhance spatial resolution of the multispectral data. First, a numerical procedure, which was developed to convert a three-band RGB (red-green-blue) display into its fundamental physiological (IHS) elements of human color perception, is performed to transform the multispectral input data into IHS space. Several algorithms have been developed for the

IHS transformation. Generally they are similar to each other with small differences in calculating the intensity component.

In data fusion (pan sharpening) of multispectral images, utilization of IHS transformation involves the substitution of one of the components with another data. Usually, it is the intensity component that is replaced by the panchromatic image. It is assumed that the intensity component is highly correlated to the panchromatic data. Panchromatic data can be contrast stretched to match the intensity component before replacement in order to reduce the spectral distortion on the final product. Finally, an inverse transformation is performed and the data are transformed back to the original RGB space. The IHS method is sufficient for SPOT imagery, but not for sensors like the Landsat TM sensor that has six or more reflection bands. The "correlation analysis of original multispectral image data and their counterparts in IHS composites indicates the need to consider carefully the potential influence alternative implementations of IHS procedures might have on the spectral characteristics of the resulting multi-resolution products."

Wavelet Method

Wavelets are an extension of Fourier analysis. Thus, to understand the wavelets and wavelet transform, it is important to look at Fourier analysis and its limitations. Fourier transform decomposes or separates a waveform or function into sinusoids of different frequency which sum to the original waveform. It recognizes or distinguishes the different frequency sinusoids and their respective amplitudes. It transforms a function f that depends on time (or on space) into a new function, $\hat{f}$, or "f hat," which depends on frequency. This new function is called the Fourier transform of the original function or, when the original function is periodic, it is called a Fourier series.

A function and its Fourier transform are two faces of the same information. The function displays time (or space) information and hides information about frequencies. Fourier transform displays information about frequencies, but information about time or space is hidden in the phases. Wavelets can provide efficient localization in both time and frequency or scale when the signal is represented as a function of time, that is, the major difference between Fourier transform and wavelet transform.

Grossmann and Morlet among others have showed that when wavelets are used to represent a signal, the "energy" of the signal is unchanged. This means that one can transform a signal into wavelet form and then get back exactly the same signal back again. It also means that a small change in the wavelet representation produces a correspondingly small change in the signal (Grossmann and Morlet 1984).

Some of wavelet applications in remote sensing include: speckle reduction, image compression, texture analysis and classification, automatic geometric registration of images, and merging multi-resolution satellite images (See the following articles: Horgan 1998; Singh 1999; Singh and Harrison 1985; Zhu and Yang 1998; Djamdji et al. 1993; Zhou et al. 1998; Yocky 1996; Núñez et al. 1999; Cakir et al. 1999; Ranchin and Wald 2000). Utilization of wavelets for data fusion

in remote sensing is best explained with the ARSIS method (Ranchin and Wald (1993); Also see Ranchin and Wald (2000)). ARSIS, the French acronym for "Amélioration de la Résolution Spatiale par Injection de Structures," enables the fusion of images having different spatial and spectral resolutions. That is, it allows the ability to insert details from high-spatial resolution data (e.g., SPOT panchromatic image) to high-spectral resolution data (e.g., SPOT multispectral). With this method, details extracted from a panchromatic image by multi-resolution wavelet decomposition are used to estimate details that are missing in successive resolutions (the spatial resolution difference between panchromatic image and multispectral image) in multispectral imagery.

Those processing methods that represent the closest colors to the original image color on the upper left side (IKONOS MS 4-m) indicate the best performance. In this case, the Cakir–Khorram (1-m) method has produced the best results. The data fusion techniques are applied to produce a 1-m multispectral data by fusing 4-m multispectral data with 1-m panchromatic data.

International Policies Governing Remotely Sensed Data

Each day, millions of individual Earth observations are collected, allowing the ability to examine, monitor, and model climate change, ecosystem health, atmospheric composition, seismic activity, terrestrial and marine resources, agricultural monitoring, weather patterns, humanitarian violations, and hundreds of other phenomena. There are now more than 27 countries with satellite-based Earth observation resources, compared with three in 1980, as well as an increasing number of countries which have their own image-receiving stations for remote sensing systems, due to the reduction in acquisition costs (de Montluc 2009).

Historically speaking, remote sensing was originally developed for military applications, with nearly 75% of satellites employed for military reconnaissance and surveillance purposes (Keeley and Huebert 2004; Jakhu 2004). The launch of the civilian American satellite, LANDSAT-1, in 1972, however, facilitated widespread Earth observation by researchers.

In 1959, shortly after launching the first satellite, the United Nations General Assembly established the Committee on the Peaceful Uses of Outer Space (COPUOS) as the only international forum for the development of international space law. Since its inception, the Committee has concluded five international legal instruments and five sets of legal principles governing space-related activities. International law and policies generally focus on three primary areas:

1. The right to acquire remotely sensed imagery/the right to launch remote sensing satellites
2. The right to disseminate remotely sensed imagery without prior consent of the sensed state
3. The right to obtain remotely sensed satellite imagery from a particular state

As Earth observation was in its infancy, the prevailing attitude was one of global scientific cooperation, particularly with regard to meteorological data sharing to warn of extreme weather events. Further, the earliest systems operated at large-scale (spatial, temporal, and spectral) resolution. As the worldwide availability of high-quality unclassified satellite imagery has increased, however, there has been increasing concern that this could be used for unsanctioned military or terrorist purposes. Consequently, international and national policies on the acquisition and distribution of remotely sensed satellite imagery have changed over time.

After extensive discussions in the COPUOS, the United Nations General Assembly unanimously adopted the Resolution, Principles Relating to Remote Sensing of the Earth from Outer Space, in 1986. Although the Principles are not a binding source of international law per se, this landmark resolution is regarded by many space law researchers as the primary international legal document that addresses issues of remote sensing (Harris 2003; Jakhu 2004; Macauley 2005; Rao and Murthi 2006).

As such, the Principles are regarded a codification of customary law and have acquired the "evidence of a general practice accepted as law" according to Article 38(1)(b), Statute of International Court of Justice (Gabrynowicz (1993); Also see Williams (2006)). The 15 Principles place international customary obligations on states and form the basis for remote sensing activities globally, regulating cooperation between sensing and sensed states. In particular, the Principles require that sensed states should have access to data without any discriminatory practices and at "reasonable cost." Further, states are urged to consult with sensed states for mutual cooperation and benefit in the use of data, and states that are engaged in remote sensing should make technical assistance accessible to other interested states on "mutually agreed terms."

In addition to the UN Principles, several other factors have contributed to a presumption of open access. Nondiscriminatory access policies have been adopted by major remote sensing nations (e.g., Japan, the United States, and Canada) and the data policies of some remote sensing missions (e.g., ENVISAT, RADARSAT) specifically incorporate nondiscriminatory access (Harris 2003). At present there are at least 65 statutes worldwide that govern access to information, of which at least 50 establish a right of access to information, rather than a mere limited right of access to documents (United Nations Resolution 41/65 1986). The right of access to environmental information is, in certain circumstances, guaranteed by the European Convention on Human Rights.

In response to a call for more international cooperation and coordination in data sharing regarding atmospheric, land, and water data, the Group on Earth Observations (GEO), a voluntary partnership of governmental and international organizations, was formally created by resolution at the third Earth Observation Summit (EOS) held in Brussels in February 2005. In creating GEO on a voluntary and legally nonbinding basis, the founding governments and international organizations represented resolved that GEO would implement the Global Earth Observation System of Systems (GEOSS) in accord with a 10-year

(2005–2015) Implementation Plan. The Plan envisions sharing of Earth observation data by GEO members and participating organizations through the Global Earth Observing System of Systems (GEOSS). A major initiative of the plan is the establishment of the GEOSS Data Collection of Open Resources for Everyone (Data-CORE). Many challenges remain to be resolved, including who pays for infrastructure, training, and administration; whether to control data access; how to include the private sector; and whether problems of collective action will continue to beleaguer the effort.

Conclusion

As high-resolution (i.e., submeter resolution) imagery continues to be widely available to the public through a multitude of sources, the divide between open (civilian) access and restricted (military) access has vanished. We have moved from an era in which a handful of developed countries had access to high-resolution imagery to one in which virtually everyone will have such access. Such transparency offers both enormous benefits and challenges. As yet, governments throughout the world are unprepared for this new era of global transparency. A harmonized international framework of international legal norms that goes beyond the scope of the UN Remote Sensing Principles will be needed to resolve these challenges.

Notes

1. Hyperspectral Sensing and Imaging, http://modis.gsfc.nasa.gov/about
2. Airbourne Visible Infrared Imaging Spectrometer, http://aviris.jpl.nasa.gov/
3. Recent AVIRIS Imaging http://aviris.jpl.nasa.gov/html/aviris.cube.html

References

T.E. Avery, G.L.L. Berlin, *Fundamentals of Remote Sensing and Air Photo Interpretation*, 5th edn. (Prentice Hall, New York, 1992), 472 p

E.J. Beh, Simple correspondence analysis: a bibliographic review. Int. Stat. Rev. **72**(2), 257–284 (2004)

J.P. Benzécri, *Correspondence Analysis Handbook* (Marcel Dekker, New York, 1992)

H.I. Cakir, S. Khorram, S.A. Nelson, Correspondence analysis for detecting land cover change. Remote Sens. Environ. **102**, 306–317 (2006)

H.I. Cakir, S. Khorram, X.L. Dai, P. de Fraipont, Merging SPOT XS and SAR imagery using the wavelet transform method to improve classification accuracy, in *Geoscience and Remote Sensing Symposium, IGARSS'99 Proceedings, IEEE International*, vol 1 (1999), pp. 71–73

J.R. Carr, K. Matanawi, Correspondence analysis for principal components transformation of multispectral and hyperspectral digital images. Photogramm. Eng. Remote Sens. **65**(8), 909–914 (1999)

X. Dai, S. Khorram, Quantification of the impact of misregistration on digital change detection accuracy, in *Proceedings of IEEE/IGARSS'97 International Geoscience and Remote Sensing Symposium,* Singapore, 1997

X. Dai, S. Khorram, Data fusion using artificial neural networks: a case study on multi-temporal change analysis. Comput. Environ. Urban Syst. **23**, 19–31 (1999)
B. de Montluc, The new international political and strategic context for space policies. Space Policy **25**, 20–28 (2009)
J.P. Djamdji, A. Bijaoui, R. Manieri, Geometrical registration of images: the multiresolution approach. Photogramm. Eng. Remote Sens. **59**, 645–653 (1993)
J.R. Dobson, E.A. Bright, R.L. Ferguson, D.W. Field, L.L. Wood, K.D. Haddad, H. Iredale, J.R. Jensen, V. Klemas, R.J. Orth, J.P. Thomas, *NOAA Coastal Change Analysis Program (C-CAP): Guidance for Regional Implementation* (National Oceanic &Atmospheric Administration, NMFS, Washington, DC, 1995), 92 p
B.C. Forster, Derivation of atmospheric correction procedures for Landsat MSS with particular reference to urban data. Int. J. Remote Sens. **5**, 799–817 (1984)
J.I. Gabrynowicz, The promise and problems of the Land Remote Sensing Policy Act of 1992. Space Policy **9**, 319–328 (1993).
H.G. Gauch, Jr., *Multivariate analysis in community structure* (Cambridge University Press, Cambridge, 1982)
M.J. Greenacre, *Theory and Application of Correspondence Analysis* (Academic, London, 1984)
A. Grossmann, J. Morlet, Decomposition of Hardy functions into square integrable wavelets of constant shape. SIAM J. Math. Anal. **15**, 723–736 (1984)
R. Harris, Current policy issues in remote sensing: report by the International Policy Advisory Committee of ISPRS. Space Policy **19**, 293–296 (2003)
M. Herold, S. Guenther, K.C. Clarke, Mapping urban areas in the Santa Barbara south coast using IKONOS and eCognition. eCognition Appl. Note. Munchen: Definiens ImgbH **4**(1), 20 p (2003)
G. Horgan, Wavelets for SAR image smoothing. Photogramm. Eng. Remote Sens. **64**(12), 1171–1177 (1998)
B. Jähne, *Digital Image Processing* (Springer, New York, 1991), pp. 219–230
R. Jakhu, International law governing the acquisition and dissemination of satellite imagery, in *Commercial Satellite Imagery and United Nations Peacekeeping: A View from Above* (Ashgate, Burlington, 2004), 259 p
J.R. Jensen, *Introductory Digital Image Processing*, 3rd edn. (Pearson Prentice Hall, Upper Saddle River, 2005), 316 p
J.F. Keeley, R.N. Huebert, *Commercial Satellite Imagery and United Nations Peacekeeping: A View from Above* (Ashgate, Burlington, 2004)
S. Khorram, Coastwatch – water quality mapping of the entire San Francisco Bay and delta from Landsat multispectral scanner data, in *Space Sciences Laboratory*, ed. by : R.N. Colwell, PI, Series 23, Issue 6. (University of California, Berkeley, 1982), 34 p
P. Legendre, L. Legendre, *Numerical Ecology: Second English Edition* (Elsevier Science, Amsterdam, 1998), 823 p
T. Lillesand, R. Kiefer, J. Chipman, *Remote Sensing and Image Interpretation*, 6th edn. (Wiley, New York, 2008), 763 p
R.S. Lunetta, C. Elvidge, *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications* (Taylor & Francis, New York, 2000), 340 p
M.K. Macauley, Is the vision of the earth observation summit realizable? Space Policy **21**, 29–39 (2005)
S.A. Nelson, P.A. Soranno, J. Qi, Land cover change in the upper Barataria Basin estuary, Louisiana, from 1972–1992: increases in wetland area. Environ. Manag. **29**(5), 716–727 (2002)
J. Núñez, X. Otazu, O. Fors, A. Prades, V. Palà, R. Arbiol, Multiresolution-based image fusion with additive wavelet decomposition. IEEE Trans. Geosci. Remote Sens. **37**(3), 1204–1211 (1999)
K. Pearson, On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* **2**, 559–579 (1901)

C. Pohl, V. Genderen, Multisensor image fusion in remote sensing: concepts, methods and applications. Int. J. Remote Sens. **19**(5), 823–854 (1998)
T. Ranchin, L. Wald, The wavelet transform for the analysis of remotely sensed images. Int. J. Remote Sens. **14**, 615–619 (1993)
T. Ranchin, L. Wald, Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation. Photogramm. Eng. Remote Sens. **66**(1), 49–61 (2000)
M. Rao, K.R.S. Murthi, Keeping up with remote sensing and GI advances – policy and legal perspectives. Space Policy **22**, 262–273 (2006)
J.C. Russ, *The Image Processing Handbook* (Boca Raton, CRC Press, 2002), 744 p
A. Singh, A. Harrison, Standardized principal components. Int. J. Remote Sens. **6**, 883–896 (1985)
V.K. Singh, Discrete wavelet transform based image compression. Int. J. Remote Sens. **20**, 3399–3405 (1999)
N.M. Short, The remote sensing tutorial. Published by NASA via Internet, (2003), http://rst.gsfc.nasa.gov/.
P.H. Swain, S.M. Davis, *Remotde Sensing: The Quantitative Approach* (McGraw-Hill, New York, 1978), pp. 166–174
J.R.G. Townshend, C.O. Justice, C. Gurney, J. McManus, The impact of misregistration on change detection. IEEE Trans. Geosci. Remote Sens. **30**, 1054–1060 (1992)
R.E. Turner, M.M. Spencer, Atmospheric model for correction of spacecraft Data, in *Proceedings of the Eighth International Symposium on Remote Sensing of the Environment*, Ann Arbor, vol. 11 (1972), pp. 895–893
United Nations Resolution 41/65. Principles relating to remote sensing of the Earth from outer space. Adopted without a vote (1986), http://www.oosa.unvienna.org/oosa/en/SpaceLaw/gares/html/gares_41_0065.html
M. Williams, Legal aspects of the privatization and commercialization of space activities, remote sensing, and national space legislation. Second report. (International Law Association, Toronto, 2006), p. 2
D.A. Yocky, Multiresolution wavelet decomposition image merger of Landsat thematic mapper and SPOT panchromatic data. Photogramm. Eng. Remote Sens. **62**(9), 1067–1074 (1996)
H. Yuan, C.F. Van Der Wiele, S. Khorram, An automated artificial neural network system for land use/land cover classification from Landsat TM imagery. J. Remote Sens., Open Source, Switzerland, ISS 2072-4292, 20 p (2009)
J. Zhou, D.L. Civco, J.A. Silander, A wavelet transform method to merge Landsat TM and SPOT panchromatic data. Int. J. Remote Sens. **19**(4), 743–757 (1998)
C. Zhu, X. Yang, Study of remote sensing image texture analysis and classification using wavelet. Int. J. Remote Sens. **19**(16), 3197–3203 (1998)

Digital Image Processing: Post-processing and Data Integration

33

Siamak Khorram, Stacy Nelson, Halil Cakir*, and Cynthia Van Der Wiele

Contents

Introduction 840
Image Processing 840
Image Post-Processing and Smoothing 853
Filtering 855
Accuracy Assessment 855
Change Detection 857
Data Integration and Spatial Modeling 858
Notes 862
References 862

*Dr. Halil Cakir did not contribute to this article as an employer of the US Environmental Protection Agency nor does this article reflect the views of this agency.

S. Khorram (✉)
Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, 94720, USA
and
Center for Earth Observation, North Carolina State University, 5123 Jordan Hall, Raleigh, NC, 27695-7106, USA
e-mail: khorram@Berkeley.edu

S. Nelson
Center for Earth Observation, North Carolina State University, 5123 Jordan Hall, Raleigh, NC, 27695-7106, USA
e-mail: Stacy_Nelson@ncsu.edu

H. Cakir
US EPA, Research Triangle Park, 109 T. W. Alexander Drive, NC, 27709, USA
e-mail: Cakir.Halil@epa.gov

C. Van Der Wiele
Cynthia Van Der Wiele and Associates, LLC, 715 Shepherd Street, Durham, NC, 27701, USA
e-mail: cynthia.vanderwiele@alumni.duke.edu

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_92, © Springer Science+Business Media New York 2013

Abstract

The main objective of this chapter is to focus on the digital image processing, post-processing, and data integration techniques as applied to remotely sensed data for the purpose of extracting useful earth resources information. Image preprocessing and data reduction tools are described in the previous chapter. The concepts discussed in this chapter include:

- Image processing techniques such as unsupervised image classifications, supervised image classifications, neural network classifiers, simulated annealing classifiers, and fuzzy logic classification systems
- The most widely accepted indices and land use/land cover classification schemes
- Post-processing techniques such as filtering and change detection
- Accuracy assessment and validation of results
- Data integration and spatial modeling including examples of integration of remotely sensed data with other conventional survey and map form data for Earth observation purposes

Keywords

Accuracy assessment • Classification schemes • Data and image integration • Digital image processing • Geospatial modeling • Image classification • Image validation • Image visualization • Post-processing • Satellite remote sensing • Spatial modeling

Introduction

In this chapter we will describe various image processing and post-processing techniques and data integration with other data types in a Geographical Information System (GIS) for spatial modeling purposes. Image processing, data reduction, and preparation of image form data for processing has been described in the previous chapter.

Image Processing

Image processing is used for the improvement of image visual display and analysis, presentation of data in an orderly and meaningfully classified form, and integration of image form data with other conventional surveys and maps. These techniques involve visual image data displays, vegetation indices, commonly used image classification systems, and widely accepted land use and land cover scheme. Data integration with other data types and spatial modeling are introduced in this section.

Data Display, Visualization, and Reduction Methods

Raw data displays are most frequently provided as true color composite (TCC) and the false color composite (FCC). True color composite displays the blue band in blue



Fig. 33.1 Two identical images illustrating false color composite (FCC) and true color composite (TCC) below (Graphic courtesy of QuickBird)

color, the green band in green color, and the red band in red color while false color composite displays any three bands in blue, green, and red colors. The standard FCC provides an infrared band displayed in red color, the red band displayed in green color, and the green band displayed in blue color. The differences between TCC and standard FCC data displays are illustrated in Fig. 33.1. In the standard FCC all vegetation appears in red, whereas in TCC vegetation appears in green (Fig. 33.2) (Hester 2008).

Band Combinations, Ratios, and Indices

Arithmetic combinations of certain bands via division, addition, subtraction, or multiplication can lead to enhanced information. Differencing and ratioing are primarily used for change detection and spectral enhancement studies. Band ratios typically include the following: infrared band over red band for vegetation distribution; green band over red band for mapping surface water bodies and wetland delineation; red band over infrared band for mapping turbid waters; red band over blue band or red band over green band for mineral mapping, such as iron-rich or iron-poor rocks.

The most frequently used index for vegetation mapping is the normalized difference vegetation index (NDVI), which is defined as:

$$NDVI = \frac{B2 - B1}{B2 + B1}$$



Fig. 33.2 True color composite (TCC) as opposed to false color composite (FCC) as presented above (Graphic courtesy of QuickBird)

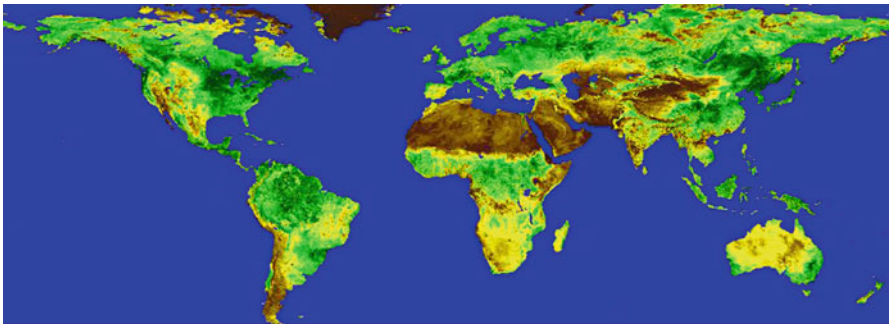


Fig. 33.3 Example of NDVI image depicting the global vegetation cover in various shades of *green* color (Graphic courtesy of NASA)

where B2 are the brightness values, or digital numbers, in an infrared band and B1 is brightness values, or digital numbers, in the red band. Figure 33.3 illustrates the extent of vegetation in various shades of green color based on NDVI.

Image Classification Systems

Categorizing the pixels of a digital image into particular land cover classes or themes is referred to as *image classification*. Most image processing techniques are

based on hard logic, utilizing both spectral and temporal spatial patterns. Several other techniques have been recently explored for processing remotely sensed data utilizing including fuzzy logic (Cho and Kim 1995) neural network and decision trees (Dai and Khorram 1997), simulated annealing, and hybrid methodologies. In this section, the focus will be on hard logic first, and then briefly describe other classifiers. Image classification based on hard logic is divided into *supervised classification* and *unsupervised classification* as defined by Hester and others.[1]

Land cover, commonly referred to as land use/land cover classification scheme, must be utilized for identifying categories of interest. Three common classification schemes are based on the objectives of analysts. These are: the US Geological Survey (USGS) classification system; the Cowardin, Carter, Golet, and LaRoe system (Cowardin et al. 1979); and the NOAA Coastal Change Analysis Program (C-CAP) system (NOAA 2004) scheme (Anderson et al. 1976; USGS 2004) is the predominant system used. It is designed to utilize remotely sensed data, and is organized in four levels of progressively detailed information. Level I and Level II of these categories are most frequently used, and shown in Table 33.1.

The other two most frequently used classification systems are primarily used for coastal and wetland analyses. These are the US Fish and Wildlife Classification System[2] and the National Oceanic and Atmospheric Administration (NOAA) Coastal Change Analysis Program (Klemas et al. 1993. Also see Khorram et al. 1996).

Usually there are overlaps and redundancies as information is contained in more than one band of multiband spectral data. The process of reducing this redundancy should be observed carefully so that one does not create an adverse impact on the accuracy of classified data through statistical and/or graphical analyses such as autocorrelation, bar graph spectral plots, feature spectral plots, and so forth.

Unsupervised Classification

Unsupervised classification methods generally use no or minimal analyst supervision to develop the resultant land use/land cover. This is a computerized process whereby each pixel is iteratively assigned to a class based on the similarity of the spectral properties of pixels in multiple bands. This is accomplished through the determination of class "spectral separability," based on means and covariance matrices. The analyst assigns the categorical information to the classified data after classification. Two most common techniques for unsupervised classification are clustering algorithms and the iterative self-organizing data analysis technique (ISODATA):

Using the *Clustering Algorithms* approach, clustering is performed in two stages. During the first stage, a number of clusters are built up sequentially. Each cluster is composed of pixels having similar spectral values (within a range that is determined by the analyst), therefore occupying a common spectral space (Jain 2009; Celik 2009) with a well-defined mean vector for each class. In the second stage, a minimum distance to means classification algorithm, described earlier, is applied to the entire data set for assigning each pixel to a cluster.

Table 33.1 Land use and land cover classification system for use with remote sensor data[a]

Level I	Level II
1. Urban or built-up land	11 Residential
	12 Commercial and services
	13 Industrial
	14 Transportation, communications, and utilities
	15 Industrial and commercial complexes
	16 Mixed urban or built-up land
	17 Other urban or built-up land
2. Agricultural land	21 Cropland and pasture
	22 Orchards, groves, vineyards, nurseries, and ornamental horticultural areas
	23 Confined feeding operations
	24 Other agricultural land
3. Rangeland	31 Herbaceous rangeland
	32 Shrub and brush rangeland
	33 Mixed rangeland
4. Forest land	41 Deciduous forest land
	42 Evergreen forest land
	43 Mixed forest land
5. Water	51 Streams and canals
	52 Lakes
	53 Reservoirs
	54 Bays and estuaries
6. Wetland	61 Forested wetland
	62 Non-forested wetland
7. Barren land	71 Dry salt flats
	72 Beaches
	73 Sandy areas other than beaches
	74 Bare exposed rock
	75 Strip mines, quarries, and gravel pits
	76 Transitional areas
	77 Mixed barren land
8. Tundra	81 Shrub and brush tundra
	82 Herbaceous tundra
	83 Bare ground tundra
	84 Wet tundra
	85 Mixed tundra
9. Perennial snow or ice	91 Perennial snowfields
	92 Glaciers

[a]Table modified from Anderson et al. 1976

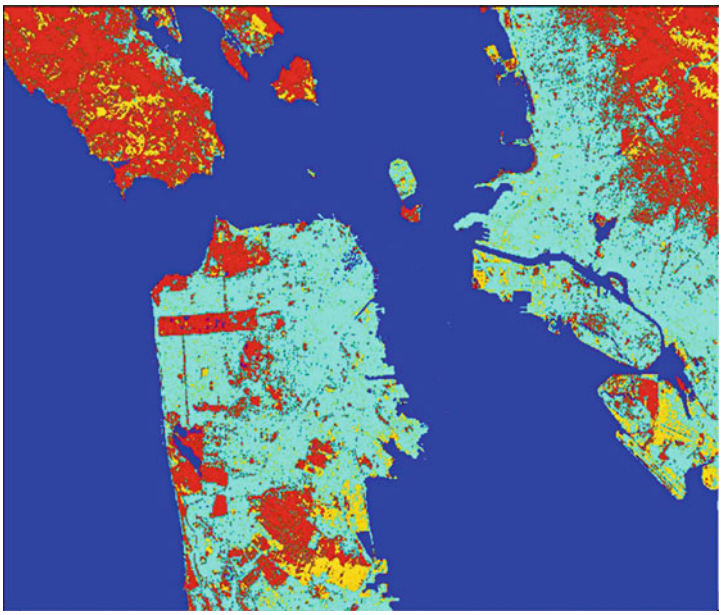


Fig. 33.4 Example of unsupervised classification depicting the City of San Francisco and surrounding area, California, USA (Graphic courtesy of Siamak Khorram)

The *Iterative Self-Organizing Data Analysis Technique (ISODATA)* (Tou and Gonzalez 1977. Also see Sabins 1987) is a very widely used multiple iterative procedure with minimal analyst supervision for processing remotely sensed digital data. ISODATA utilizes the mean and standard deviation in a number of bands in n-dimensional space. The parameters determined by the analyst initially to guide the algorithm include maximum number of clusters; maximum percentage of pixels that can remain unchanged between iterations; maximum number of iterations; minimum percentage of pixels assigned to each cluster; maximum standard deviation; and the minimum distance between clusters (Jensen 2005. Also see Hester 2008). An example of an application of unsupervised classification for the City of San Francisco and surrounding areas is shown in Fig. 33.4.

Supervised Classification Systems

Supervised classifications are a three-stage process performed by the analyst comprised of training, classification, and output.

During the *training stage*, training sites are selected by the analyst to represent areas with known cover types. In this stage, the analyst establishes the relationships between the cover types and the spectral data in multiple wavelength bands. Other data types such as aerial photography, ancillary spatial data, or field visits are typically used for establishing these relationships. The number of training sites

should be at least three times the number of categories of interest. Training sites are usually selected to represent the spectral distribution of each cover type of interest and are randomly or systematically distributed throughout the study area. The training sites should each represent a completely homogenous window, which is a small portion of the full scene covering a given cover type.

The second stage of supervised classification is the *classification stage*: In order to differentiate spectral bands into accurate land use/land cover categories, a number of classification and pattern recognition algorithms have been developed for supervised classification. The most widely used classification algorithms include minimum-distance-to-means, parallelepiped, and maximum likelihood, and are briefly described.

The *Minimum-Distance-to-Means Classification Algorithm* determines the mean spectral values for each category in each band, and then the spectral domains for multiple bands are computed. Each land cover category has a well-identified spectral space. A pixel of unknown identity (unassigned pixel) can be assigned to the category which has the minimum distance between the unknown pixel value and the mean vector in the category of interest. Although this method is simple and computationally efficient, it is insensitive to various degrees of variance in spectral response data (Lillesand 2008), which eventually results in decreased classification accuracy.

Spectral variance is taken into account in the *Parallelepiped Classification Algorithm* by establishing a range for each category in multiple bands to establish a decision region. The ranges are defined as the lowest and highest brightness values arranged in a rectangular or stepped rectangular scattergrams called parallelepipeds. This method is also simple, computationally efficient, and allowed for variance. But covariance, i.e., the tendency of spectral values to vary similarly in two or more bands, can be poorly taken into account at best. In this case, the interdependency of bands to each other is removed to an extent (Lillesand 2008).

The *Maximum Likelihood Classification Algorithm* assumes the normal, Gaussian, distribution of training data statistics for each class in each band (Blaisdell 1993). Given this assumption, the distribution of spectral responses in each category can be described in mean vector and covariance matrix. The statistical probability of a given pixel belonging to a category of interest can then be determined. For a three-band data, the probability density functions. The unknown (unassigned) pixel can then be assigned the category of the highest probability value, which is maximum likelihood of belonging. The maximum likelihood classifier is based on equiprobability contours. The shape of these contour lines is due to their sensitivity to covariance.[3] The maximum likelihood classification algorithm is the most commonly used method, but it is not as computationally efficient as the other methods described above. Figure 33.5 illustrates an application of the maximum likelihood classifier over Raleigh, North Carolina, USA as compared to the TCC of the same area based on the QuickBird satellite data.

Finally, a more complex version of the maximum likelihood classifier is the Bayesian Classifier, in which a certain weight is assigned to probability estimates

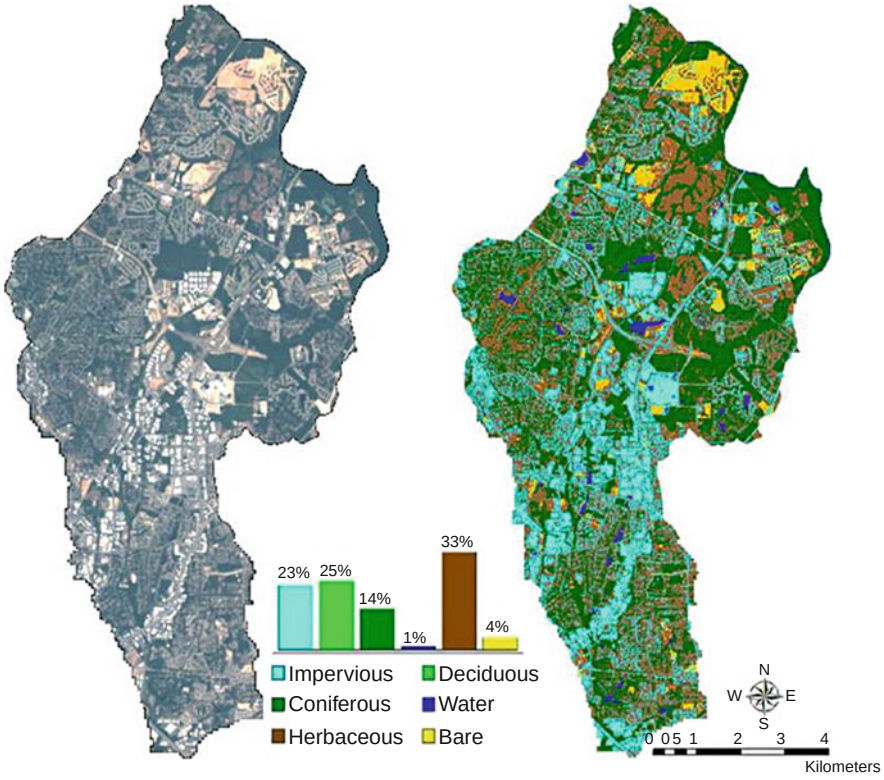


Fig. 33.5 Quickbird satellite imagery (*left*) and the associated land use/land cover classification map (*right*) based on maximum likelihood classifier (Graphic courtesy of Siamak Khorram)

based on a prior knowledge of the anticipated likelihood of occurrence (Hester et al. 2008; Also see Hord 1982).

The final stage is known as the *output stage*. Output products are used for presentation and visualization of the results. This is done by regrouping the results of classified data into a desired group of classes as determined by the analyst and presented in digital form, hard copy output product form (film, transparencies, paper, slides, etc.), or in graphic and tabular forms. The output product is accompanied by the statistical parameters, accuracy assessment table, and other supporting information.

Figures 33.6 and 33.7 are examples of a classified image based on the Supervised Maximum Likelihood classification system showing progressively more detailed information as compared to the TCC of the same areas.

Neural Network Classifiers

Artificial Neural Network (ANN) approaches have been widely used for image classification in remote sensing since the 1990s (Qiu 2004; Dai and Khorram 1999; Yang and Chung 2002). The automated system consists of two ANN classification

modules: (1) a Kohonen's Self-Organizing Mapping (KSOM) module based on *unsupervised* KSOM neural networks and (2) a Multilayer Perceptron (MLP) module based on *supervised* MLP neural networks that uses single or multilayer perceptrons to approximate the inherent input–output relationships is the most commonly used network model for image classification in remote sensing (Kanellopoulos 1997). MLP networks are typically trained with the supervised back propagation (BP) algorithm (Rumelhart et al. 1986) and consist of one input layer, one or more hidden layers, and one output layer. In the ANN system, each module is composed of several sub-modules: pattern conversion, network training, and the network generalization sub-module. A working flowchart of the ANN-based classification system and functions of each sub-module are summarized in Fig. 33.7 (Nogami et al. 1997. Also see Xu and Wunsch 2005).

The pattern conversion sub-module performs the following functions: (1) sampling a certain number of training and testing patterns from a number of selected image subsets; (2) scaling the input pattern into the network operational interval; and (3) generating training or testing pattern files. In the pattern conversion sub-module of the supervised MLP network, the corresponding class label must be provided for each pattern. Network training sub-modules provide the graphical user interfaces to allow the user to interactively define the architecture and parameters needed and to perform the training once all the parameters are set. In the KSOM module, two training sub-modules are provided: the standard KSOM and the KSOM-SA training sub-module. In the MLP module, the BP training sub-module is used. During a network training trial using each of the training sub-modules, an error file is generated to record the training MSEs to assist in monitoring the training behavior and selecting the appropriate network and parameters. After training is completed, network generalization sub-modules are implemented to generalize the entire image using the trained network and to produce the classified map. The application of this methodology for land use/land cover classification is illustrated in Fig. 33.9a–c. using an example in Coastal North Carolina, USA.

Simulated Annealing Classifiers

The commonly used *K*-means type algorithms are known to produce good results only if the clusters are well-separated in the feature space, and are hyper-spherical in shape when *Euclidean* distance is used. However, in a complex problem that cannot be solved by a simple convex cost function, the local minima problem is inevitable.

Alternatively, Simulated Annealing (SA) was developed on the basis of an analogy between the physical annealing process of solids and the large combinatorial optimization problems (Das and Chakrabarti 2005. Also see De Vincente et al. 2003). SA was proven to have great potential to find or approximate the global or near-global optimal in a combinatorial optimization problem. The premise of SA is to incorporate some randomness in the assignments of cluster labels to pixels in the clustering procedure, thus reducing the limitation of local minima. As a result, using the SA-based approach to classification has the potential to improve the



Fig. 33.6 A comparison of TCC and a classified image of the same area based on maximum likelihood classifier (Graphic courtesy of Siamak Khorram)

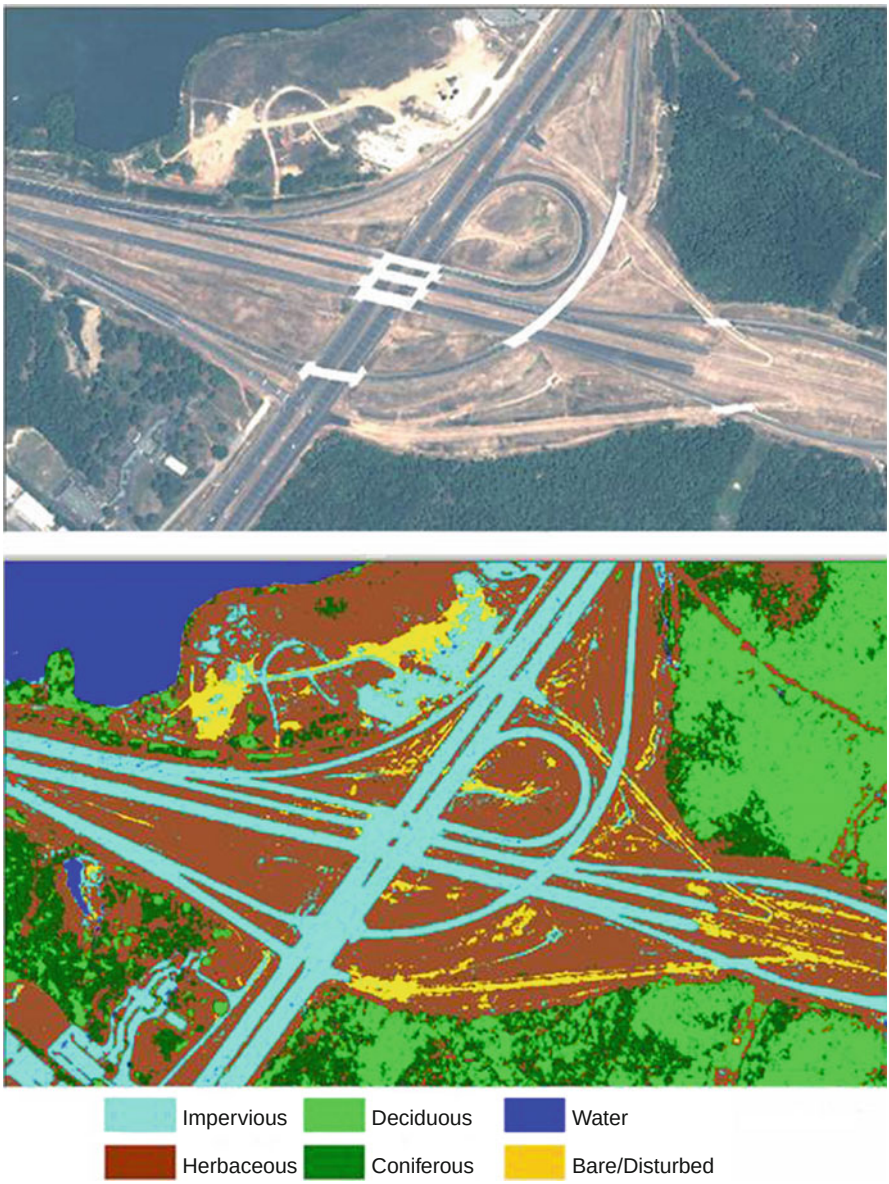


Fig. 33.7 An illustration of a more detailed classification based on maximum likelihood classifier as compared to a TCC of the same area (Graphic courtesy of Siamak Khorram)

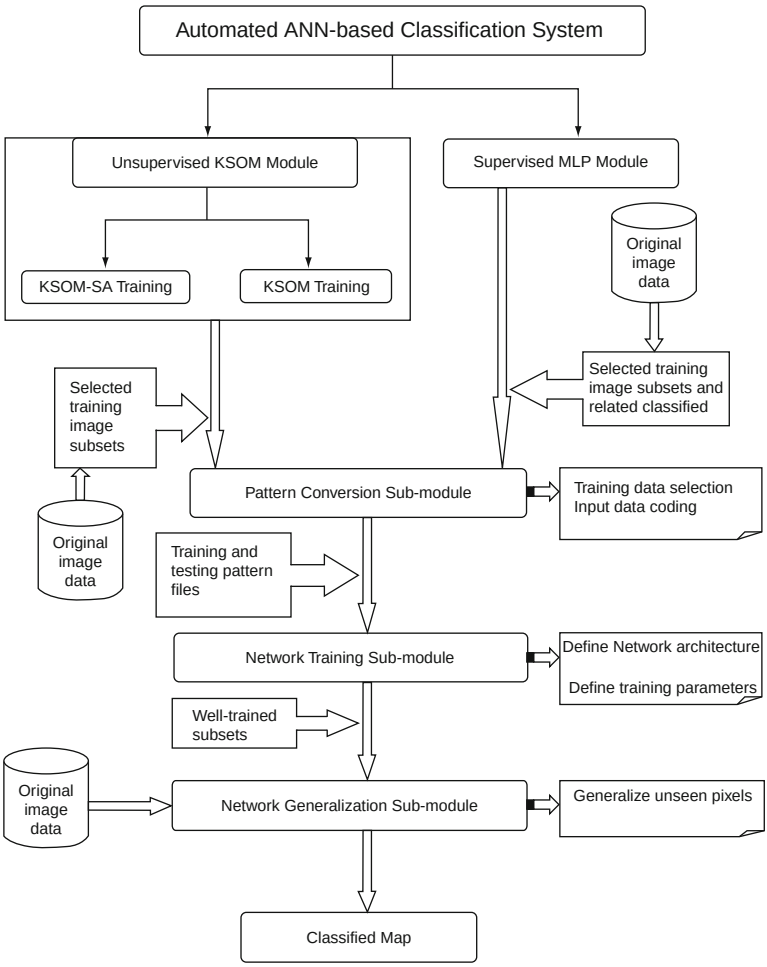


Fig. 33.8 Flowchart of an automated ANN-based classification system

accuracy for land cover classification. Simulated Annealing (SA) was developed on the basis of an analogy between the physical annealing process of solids and large combinatorial optimization problems (Kirkpatrick et al. 1983; See also Cerny 1985).

SA has the potential to find or approximate the global or near-global optimal in a combinatorial optimization problem. Although SA often requires greater computational time, the improved network can self-adapt to choose momentum parameters according to annealing temperature, thus enabling the network to escape from local minimum spots and converge stably. Mathematically, SA can be modeled by means of a Markov chain. The basic procedure involves a cooling schedule, in which a parameter called temperature T starts out sufficiently high and is gradually

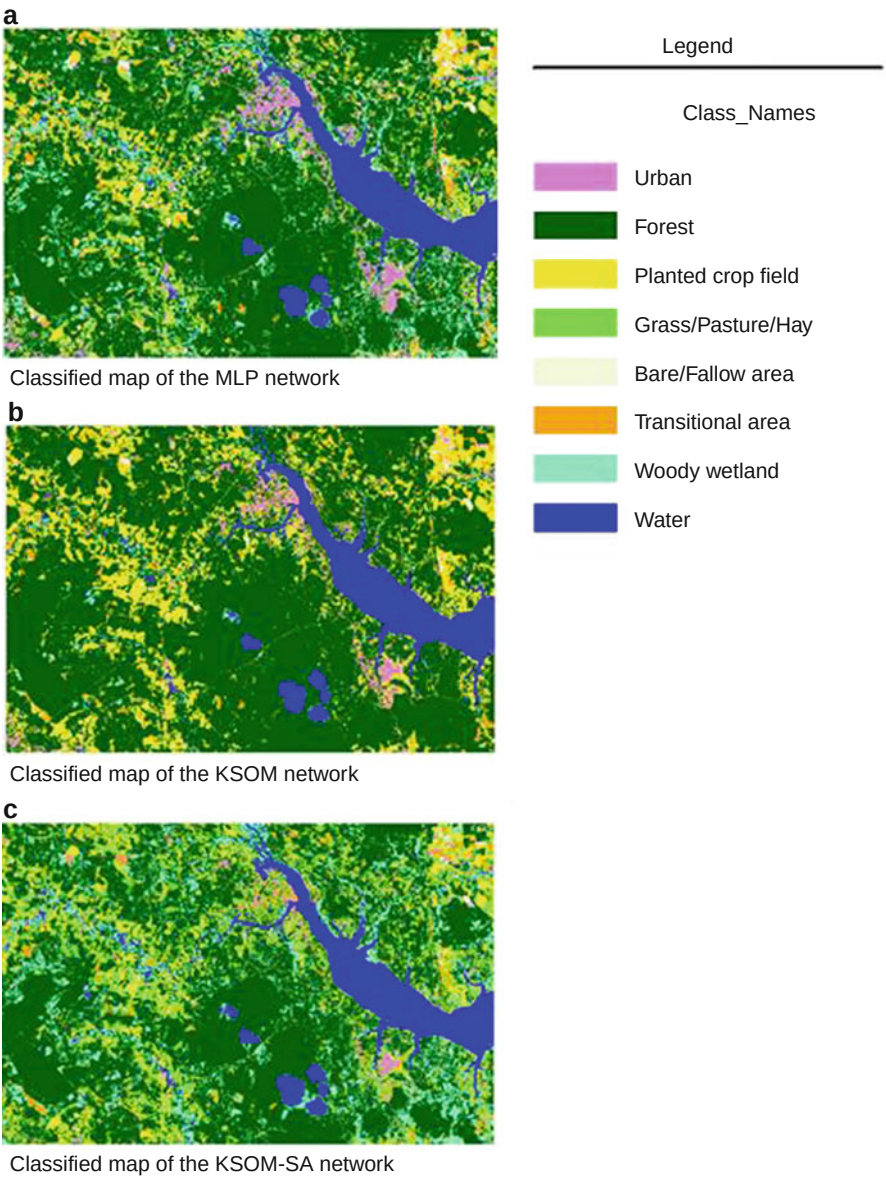


Fig. 33.9 (**a**–**c**) Classified images of Coastal North Carolina, USA, based on three versions of the ANN classification method (Graphic courtesy of USGS)

lowered in a given schedule to minimize the energy or cost function associated with a specific problem formulation. At each temperature T, a small, randomly generated, perturbation is repetitively applied until the system reaches thermal equilibrium at that temperature. Then the algorithm moves to the next temperature in the

given schedule. The rule of accepting a perturbation is based upon the *Metropolis criterion* (Shurr 1997). The original image and single Simulated Annealing and K-Means clustering classified maps, the Integrated K-means and Simulated Annealing (I-SA) are shown in Fig. 33.10a–d.

Fuzzy Logic Classifiers

In order for remotely sensed land use/land cover classification data to be the most useful, the post-classification results should resemble accurate representation of the ground features being investigated. The incorporation of unsupervised and supervised classification techniques, based on training sets and maximum likelihood or other spatial algorithms, develops classification groupings of objects belonging to exclusive categories. Often for general or moderate resolution land cover analysis these methods are acceptable. However, in situations where highly accurate land cover data is required, it is important to note that the real world often does not conform to the "hard" rules imposed by the algorithms or boundaries that result from processes such as maximum likelihood, nearest neighbor, etc.

Fuzzy logic classifiers (Fig. 33.11) may be used to circumvent the "hard" classification boundaries by allowing for a transition or "fuzzy" region to be established between classes when a pixel has membership in two or more classification categories (Jensen 2005). The fuzzy region is established based on the vector direction of the spectral measurement space distance from the means of all established classes. This gradation of class values allows for the class transition region or fuzzy class to represent the value of member each pixel has in the established classes.

Studies by Hester found the capacity of fuzzy logic sets was able to overcome post-classification accuracy limitations in the analysis of an urban watershed utilizing high-resolution data (Hester 2008; Hester et al. 2010). The use of fuzzy classifiers addressed two major sources of change detection error: individual-date map misclassification and image "misregistration." The calculated overall "from-to accuracy" achieved in this study was over 78% for all classes.

Each pixel is described as a "fuzzy neighborhood vector" and those vectors are compared between maps based on neighborhood size, distance-decay function, and neighborhood vector comparisons. The result is a "similarity value" that is a measure of each pixel's inter-map neighborhood similarity ranging from 0 to 1 (Hagen 2003).

Image Post-Processing and Smoothing

The post-processing phase is often conducted based on the objectives of the project. At times, filtering is needed for smoothing the data and removing speckles in the classified image for better correspondence to conventional survey maps. Other techniques utilized in post-processing include accuracy assessment of the classified images for verification and validation of the results; change detection and

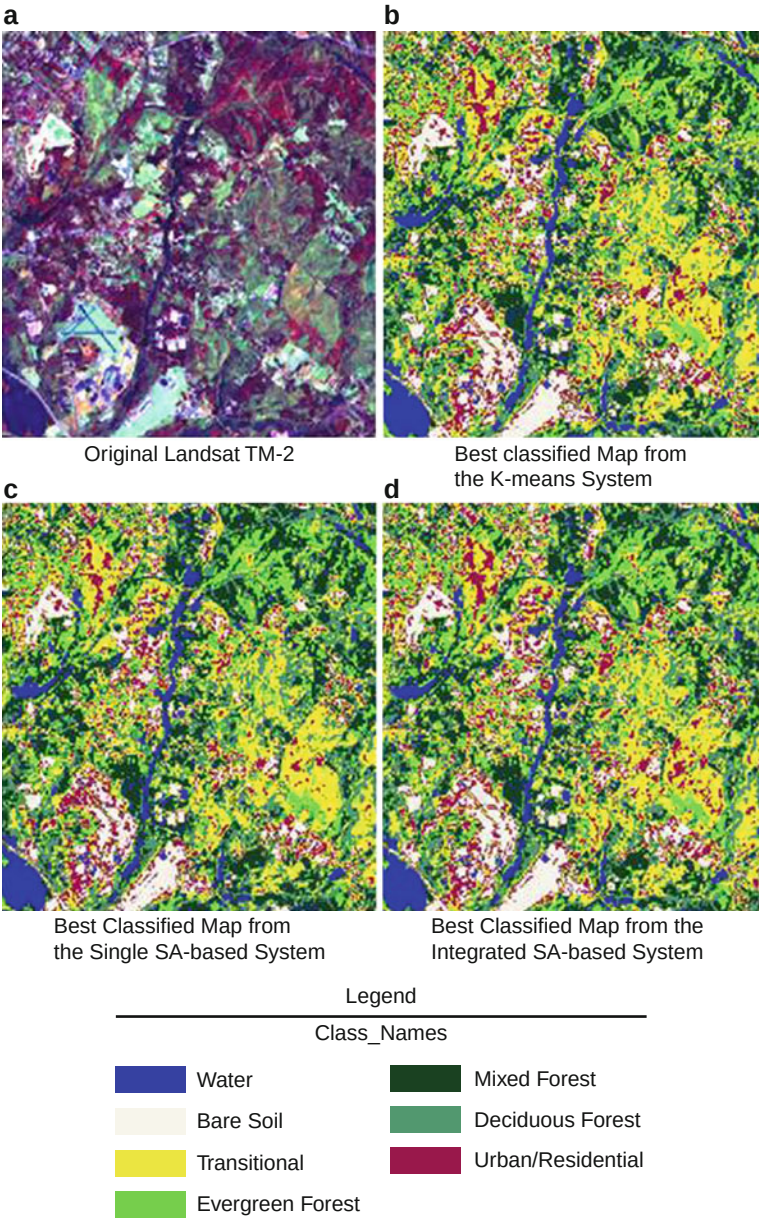


Fig. 33.10 (**a**–**d**) Land use/land cover using simulated annealing techniques (Graphic courtesy of NOAA)

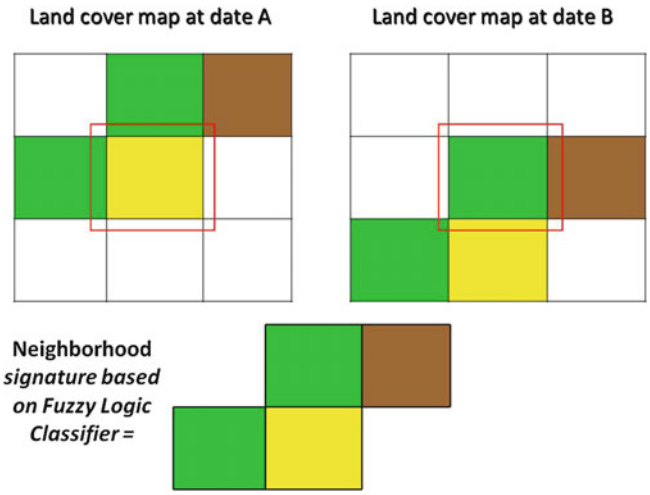


Fig. 33.11 Illustration of an application of fuzzy classification technique for land use/land cover mapping

monitoring based on two or more dates of classified images; and integration of the classified images with other conventional survey maps and data layers in a geographical information system (GIS) for geospatial modeling.

Filtering

After the data is classified, a number of algorithms can be applied to the classified image for better presentation and visualization. Kernels, which can be thought as moving windows, are applied to classified images to reduce noise and filter unwanted information. An example of a 5×5 median filter is shown in Fig. 33.12 which has been applied to a classified map. Median values are calculated within a moving 5 pixel by 5 pixel window and assigned into the central pixel. Mean and median filters usually have a smoothing effect. Other filters that have been frequently used in image processing use functions that highlight differences between pixel values. They are usually used for sharpening the edges of objects in an image.

Accuracy Assessment

Accuracy assessment has been a key component and the focus of a significant number of remote sensing studies (Van Genderen and Lock 1977; Congalton et al. 1983; Goodchild et al. 1992; Khorram et al. 1992; Congalton and Green 1999; Paine and Kiser 2003). Without assessing the accuracy of a classified data, the reliability and repeatability of the output products are in question.

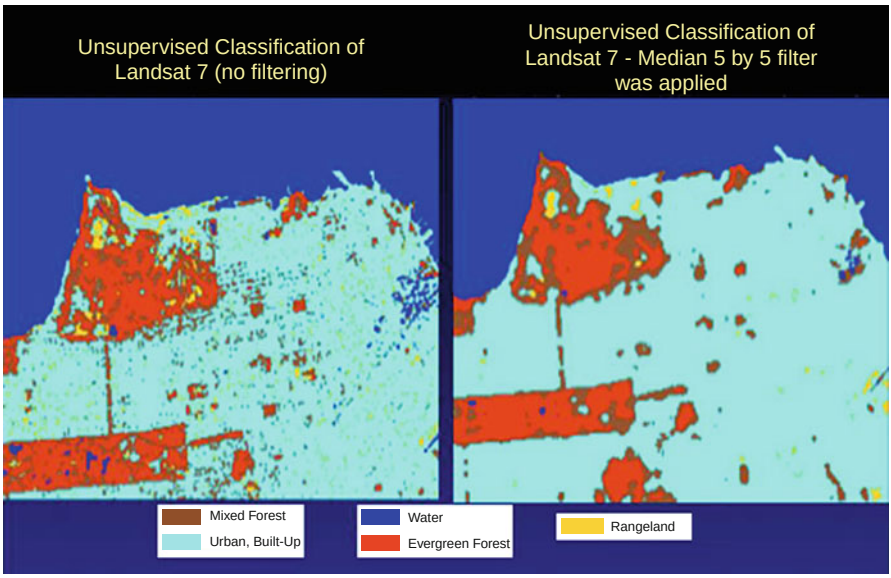


Fig. 33.12 A filtering technique has been applied to classified land use/land cover images on the *right* side as compared to a classified original image on the *left* side (Graphic courtesy of Siamak Khorram)

Sophisticated statistical procedures in the analysis of error matrices are developed for the accuracy assessment of land cover classifications for a single date (Congalton et al. 1983; Aaronoff 1985; Khorram et al. 1999; Stehman 2001; Foody 2002; Lunetta and Lyons 2003; Pal and Mather 2003; Morisette and Khorram 2007). However, the accuracy assessment of change detection procedures are not in operational stage yet and involve issues such as not yet widely accepted sampling techniques, image registration, boundary problems, and reference data. The error sources involved in the accuracy assessment include registration differences between reference data and remotely sensed data, delineation errors in digitizing, data entry errors, errors in image classification and delineation, and errors involved in sampling, collection, and interpretation of reference data.

The most commonly used procedure for accuracy assessment is error matrix analysis. An error matrix can be constructed by the results based on reference data (e.g., data collected on the ground or from substantially higher spatial resolution image) on one side of a table, and the results based on the classified image on the other side of the table. An adequate number of samples are identified on the classified image and corresponding reference data are collected using various sampling strategies. The accuracy is determined in terms of percent correctly classified sample sites, as compared to their corresponding reference data, for each category of interest as well as the overall classification accuracy involving all categories. Traditionally, the total number of correct samples in a given category

is divided by the total number of samples in that category based on reference data. This accuracy measure indicates omission errors and is often referred to as "producer's accuracy" because the producer of image classification is interested in how well he has classified a certain category.

If the total number of correct samples in a given category is divided by the total number of samples based on classified data, then this indicates the commission error. This measure is called the "user's accuracy" or reliability because the user is interested in the probability that a classified sample represents the actual category on the ground (Story and Congalton 1986). Multivariate statistical procedures have also been used for accuracy assessment. The most commonly used is a discrete multivariate technique, called KAPPA. The result of KAPPA analysis is a KHAT statistics, which is a measure of agreement or accuracy (Cohen 1960).

Change Detection

Global environmental change has become a major national and international policy issue. Land cover change provides an important component to estimate and model changes in environmental and socioeconomic conditions resulting from regulatory or land use policy changes and is a potentially important indicator of the effects of local, national, and international policies on environmental quality and even human health. Remote sensing, combined with supporting data, provides the most feasible approach to land cover change detection. Issues specific to change analysis include the land cover classification scheme, hard logic versus fuzzy logic, large number of categories of change, the difficulty of field verification for past time periods, errors involved in classifications for both dates, remote sensor systems characteristics, digital image processing techniques, and environmental characteristics.

There are various methods for change detection. Some changes are easily detected through the visual analysis of color composites for the dates under consideration. Figure 33.13 illustrates change detection based on visual analysis of Advanced Very High Resolution Radiometry (AVHRR) satellite data using an example of before and after oil wells were set on fire in Kuwait.

The most common change detection techniques are image algebra change detection and post-classification comparison change detection.

Image Algebra Change Detection

The changes between two dates of remotely sensed data may be quantified by using a simple image algebra band "ratioing" or differencing technique for corresponding bands in two dates. The resulting change image produced has positive and negative values (that can be transformed to positive values by adding or constant). A threshold for determining "change" and "no change" can be applied to the change image after examining the histogram (Elfishawy and Kesler 1991). Although this method does not provide "from-to" change information, it is simple and widely used.

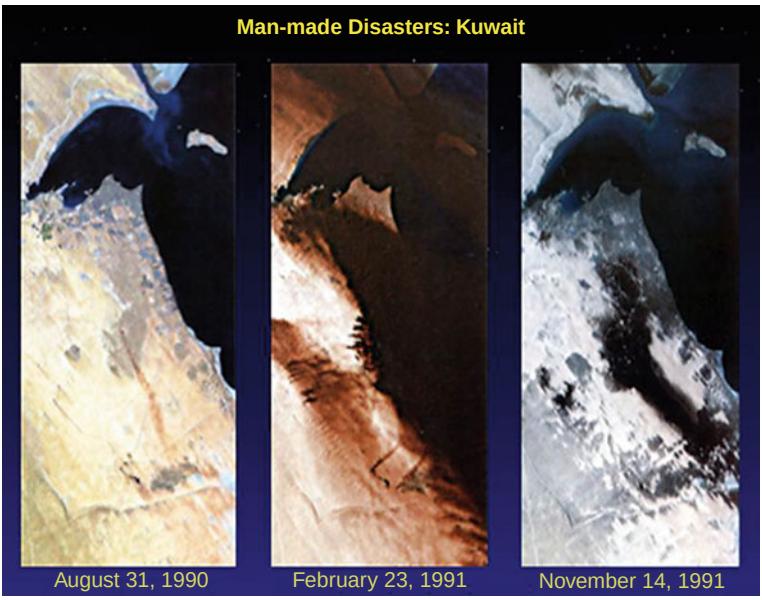


Fig. 33.13 Change detection as illustrated with images of before and after oil wells were set on fire in Kuwait (courtesy of Dai and Khorram 1997)

Post-Classification Comparison Change Detection

Post-classification comparison change detection methods are easy to use and require a complete classification of each individual date involved (Rutchey and Velcheck 1994. Also see Jensen 2005). Unfortunately, any error present in the classification of individual dates will be incorporated in the change detection map. The number of classes increases geometrically. For example, for a six-category image for each date, $6 \times 6 = 36$ category of change is formed for the change map. The analyst can highlight very specific changes from a few desired categories to another few special interest categories and ignore the other land use/land cover categories in order to get a understanding of the change analysis. This is a commonly used technique and provides detailed "from-to" change information. Figure 33.14 shows the changes in land use and land cover that was produced using the post-classification method (Dobson et al. 1995).

In this figure all changes are represented in various shades of gray.

Figure 33.15 illustrates detailed post-classification change detection as compared to the TCC of the same area.

Data Integration and Spatial Modeling

The integration of remotely sensed data into available spatial analyses packages, such as ESRI's ArcGIS, and powerful statistical packages offer a tremendous

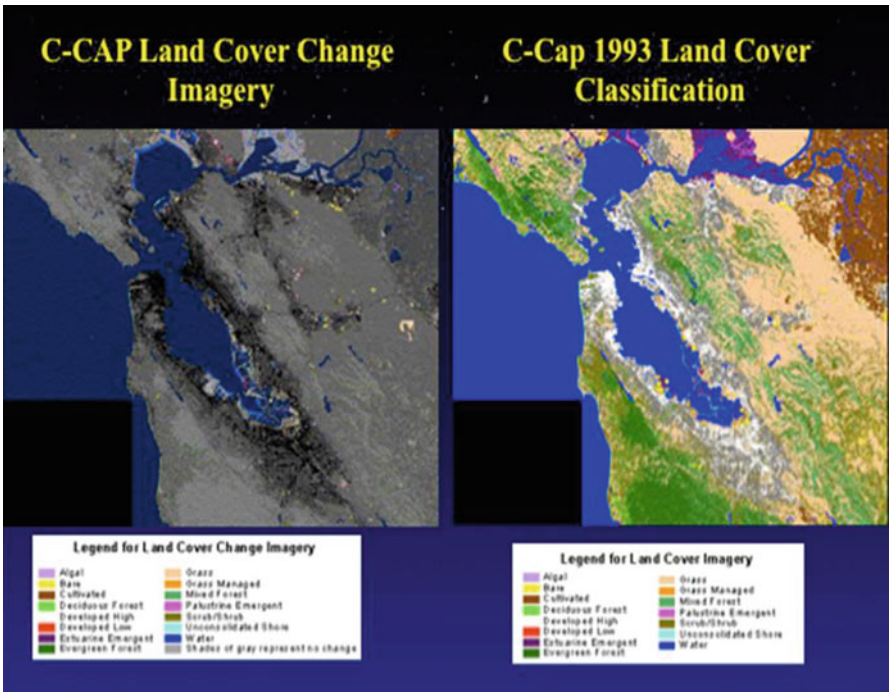


Fig. 33.14 Image of San Francisco Bay Area (courtesy of the coastal change analysis program of the U.S. National Oceanic and Atmospheric Administration)

opportunity for achieving historical and current information that is necessary for resource monitoring, mapping, and spatial modeling. The advantages provided by remote sensing include the ability to provide data collection and analysis over large and/or remote areas; a rapid, cost-effective method for monitoring, collecting, and analyzing data; the ability to collect, model, and analyze current and archival environmental data. Combining remotely sensed data with geographic information systems (GIS) applications further provide the possibility of developing spatial analysis procedures that take advantage of temporal and spatial information on the structure of the landscape. Additionally, combining remotely sensed image data with geographic referenced data sets within a GIS allow users to take advantage of the extensive spatial database management system that lends itself easily to combining other geographic and tabular data of the respective area. Spatial and statistical models may be developed within a GIS or statistics program to determine likelihood of occurrence, probabilities, pattern and cluster analysis, and density and distribution functions that utilize classified image data as a base data source.

The true power of using remotely sensed data in a GIS is the fact that classified (processed) image data may serve as an input data source within the GIS. The GIS is able to use data in the form of layers to visualize patterns and trends that exist within the data, as shown in Fig. 33.16. The interactions and relationships that occur

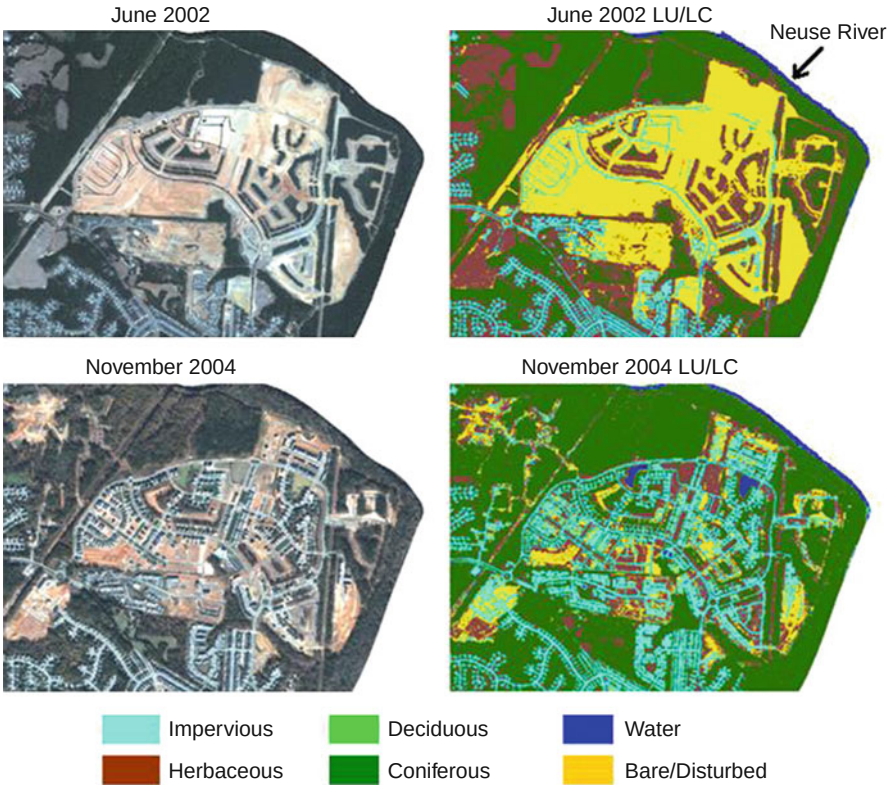


Fig. 33.15 Change detection using post-classification method (Graphic courtesy of Siamak Khorram)

between these data layers can then be taken into account in a geospatial modeling context. Geospatial modeling results are then used by researchers and resource managers to solve problems and make appropriate decisions based on location-derived attributes.

Despite the potential analysis and modeling power provided by combination of remotely sensed image data with GIS and statistical packages, development of integration protocols have only recently gained widespread attention. This may be due to differences in the data structures used to acquire and store data in the remote sensing and GIS environments. Digital imagery from spectral remote sensing detectors are stored as a raster data model. This data model is usually composed of a series of homogenous pixels organized in a row and column arrangement to represent the area of the earth or targeted feature that the image data has been acquired over. The most popular data structures within a GIS are the vector file-based data models. Although many GIS allow for the use of raster data, problems may arise if the raster data is not available in a format that is readily usable for spatial analysis purposes. For example, unclassified (raw) raster data may require

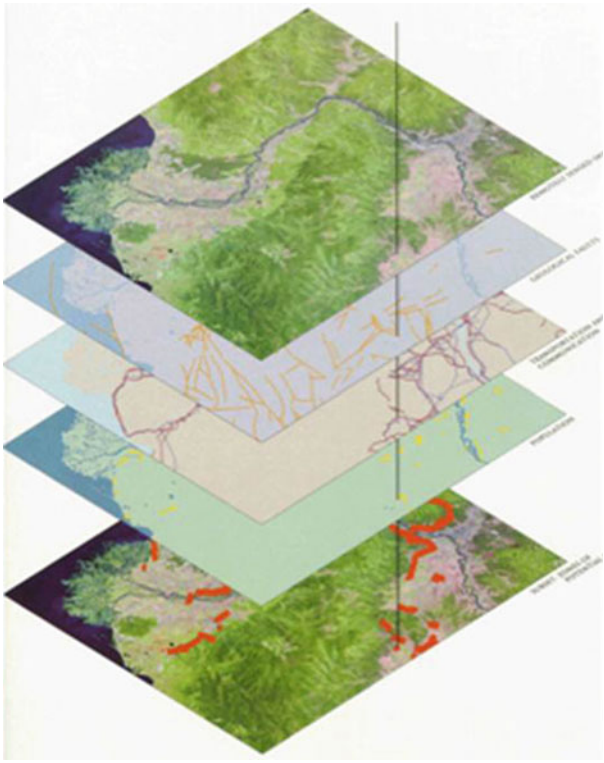


Fig. 33.16 Data input layers for use in GIS (Graphic courtesy of Siamak Khorram)

a considerable amount of processing outside of the GIS environment before it can be used for spatial modeling purposes.

Additionally, errors may arise from interpreting data accuracy of models developed from the combination of raster and vector data sources. This is largely due to the fact that accuracy assessments for remotely sensed data classifications are generated using the error matrix method. This method provides accuracy information on a global-image scale. Accuracy information developed within a GIS typically provides an analysis of spatial error on a local scale. Given this difference, care must be used to ensure accuracy assessment techniques are appropriate for all data types used in the analysis (Wang 1991. Also see Khorram et al. 1999).

Conclusion

The techniques described in this chapter are commonly used for processing airborne and satellite data for use by a wide range of end users. Image processing techniques evolve as the higher resolution data becomes more readily available from a variety of airborne and spaceborne platforms. With rapid technological advances in computer processing and cloud computing capabilities, removal of communication bottlenecks, improvements in screen technology, and the

increase in high-volume data storage capacity, the future of remote sensing and geospatial technologies provide exciting opportunities for new applications in Earth observation and planetary exploration.

Notes

1. Op cit. Hester 2008
2. Op cit. Cowardin et al.
3. Op cit. Lillesand et al.

References

S.G. Aaronoff, The minimum accuracy value as an index of classification accuracy. Photogrammetr. Eng. Remote Sens. **57**(5), 501–509 (1985)

J.R. Anderson, E. Hardy, J. Roach, R. Witmer, A land use and land cover classification system for use with remote sensing data, US Geological Survey Professional Paper 964, Washington, DC, 1976, p. 28ff.

E.A. Blaisdell, *Statistics in Practice* (Harcourt Brace Javanovich, New York, 1993), p. 653ff

T. Celik, Unsupervised change detection in satellite images using principal component analysis and k-means clustering. Geoscien. Remote Sens. Lett. IEEE **6**(4), 772–776 (2009)

V. Cerny, Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm. J. Optim. Theory Appl. **45**, 45–51 (1985)

S.B. Cho, J.H. Kim, Combining multiple neural networks by fuzzy integral for robust classification. IEEE Trans. Syst. Man Cybern. **25**(2), 380–384 (1995)

J.A. Cohen, A coefficient of agreement for nominal scales. Educ. Psychol. Meas. **20**, 37–46 (1960)

R.G. Congalton, K. Green, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices* (Lewis, Boca Raton, 1999), 137 p

R.G. Congalton, R.G. Oderwald, R.A. Mead, Assessing landsat classification accuracy using discrete multivariate statistical techniques. Photogrammetr. Eng. Remote Sens. **49**(12), 1671–1678 (1983)

L.M. Cowardin, V. Carter, F.C. Golet, E.T. LaRoe, *Classification of Wetlands and Deepwater Habitats of the United States* (U.S. Fish and Wildlife Service, Washington, 1979), p. 103ff. FWS/OBS-79/31

X. Dai, S. Khorram, A new automated land cover change detection system for remotely-sensed imagery based on artifical neural networks, in *Proceedings of the IEEE/IGARSS 1997 International Geoscience and Remote Sensing Symposium*, Singapore, 1997

X. Dai, S. Khorram, Quantification of the impact of misregistration on digital change detection accuracy, in *Proceedings of the IEEE/IGARSS 1997 International Geoscience and Remote Sensing Symposium*, Singapore, 1997

X. Dai, S. Khorram, Data fusion using artificial neural networks: a case study on multitemporal change analysis. Comput. Environ. Urban Syst. **23**, 19–31 (1999)

A. Das, B.K. Chakrabarti, Quantum annealing and related optimization methods. Lect. Notes Phys. **679**, 239–257 (2005)

J. De Vincente, J. Lanchares, J. Hermida, Placement by thermodynamic simulated annealing. Phys. Lett. A **317**, 415–423 (2003)

J.R. Dobson, E.A. Bright, R.L. Ferguson, D.W. Field, L.L. Wood, K.D. Haddad, H. Iredale, J.R. Jensen, V. Klemas, R.J. Orth, J. P. Thomas, NOAA Coastal Change Analysis Program (C-CAP); Guidance for Regional Implementation, National Oceanic & Atmospheric Administration, Washington, NMFS **123**, 1995, p. 92ff

A.S. Elfishawy, S.B. Kesler, Adaptive algorithms for change detection in image sequence. Signal Process. **23**, 179–191 (1991)
G.M. Foody, Status of land cover classification accuracy assessment. Remote Sens. Environ. **80**, 185–201 (2002)
M.F. Goodchild, G.Q. Sun, S. Yang, Development and test of an error model for categorical data. Int. J. Geogr. Inf. Syst. **6**(2), 87–104 (1992)
A. Hagen, Fuzzy set approach to assessing similarity of categorical maps. Int. J. Geogr. Inf. Sci. **17**, 235–249 (2003)
D.B. Hester, Land cover mapping and change detection in urban watersheds using quick bird high spatial resolution satellite imagery, Ph.D. Dissertation, North Carolina State University, 2008, p. 148
D.B. Hester, H.I. Cakir, S.A.C. Nelson, S. Khorram, Per-pixel classification of high spatial resolution satellite imagery for urban land cover mapping. Photogramm. Eng. Remote Sens. **74**, 463–471 (2008)
D.B. Hester, S.A.C. Nelson, H.I. Cakir, S. Khorram, H. Cheshire, High resolution land cover change detection based on fuzzy uncertainty analysis and change reasoning. Int. J Remote Sens. **31**, 455–475 (2010)
R.M. Hord, *Digital Image Processing of Remotely-Sensed Data* (Academic, New York, 1982), p. 256
A.K. Jain, *Fundamentals of Digital Image Processing* (Prentice Hall, Englewood Cliffs, 1989), pp. 418–421
J.R. Jensen, *Introductory Digital Image Processing*, Third edn. (Pearson Prentice Hall, Upper Saddle River, 2005). 316p
I. Kanellopoulos, G.G. Wilkinson, Strategies and best practice for neural network image classification. Int. J. Remote Sens. **18**, 711–725 (1997)
S. Khorram, H.M. Cheshire, K. Sidrellis, Z. Nagy. Mapping and GIS Development of Land Use/Land Cover Categories for the Albemarle-Pamlico Drainage Basin, NC Department of Environmental, Health, and Natural Resources, Dept. No. 91–08, 1992, p. 55ff
S. Khorram, H. Cheshire, X. Dai, J. Morisette, Land cover inventory and change detection of coastal North Carolina using landsat thematic mapper data. ASPRS/ACSM – Annu. Conv. Expos. **1**, 245–250 (1996). Remote Sensing and Photogrammetry
S. Khorram, G.S. Biging, N.R. Chrisman, D.R. Colby, R.G. Congalton, J.E. Dobson, R.L. Ferguson, M.F. Goodchild, J.R. Jensen, T.H. Mace, *Accuracy Assessment of Remote Sensing-Derived Change Detection*. American Society of Photogrammetry and Remote Sensing, Monograph (1999)
S. Kirkpatrick, C.D. Gelatt Jr., M.P. Vecchi, Optimization by simulated annealing. Science **220**, 671–688 (1983)
V.V. Klemas, J.E. Dobson, R.L. Ferguson, K.D. Haddad, A coastal land cover classification system for the NOAA coastWatch change analysis program. J. Coast. Res. **9**(3), 862–872 (1993)
T. Lillesand, R. Kiefer, J. Chipman, *Remote Sensing and Image Interpretation*, 6th edn. (Wiley, New York, 2008). 763p
R.L. Lunetta, J.G. Lyons (eds.), Geospatial data accuracy assessment. Report No. EPA/600/R-03/064 (US Environmental Protection Agency, Las Vegas, 2003), 335 p
J.T. Morisette, S. Khorram, Exact confidence interval with portions. Photogrammetr. Eng. Remote Sens. 5 (2007)
NOAA, Coastal Change Analysis Program (C-CAP), (Charleston: NOAA Coastal Services Center, 2004), http://www.csc.noaa-gov/crs/lca/ccap_program.html.
Y. Nogami, Y. Jyo, M. Yoshioka, S. Omatu, Remote sensing data analysis by Kohonen feature map and competitive learning. IEEE SMC'97 **1**, 524–529 (1997)
D.P. Paine, J.D. Kiser, Chapter 23: mapping accuracy assessment, in *Aerial Photography and Image Interpretation*, 2nd edn. (Wiley, New York, 2003), pp. 465–480
M. Pal, P.M. Mather, An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sens. Environ. **86**, 554–565 (2003)

F. Qiu, J.R. Jensen, Opening of black box of neural networks for remote sensing image classification. Int. J. Remote Sens. **9**, 1749–1768 (2004)
D.E. Rumelhart, G.E. Hinton, R.J. Williams, *Parallel Distributed Processing* (MIT Press, Cambridge, MA, 1986)
K. Rutchey, L. Velcheck, Development of an everglades vegetation map using a SPOT image and global positioning system. Photogramm. Eng. Remote Sens. **60**(6), 767–775 (1994)
M.J. Sabins, Convergence and consistency of fuzzy C-means/ISODATA algorithms. IEEE Trans. Pattern Anal. Mach. Intell. **9**, 661–668 (1987)
P.C. Shurr, Acceptance of the acceptance criteria for the simulated annealing algorithm. Math. Oper. Res. **22**(2), 266ff (1997)
S.V. Stehman, Statistical rigor and practical utility in thematic map accuracy assessment. Photogrammetr. Eng. Remote Sens. **67**, 727–734 (2001)
M. Story, R.G. Congalton, Accuracy assessment: a user's perspective. Photogrammetr. Eng. Remote Sens. **52**(3), 397–399 (1986)
J.T. Tou, R.C. Gonzalez, *Pattern Recognition Principles* (Addison-Wesley, Readings, 1977). 377p
J.L. Van Genderen, B.F. Lock, Testing land use map accuracy. Photogramm. Eng. Remote Sens. **43**(9), 1135–1137 (1977)
USGS, USGS National Land Cover Data, (EROS Data Center, Sioux Falls, 2004), http://landcover.usgs.gov/prodescription.html.
F. Wang, Integrating GIS and remote sensing image analysis systems by unifying knowledge representation scheme. IEEE Trans. Geosci. Remote Sens. **29**, 656–664 (1991)
R. Xu, D. Wunsch, Survey of clustering algorithms. IEEE Trans. Neural Netw. **16**, 32f (2005)
C. Yang, P. Chung, Knowledge-based automatic change detection positioning system for complex heterogeneous environments. J. Intell. Robotic Syst. **33**, 85–98 (2002)

Remote Sensing Data Applications

34

Haruhisa Shimoda

Contents

Introduction 866
Atmospheric Applications 867
Radiative Transfer and Inversion Problem 867
Temperature and Water Vapor 868
Aerosols and Clouds 870
Atmospheric Constituents 872
Greenhouse Gases 877
Oceanic Applications 881
Sea Surface Temperature 881
Sea Surface Salinity 882
Sea Surface Wind 882
Sea Surface Height 886
Ocean Color 887
Land Applications 891
Topography 891
Geometric Corrections and Map Projection 893
Radiometric Corrections 894
Land Cover and Land Use 896
Geological Applications 905
Soil Moisture 906
Carbon Cycle 909
Cryospheric Applications 914
Sea Ice 914
Snow and Glaciers 915
Operational Applications 917
NWP and Weather Forecasting 917

H. Shimoda
Earth Observation Research Center, Aerospace Exploration Agency, 2-1-1, Sengen, Tsukuba, 305-8505, Japan
and
Research & Information Center, Tokai University, Shibuya-ku, Tokyo, 151-0063, Japan
e-mail: smd@keyaki.cc.u-tokai.ac.jp

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_47, © Springer Science+Business Media New York 2013

Fisheries 918
Disasters 920
Ship Navigation 922
Agriculture 924
Cross-References 929
Notes 929
References 930

Abstract

Application areas of remote sensing are very wide. They can be divided into two areas. One is applications in the Earth environmental monitoring and process studies of the Earth system and another is operational applications. The former can be divided into atmosphere, ocean, land, cryosphere, and their interactions. In this chapter, temperature, water vapor, aerosols and clouds, atmospheric constituents, greenhouse gases, sea surface temperature, sea surface salinity, sea surface wind, ocean color, sea surface height, topography, land cover, soil moisture, carbon cycle, sea ice, snow, and glaciers are described. The latter has wide variety. This chapter cannot cover all the operational application areas. Among them, Numerical Weather Prediction (NWP) and weather forecasting, fisheries, disasters such as biomass burnings, floods, ship navigations and agriculture are described. In addition to these application areas, some basic processings for applications are also described. These processings include radiative transfer and inversion problem, geometric and radiometric corrections, and classification algorithms.

Keywords

Aerosol • Atmospheric constituents • Carbon cycle • Cloud • Geometric correction • Greenhouse gases • Land cover • Ocean color • Radiative transfer • Radiometric correction • Sea ice • Sea surface height • Sea surface salinity • Sea surface temperature • Sea surface wind • Snow • Soil moisture • Temperature • Topography • Water vapor • Weather forecasting

Introduction

Application areas of remote sensing are very wide. The application areas at the start of remote sensing in the 1960s were mostly operational applications, i.e., land cover/use, mineral explorations, agriculture, forest industry, fisheries, and disasters. However, recent development of global change has created new application fields, i.e., Earth environmental monitoring. In order to understand the global change, global geophysical parameters (e.g., atmospheric constituents like greenhouse gases, ozone, etc.), large-scale forest decrease, desertification, etc., should be measured and monitored for a long time. It is almost impossible to monitor these global phenomena using conventional in situ measurements. It is critical to use remote sensing for these purposes.

In this chapter, measurement and monitoring techniques for each sphere, i.e., atmosphere, ocean, land, and cryosphere, are described. After these descriptions, some of the operational fields, like weather forecast, fisheries, disaster, etc., are described.

In addition to these application areas, some basic processings for applications are also described. These processings include radiative transfer and inversion problem, geometric and radiometric corrections, and classification algorithms.

Atmospheric Applications

Radiative Transfer and Inversion Problem

In remote sensing, sensors on satellites receive electromagnetic waves from the Earth. However, these electromagnetic waves are reflected or emitted from the ground or from the atmosphere and absorbed or scattered by the atmosphere and finally arrive at the sensor. So, the received electromagnetic waves by the sensor can be described using radiative transfer codes. When light of spectral radiance I_λ propagates ds in a medium, radiative transfer equation can be written as follows.

$$\frac{dI_\lambda}{K_\lambda \rho ds} = -I_\lambda + S_\lambda$$

Here,

I_λ: Spectral radiance of wavelength λ

K_λ: Mass attenuation coefficient

ρ: Density of medium in which wave propagates

S_λ: Source function of radiation

and

$$K_\lambda = \chi_\lambda + \sigma_\lambda$$

χ_λ: Mass absorption coefficient

σ_λ: Mass scattering coefficient

$$S_\lambda \equiv j_\lambda / K_\lambda$$

where

j_λ: Mass emission coefficient

Unfortunately, this equation cannot be solved analytically. So, many kinds of approximation codes are developed and used. Most popular codes are MODTRAN (Berk et al. 1998), 6S (Kotchenova et al. 2006), RSTAR (Nakajima et al. 2004), and LBLRTM (Clough et al. 2005). LBLRTM is used for thermal infrared region, while other codes are mainly used in visible and near-infrared region. AtmosphericRadiative Transfer Simulator (ARTS) (Buehler et al. 2005) is sometimes used for the microwave region.

In the visible and near-infrared region, main atmospheric effects are scattering by atmospheric molecules and aerosols. The former can be modeled by Rayleigh scattering (Rayleigh 1871) and the latter can be modeled by Mie scattering (Mie 1908). In order to calculate atmospheric attenuation of light by these scatterings, characteristics of atmospheric molecules and aerosols are necessary. These characteristics are usually given from MODTRAN. On the other hand, molecular absorption characteristics are necessary to calculate in the infrared region. These characteristics are given by databases. Most popular databases are HITRAN (Rothman et al. 2009) and GEISA (Jacquinet-Husson et al. 2009).

In order to retrieve geophysical parameters from data obtained by satellite-borne sensors, it is necessary to invert the radiative transfer equation. Usually, this process is an ill-posed problem and many kinds of inversion algorithms are proposed. The most popular algorithms are MAP (maximum a posteriori probability) (Rodgers 2000) and artificial neural networks. Here, a brief description of MAP algorithm is described.

$$\hat{\boldsymbol{x}} = \boldsymbol{x}_a + \left(\boldsymbol{K}^T \boldsymbol{S}_\varepsilon^{-1} \boldsymbol{K} + \boldsymbol{S}_a^{-1}\right)^{-1} \boldsymbol{K}^T \boldsymbol{S}_\varepsilon^{-1} (\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}_a)$$

Here,

$\hat{\boldsymbol{x}}$: Retrieved geophysical parameter (vector)
$\boldsymbol{x}$: A priori parameter (vector)
$\boldsymbol{y}$: Observed data (vector)
$\boldsymbol{S}_\varepsilon$: Variance–covariance matrix of observation noise
$\boldsymbol{S}_a$: Variance–covariance matrix of the parameter
$\boldsymbol{K}$: Jacobian matrix

$$\boldsymbol{K} = \partial \boldsymbol{F}(\boldsymbol{x}) / \partial \boldsymbol{x}$$

where,

$\boldsymbol{F}(\boldsymbol{x})$: Observed radiance spectra

Temperature and Water Vapor

Atmospheric temperature and water vapor are the most important geophysical parameters for weather forecasting. There are two ways to measure these variables from space. One is to use optical sensors and another is to use microwave sensors. For the optical sensor, high spectral resolution IR optical sensors are usually used. Two kinds of sensors are now used. One is a grating spectrometer and another is a Fourier transform IR spectrometer. AIRS (▶ Chap. 1, "Satellite Applications Handbook: The Complete Guide to Satellite Communications, Remote Sensing, Navigation, and Meteorology") on EOS Aqua (▶ Chap. 1) is an example of the former instrument, while IASI (▶ Chap. 1) on METOP (▶ Chap. 1) is an example of the latter instrument. Both instruments can measure temperature with 0.1 K accuracy and water

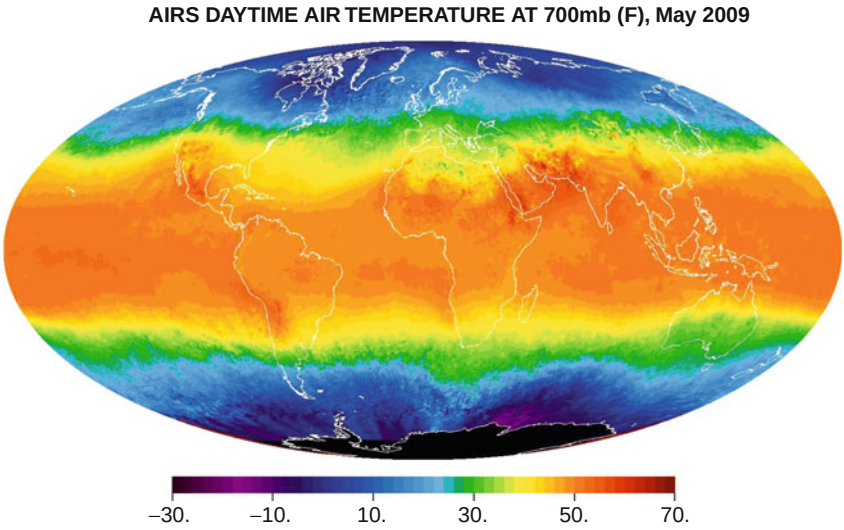


Fig. 34.1 Global atmospheric temperature at 700 hPa measured by AIRS on Aqua (Global atmospheric temperature at 700 hPa measured by AIRS on Aqua, http://photojournal.jpl.nasa.gov/catalog/PIA12098) (Courtesy of NASA)

vapor in 10% accuracy with 1 km vertical resolution. Another kind of optical sensor which can measure column density of water vapor over land uses near-infrared water vapor absorption lines. The disadvantage of optical sensors is that they cannot measure parameters under clouds. Figures 34.1 and 34.2 show global atmospheric temperature and water vapor measured by AIRS (▶ Chap. 1) on Aqua (▶ Chap. 1).

Vertical profiles of temperature and water vapor also can be retrieved by microwave sensors like AMSU (▶ Chap. 1) on METOP (▶ Chap. 1). Temperature is measured mainly using 50 GHz band, while water vapor is measured using 183 GHz band. The accuracy of microwave sounders is not so good as optical sounders. The vertical resolution depends on the number of channels in each absorption band. The advantage of microwave sounders is that it can measure over clouds. Also, microwave instruments can measure column water vapor using 23 and 31 GHz bands.

Another way of retrieving temperature and water vapor vertical distribution is to use GPS occultation technologies. GPS signal received at the satellites through atmosphere are refracted by atmosphere. The extent of this refraction depends on temperature and water vapor concentrations of the atmosphere. Hence, temperature and water vapor concentrations can be inferred from the extent of the refraction (Melbourne, et al. 1994; Kursinski et al. 1997). COSMIC (▶ Chap. 1) system on FORMOSAT (▶ Chap. 1) are now composed of six satellites carrying GPS occultation instruments, and several other satellites also carry GPS occultation instruments. From these instruments several thousand measurements are done in 1 day now.

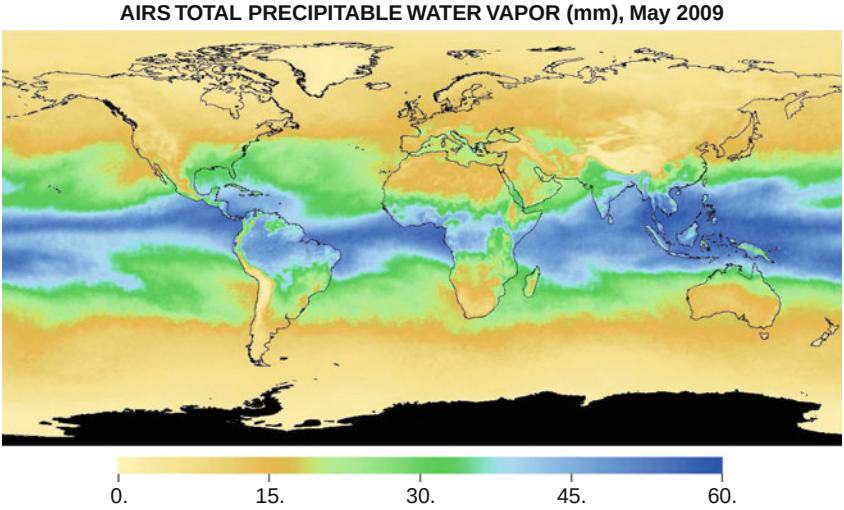


Fig. 34.2 Global atmospheric water vapor (total precipitable water) measured by AIRS on Aqua (Global atmospheric water vapor (total precipitable water) measured by AIRS on Aqua, http://photojournal.jpl.nasa.gov/catalog/PIA12097) (Courtesy of NASA)

Aerosols and Clouds

Aerosols and clouds are one of the most important parameters which can be measured from space. In the last IPCC report, aerosols and clouds are still the most uncertain radiative forcings in the climate models. There are two ways to measure aerosols and clouds from space. One is to use passive optical sensors and the other is to use active sensors, i.e., lidar and radar. In the passive system, aerosols are usually measured using visible wavelengths. Over the ocean, it is rather easy. In most of the methods, near-infrared or short wave infrared wavelength is used to acquire aerosol-free reflection from the ocean and this information is used to infer visible wavelength Mie scattering properties. In most cases, retrieved geophysical parameters are optical thickness and Angstrom exponent. The former is the total quantity of aerosols, and the latter corresponds to aerosol particle radius. Figure 34.3 shows a typical example of this type of aerosol retrieval obtained from visible and near-infrared channels of GLI (▶ Chap. 1) on ADEOS2 (▶ Chap. 1). Other important parameters are kinds of aerosols, refractive indices, and height of aerosols.

Over land, the retrievals are rather difficult. Ocean is very homogeneous and dark, but land is very inhomogeneous and bright. Two kinds of aerosol retrievals over land are tried using ultraviolet wavelength and polarization. Ultraviolet absorbing aerosols such as dust and soot can be retrieved using ultraviolet wavelengths. Figure 34.4 shows aerosol distribution retrieved from TOMS (▶ Chap. 1) using ultraviolet and Fig. 34.5 shows aerosol retrievals over land using POLDER (▶ Chap. 1) on ADEOS (▶ Chap. 1) using polarizations.

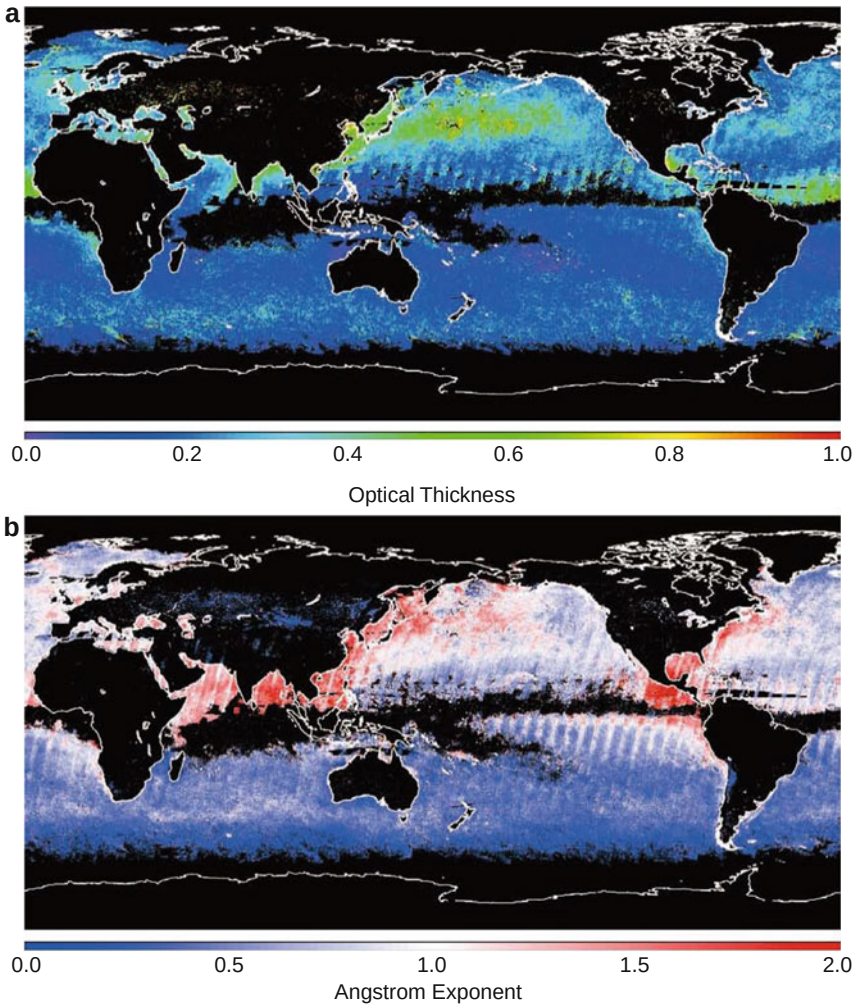


Fig. 34.3 Global aerosol distribution over ocean retrieved from GLI on ADEOS2 (Nakajima et al. 2009). (**a**) Aerosol optical thickness, (**b**) Aerosol Angstrom exponent (Courtesy of JAXA)

Active sensors, i.e., lidar and radar, can retrieve vertical distribution of aerosols and clouds. However, it is very difficult to retrieve horizontal distributions using these sensors. For aerosol retrieval, Mie scattering lidar is used, while for cloud retrieval, cloud profiling radar (CPR) is used. CPR is a very high-frequency radar using around 94 GHz frequency. Figure 34.6 shows an example of aerosol vertical distribution retrieved from Mie lidar CALIOP (▶ Chap. 1) on CALIPSO (▶ Chap. 1) data and Fig. 34.7 shows an example of cloud vertical distribution retrieved from CPR (▶ Chap. 1) on CLOUDSAT (▶ Chap. 1).

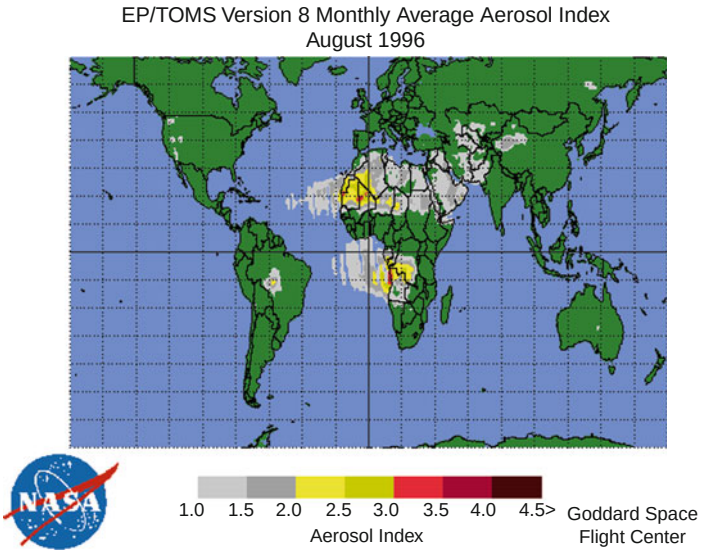


Fig. 34.4 Aerosols over land from TOMS (EP/TOMS version 8 monthly average aerosol index, ftp://toms.gsfc.nasa.gov/pub/eptoms/images/monthly_averages/aerosol/IM_aiavg_ept_199608.png) (Courtesy of NASA)

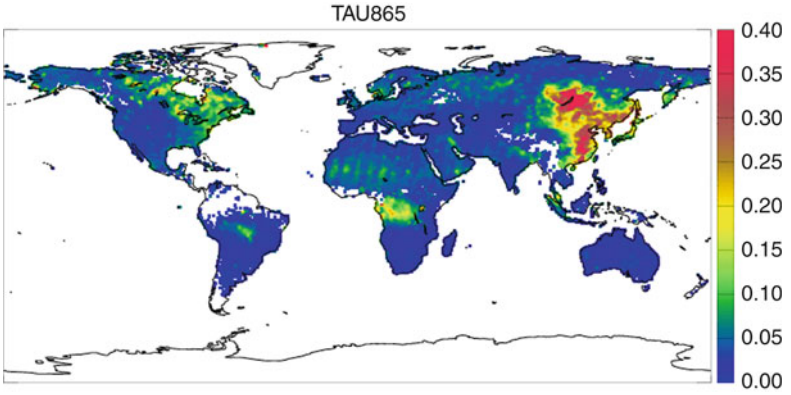


Fig. 34.5 Aerosols over land and ocean using POLDER(Aerosols results over land, http://smsc.cnes.fr/POLDER/A_produits_scie.htm) (Courtesy of CNES)

Atmospheric Constituents

Many kinds of atmospheric constituents can be measured from space. Figure 34.8 shows the atmospheric absorption spectrum of major constituents. From this figure, wide range of electromagnetic wavelength can be used for the

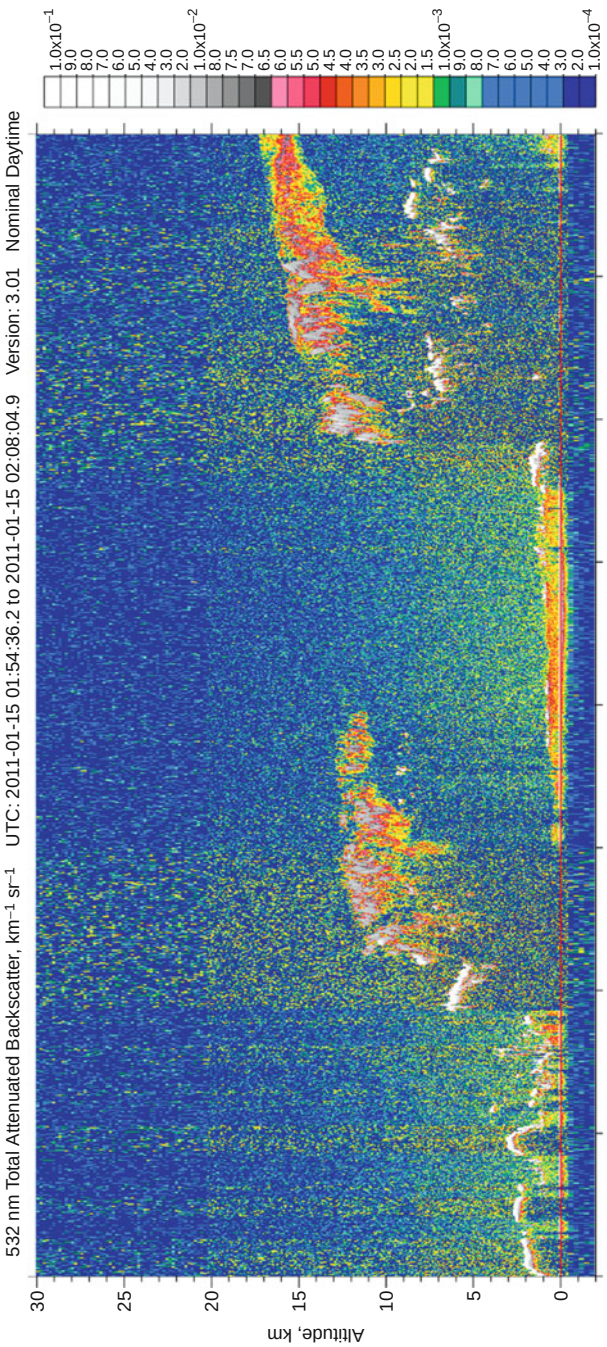


Fig. 34.6 Aerosol *vertical* distribution from CALIOP (Lidar level 1 browse images – 2011-01-15 01:41:08Z – Section 2, http://www-calipso.larc.nasa.gov/products/lidar/browse_images/show_detail.php?s=production&v=V3-01&browse_date=2011-01-15&orbit_time=01-41-08&page=2&granule_name=CAL_LID_L1-ValStage1-V3-01. 2011-01-15T01-41-08ZD.hdf) (Courtesy of NASA)

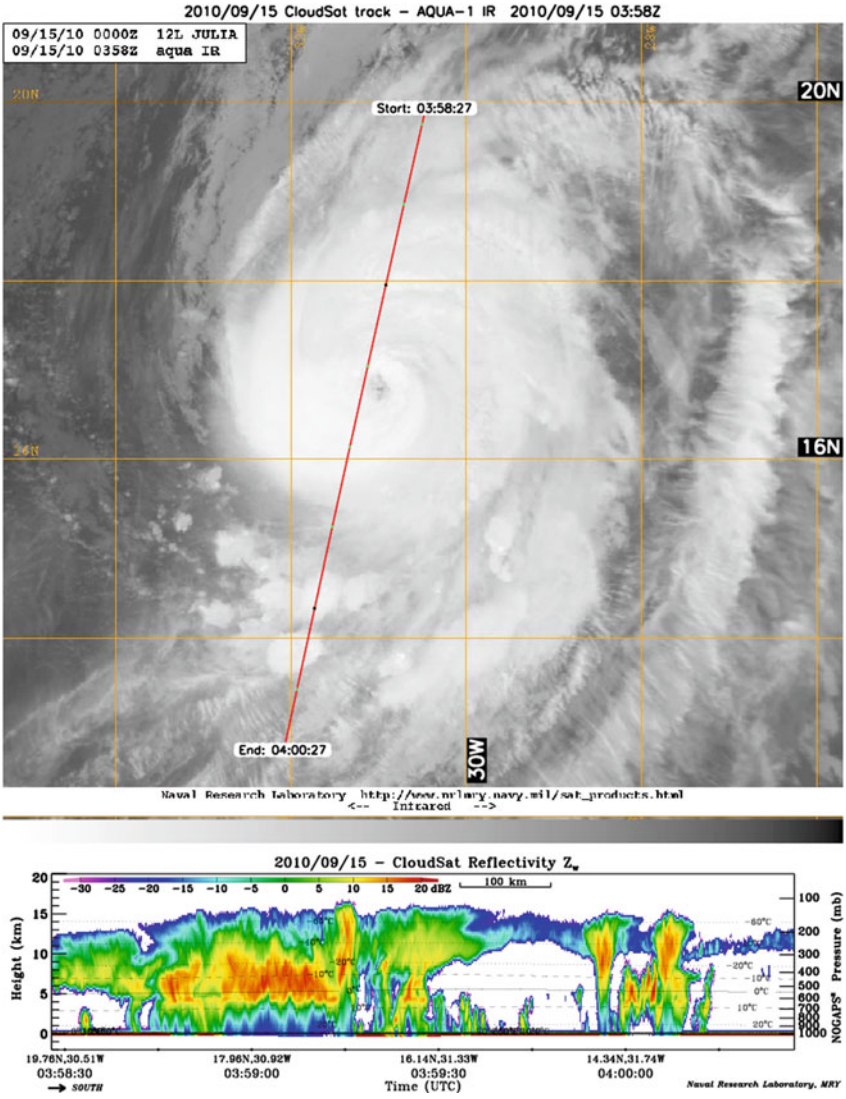


Fig. 34.7 Cloud *vertical* distribution from Cloudsat (http://cloudsat.atmos.colostate.edu/2010tcs/20100915julia.jpg) (Courtesy of Colorado State Univ.)

measurements. Usually, ultraviolet (UV), visible, and near-infrared, thermal infrared, microwave, millimeter, and submillimeter regions are used for measuring atmospheric constituents.

There are two ways of measurements. One is to observe the Earth in the vertical direction (nadir observation) and another is to observe in the limb direction (limb observation) as shown in Fig. 34.9. The latter can be divided into two methods

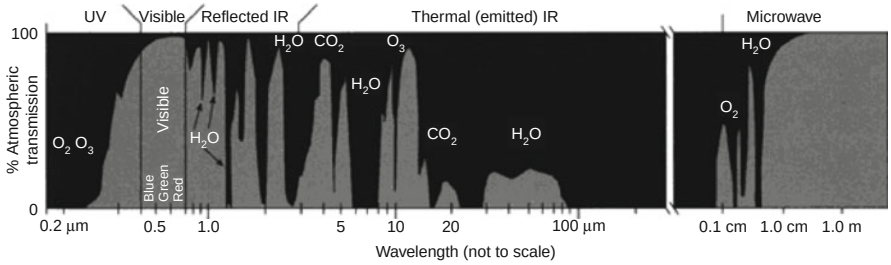


Fig. 34.8 Atmospheric absorption spectra by major atmospheric constituents (Earth observatory, absorption bands and atmospheric windows, http://earthobservatory.nasa.gov/Features/RemoteSensing/remote_04.php) (Courtesy of NASA)

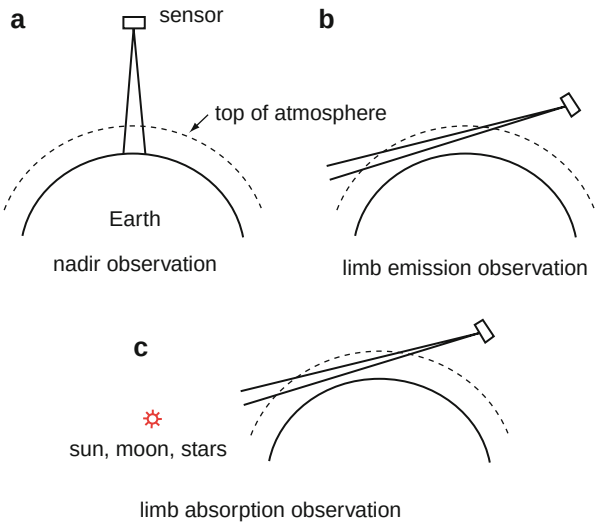


Fig. 34.9 Observation method of atmosphere

depending on the source light. One is to measure the emissions from the atmosphere and another is to observe absorption by the atmosphere from light sources, e.g., sun, moon, and stars. The latter method is called an occultation measurement. Each observation method has advantages and disadvantages. Nadir observation is good for observing horizontal distributions, while its vertical resolution is not so good. Limb observation has very high vertical resolution and is also very sensitive, but its horizontal resolution is bad, and it is very difficult to observe middle and lower troposphere. Sun occultation measurements are very stable and have high signal-to-noise ratio, but the observation areas are very limited when used from sun-synchronous orbit satellites. Limb emission measurements can measure in day and night and can cover very wide areas, but its sensitivity is limited by the instrument sensitivity.

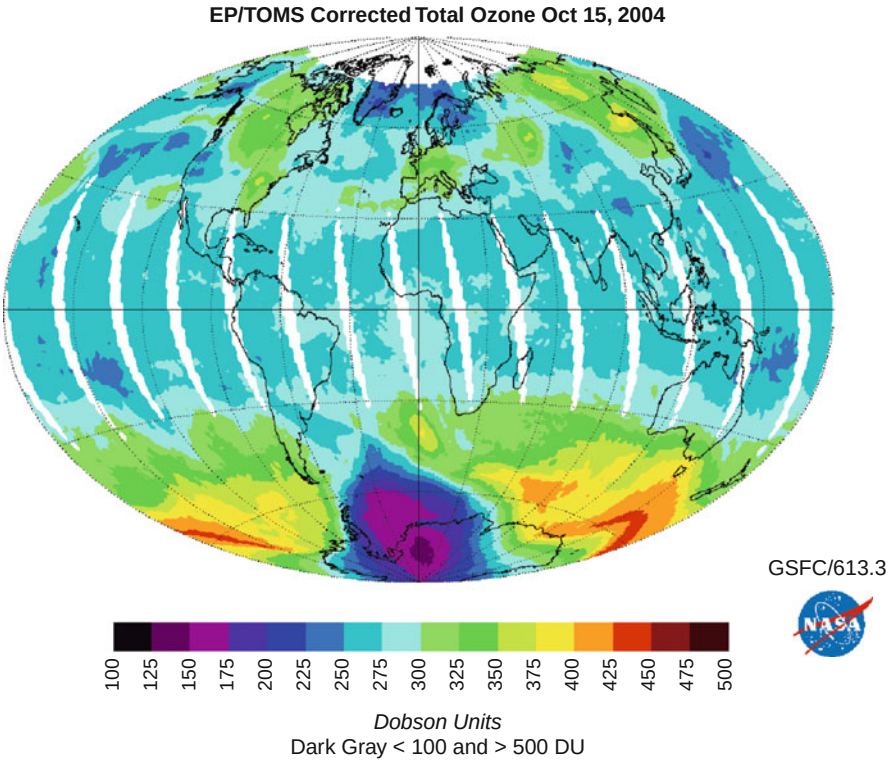


Fig. 34.10 Total ozone measured by TOMS on EP (Earth probe TOMS data & images http://toms.gsfc.nasa.gov/eptoms/ep_v8.html) (Courtesy of NASA)

In the UV region, the main target is Ozone, sulfur dioxide, and NO_2. Figure 34.10 shows the total ozone measured by EP/TOMS (▶ Chap. 1). The antarctic ozone hole can be clearly seen in this image. In the visible and near-infrared region, there are very few steep absorption lines. There are some lines of ozone and detailed absorption lines of oxygen. Oxygen lines are used to retrieve pressures. Figure 34.11 shows the vertical distribution of ozone over Antarctica obtained from sun occultation sensor ILAS (▶ Chap. 1) on ADEOS (▶ Chap. 1).

In the infrared region, there are many absorption lines from many molecules. For the nadir measurement, major molecules which can be measured are H_2O, CO, N_2O, O_3, and CH_4. Figure 34.12 shows the distribution of CO measured by SCIAMACHY (▶ Chap. 1) on ENVISAT (▶ Chap. 1).

Limb emission measurement can measure many kinds of atmospheric molecules. For instance, MIPAS (▶ Chap. 1) on ENVISAT (▶ Chap. 1) can measure following molecules: C_2H_2, C_2H_6, CH_4, ClO, $ClONO_2$, CO, F11, F12, F22, H_2O, H_2CO, HCN, HCOOH, HNO_3, HNO_4, HOCl, N_2O_5, N_2O, NO_2, NO, O_3.

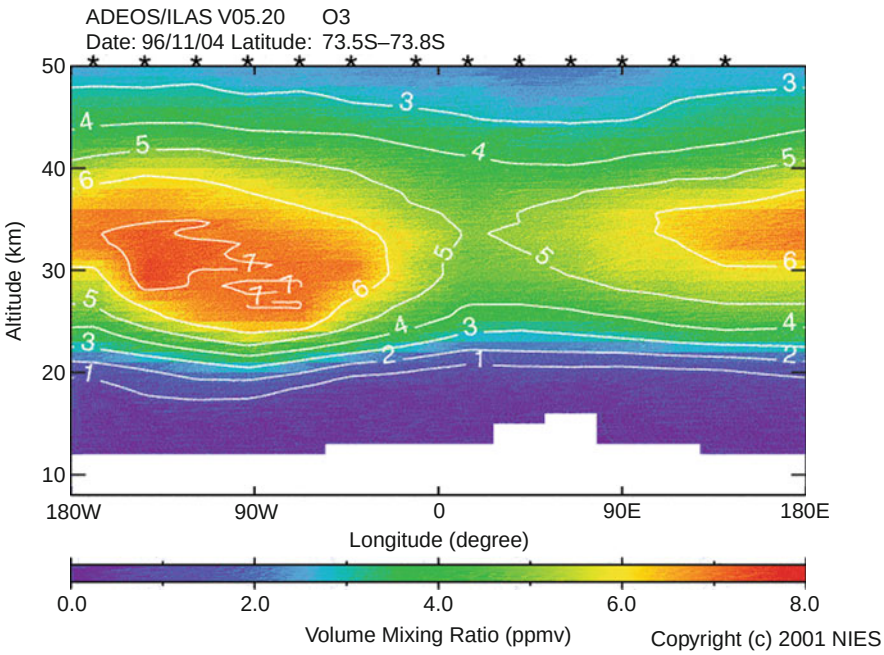


Fig. 34.11 Vertical distribution of ozone over Antarctica obtained from sun occultation sensor ILAS on ADEOS (ADEOS earthview, Three-dimensional distribution of atmospheric ozone, http://suzaku.eorc.jaxa.jp/GLI2/adeos/Earth_View/jap/adeos16j.pdf) (Courtesy of NIES)

Figure 34.13 shows global distributions of HCN (top) and C_2H_6 (bottom) at 200hPa (10.5–12.6km) emitted from biomass burning observed by MIPAS (► Chap. 1) on ENVISAT in September 2003 (► Chap. 1).

White areas are data gaps due to cloud contamination (or insensitive values in case of C_2H_6 south of 60°S). Red solid lines show the monthly averaged tropopause intersection from the NCEP reanalysis. Strongly enhanced values of HCN between South America, Africa, and Australia reflect the southern hemispheric biomass burning plume. Enhanced C_2H_6, an indicator for both biomass burning and industrial/urban pollution, is also visible in this region, but additionally west of Peru and in the northern tropics and subtropics. Trajectory calculations (not shown) hint toward pollution sources in Northern South America. For further details see: http://www.atmos-chem-phys.net/9/9619/2009/acp-9-9619-2009.html

Greenhouse Gases

Greenhouse gases, especially distributions of CO_2 and CH_4 and their fluxes are very important parameters to understand the carbon cycle. Atmospheric greenhouse

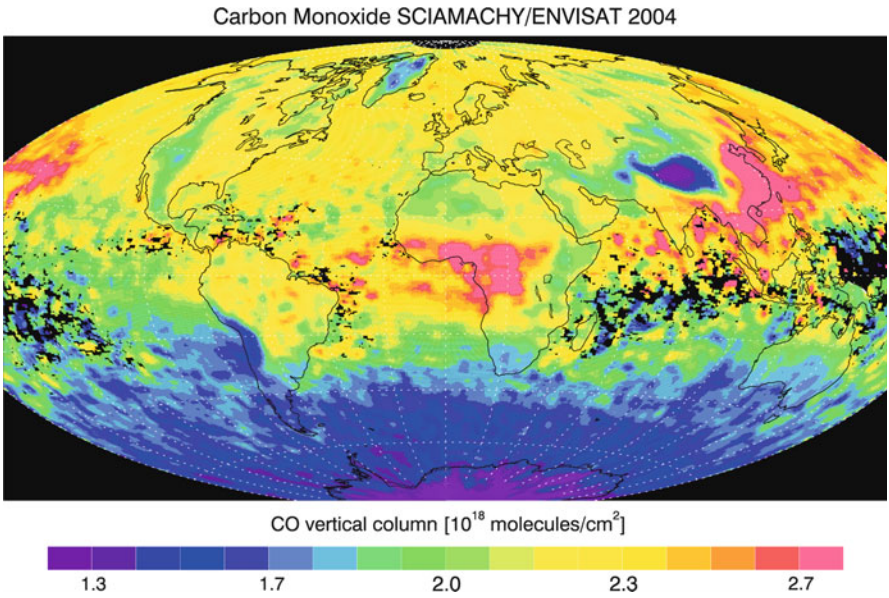


Fig. 34.12 Global CO distribution measured by SCIAMACHY on ENVISAT (Carbon monoxide SCIAMACHY/ENVISAT 2004, http://www.iup.uni-bremen.de/sciamachy/NIR_NADIR_WFM_DOAS/SCIA_CO_glo_2004.png) (Courtesy of Univ. Bremen)

gases, especially carbon dioxide is very difficult to measure. In order to retrieve useful CO_2 concentrations, required accuracy is at least 1%, and 1 ppm accuracy is desirable. On the contrary, CO_2 measurement is very sensitive to aerosols, clouds, and surface pressures. This is the reason why there were no spaceborne sensors to measure CO_2 until recently. There are several CO_2 absorption lines in short wave infrared and thermal infrared regions. In the short wave infrared region, absorption lines are in 1.6 μm and 2.0 μm regions. In the thermal infrared region, CO_2 absorption lines are in 11 μm and 15 μm regions. Figure 34.14 shows CO_2 and CH_4 absorption lines in shortwave and thermal infrared regions. As for the vertical sensitivity, short wave infrared region has almost flat sensitivity in the troposphere, while thermal infrared region has high sensitivity in mid-troposphere, but very low sensitivity in lower troposphere. From short wave infrared spectra, only total column density can be retrieved, while in the thermal infrared region, vertical profile from around 3,000 m and above can be retrieved. Figure 34.15 shows CO_2 distribution retrieved from short wave infrared bands of TANSO-FTS (▶ Chap. 1) on GOSAT (▶ Chap. 1). TANSO-FTS is a Fourier transform spectrometer with short wave infrared bands and thermal infrared bands. GOSAT can also retrieve CH_4.

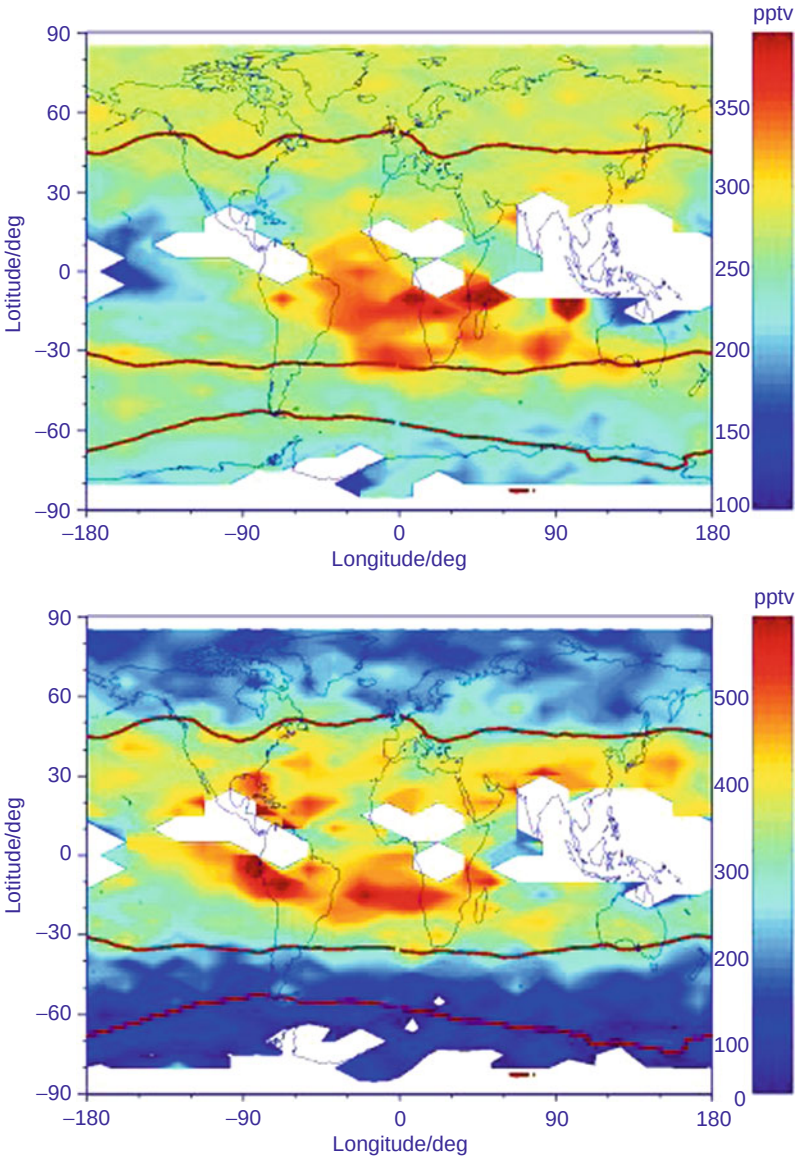


Fig. 34.13 Global distributions of HCN and C_2H_6 showing biomass burning (Global distributions of HCN and C2H6 showing biomass burning, http://www.imk-asf.kit.edu/english/673.php) (Courtesy of IMK)

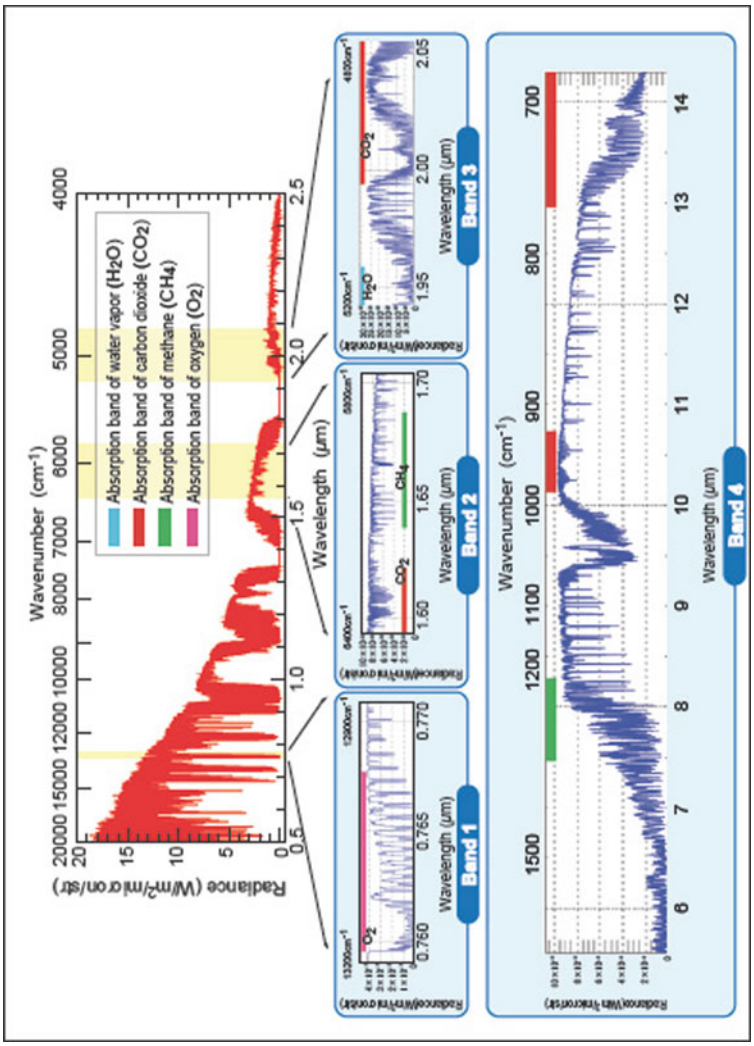


Fig. 34.14 CO_2 and CH_4 absorption lines in shortwave and thermal infrared region (GOSAT project, http://www.gosat.nies.go.jp/eng/gosat/zu8.htm) (Courtesy of NIES)

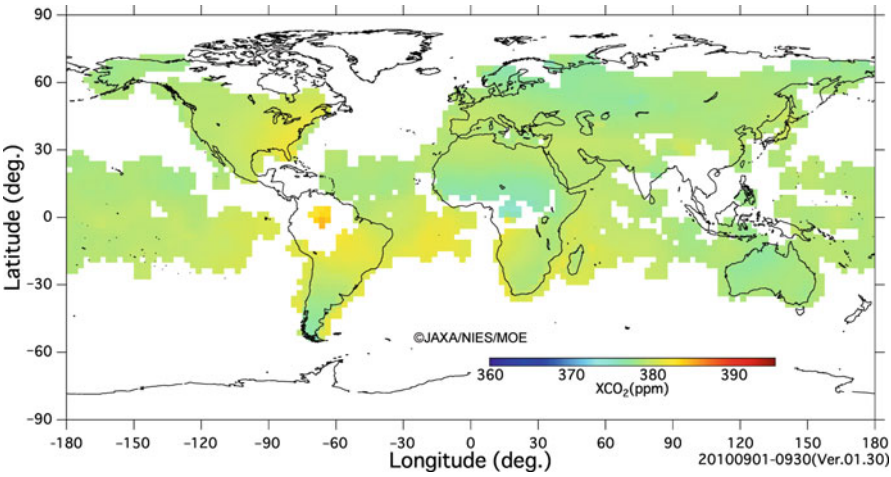


Fig. 34.15 CO_2 distribution from TANSO on GOSAT (XCO2 distribution level3, http://www.gosat.nies.go.jp/jp/related/2010/img/XCO2_L3_20100 9010930_v01_30.png) (Courtesy of NIES)

Oceanic Applications

Sea Surface Temperature

Sea surface temperature (SST) is a very important parameter to understand the surface ocean circulations, as well as to monitor the global warming trend. It is also very useful for efficient fisheries. SST is the first geophysical parameter retrieved from satellite-borne sensors. SST can be measured using thermal infrared spectra, but these spectra are also sensitive to atmospheric water vapor. In order to eliminate the water vapor absorption, two or three thermal infrared bands are used. The most popular algorithm to retrieve SST was developed by NOAA for AVHRR (▶ Chap. 1) on NOAA (▶ Chap. 1) satellite. It is an empirical algorithm and is called multichannel SST (MCSST) algorithm (McClain et al. 1985). The two-band MCSST algorithm is shown as follows.

$$T_S = a_1 T_4 + a_2 (T_4 - T_5) + a_3 (T_4 - T_5)(\sec \theta_z - 1) + a_4$$

Here,

T_S: Sea surface temperature
T_4: Brightness temperature of 11 μm band
T_4: Brightness temperature of 12 μm band
θ_z: Satellite zenith angle at the ground
a_1–a_4: Coefficients which are determined experimentally

This algorithm works well and its accuracy is around 0.5 K. However, it fails when the sea surface is very calm.

One of the problems of SST measured by thermal IR spectra is that it measures sea skin temperature, i.e., less than 1 mm depth temperature, while in situ measurement usually measures 1 m depth temperature. Therefore, the SST retrieved from satellite IR sensors are inherently different from in situ measurements. However, usually, the differences are within the accuracy.

Another way of measuring SST from space is to use microwave region. IR sensors cannot be used over clouded areas, but microwave sensors can measure SST under clouds. In the microwave region, lower frequency is more sensitive to SST, and C band is the most appropriate frequency. Several corrections are necessary to retrieve SST from microwave radiometers. Refer to the AMSR-E case.[1] The problem is that the spatial resolution of low frequency microwave is very low. In the case of AMSR-E (▶ Chap. 1), the spatial resolution of SST is around 50 km compared to 1 km resolution of MODIS (▶ Chap. 1) which uses thermal IR.

Figure 34.16a shows 8-day composite of global SST retrieved from MODIS on Aqua, and Fig. 34.16b shows 8-day composite of global SST retrieved from AMSR-E on Aqua.

Sea Surface Salinity

Sea surface salinity (SSS) is one of the key parameters which controls the deep sea water circulation (Lagerloef 2002; Lagerloef and Schmitt 2006; US CLIVAR 2008). However, only low frequency microwave, i.e., L-band microwave, is sensitive to SSS. Also it depends on sea surface winds and column water vapor. Aquarius (▶ Chap. 1) satellite which can simultaneously measure SSS, sea surface winds, and water vapor was launched this year.

Sea Surface Wind

Sea surface winds are very important parameters which decide surface ocean circulations. It is also very important to find and track cyclones. Sea surface winds can be retrieved from microwave scatterometers, microwave altimeters, and microwave radiometers. Microwave scatterometers measure Bragg scatterings from capillary waves on the sea surface. This scattering depends on surface wind speeds, directions, and incidence angles. The relationship between these parameters and back scattering coefficient can be described by geophysical model functions, which is determined by empirical data (Jones et al. 1981; Naderi and Freilich 1991). Figure 34.17 shows a sample of geophysical model function. As can be seen from this figure, there are four ambiguities for the surface wind directions. In order to eliminate this ambiguity, back scatterings are measured from three directions. Still, there remain 180° ambiguities and these ambiguities are removed by post-processings. Figure 34.18 shows an

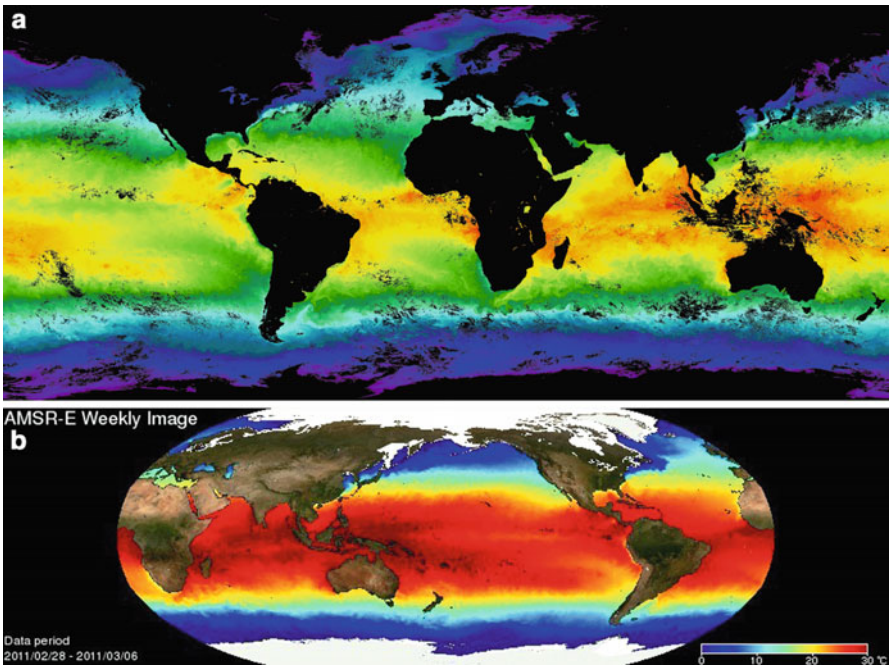


Fig. 34.16 Sea surface temperature retrieved from thermal infrared (**a**) (SST retrieved from MODIS on Aqua (http://oceancolor.gsfc.nasa.gov/cgi/l3/A20110572011064.L3m_8D_SST_9.png?sub=img); Courtesy of NASA) and C-band microwave radiometer (**b**) (SST retrieved from AMSR-E on Aqua (AMSR-E weekly image, http://sharaku.eorc.jaxa.jp/AMSR/topimage/SST.jpg); Courtesy of JAXA)

example of sea surface wind vectors measured by QuikScat (► Chap. 1). Until now, two frequencies are used for microwave scatterometer, i.e., Ku band and C band. Ku band scatterometer is more accurate and can measure low wind speed, while C-band scatterometer can measure higher wind speed and is more insensitive to rainfall.

Measurement of sea surface winds using microwave altimeter depends on another mechanism. It measures quasi-specular scatterings from the sea surface, and these scatterings are inversely proportional to the wind speed. Microwave altimeter cannot measure wind directions.

Microwave radiometer also can measure sea surface wind speed. Most of microwave radiometers can measure only wind speed and not wind directions. Wind speed modifies the emissivity of the sea surface, and these modifications depend on frequencies, and wind speed also modifies the polarization. Depending upon these modifications, sea surface wind speed can be retrieved from microwave radiometer data. However, wind direction also slightly modifies brightness temperature and

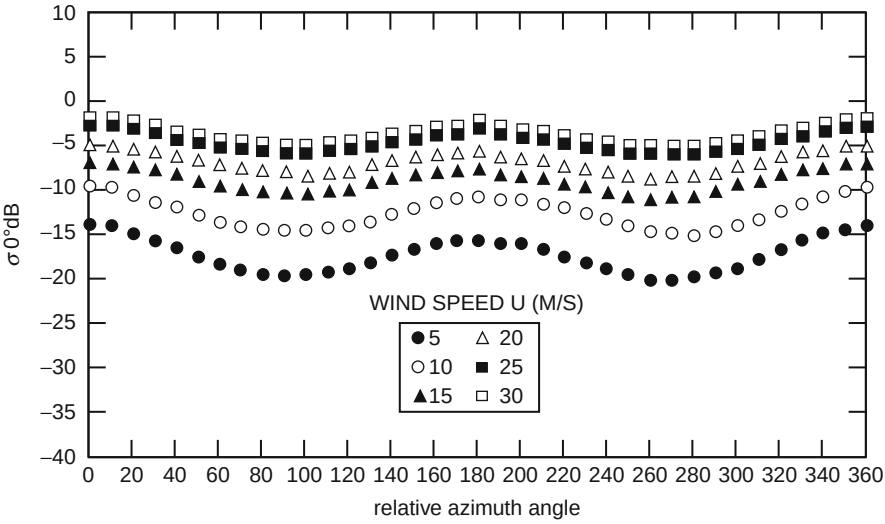


Fig. 34.17 A sample of geophysical model function (Japan Association of Remote Sensing (eds.), *Remote Sensing Graphical Explanation* (in Japanese), Fig. 4.11.3, p. 91 (Japan Association of Surveyers, 2001))

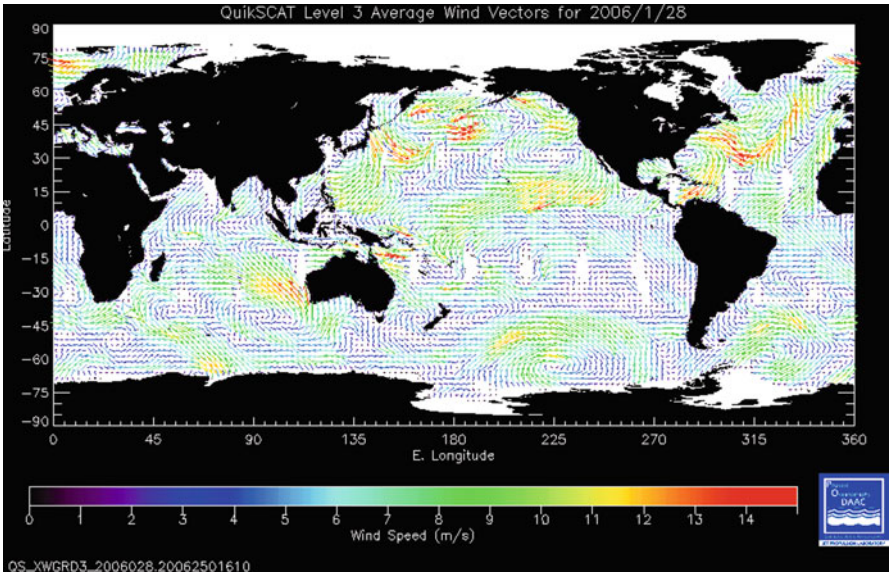


Fig. 34.18 A sample of sea surface wind vector measured by SeaWinds on QuikScat (http://podaac.jpl.nasa.gov/cgi-bin/quikscat/l3toRev.pl?image=ftp://podaac.jpl.nasa.gov/pub/ocean_wind/quikscat/L3/browse/2006/QS_XWGRD3_2006028.gif) (Courtesy of NASA/JPL)

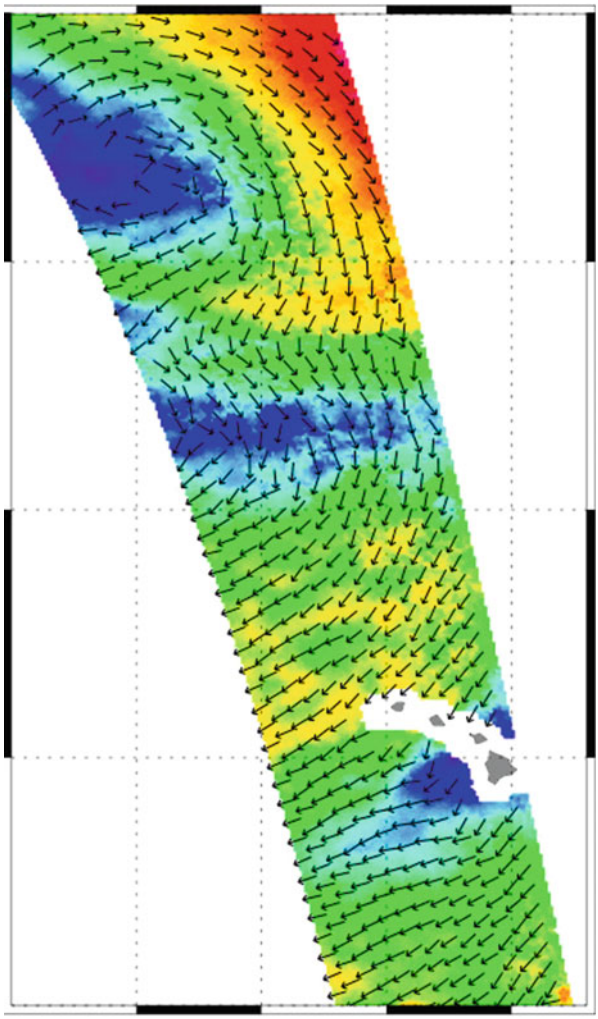


Fig. 34.19 A sample of sea surface wind vector measured by Windsat (Windsat images, http://www.nrl.navy.mil/WindSat/images/wndmap.jpg) (Courtesy of NRL)

polarizations. With the Stokes vector measurement (i.e., four directions of polarization measurement), wind direction also can be retrieved (Meissner and Wentz 2002; Jelenak et al. 2004). Windsat (► Chap. 1) has this capability and measures sea surface wind vector. However, wind vector measurement using microwave radiometer can be done over rather small range of wind speed (cannot measure less than 6 m/s), and also it is largely affected by rainfall. Figure 34.19 shows an example of sea surface wind vector retrieved from Windsat.

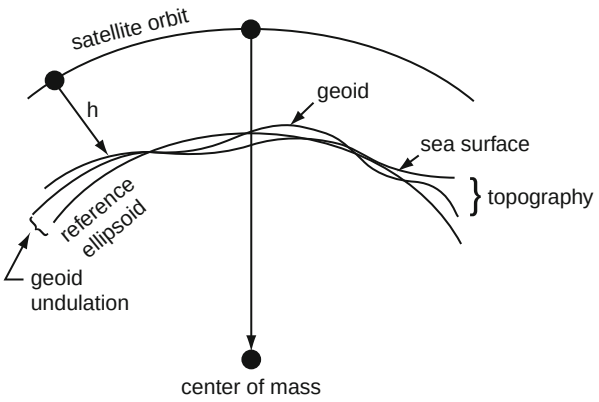


Fig. 34.20 The principle of satellite altimetry (Japan Association of Remote Sensing (eds.), Remote sensing note, Fig. 4.10.1, http://www.dgeo.udec.cl/yuras/tutoriales/rsnote/cp4/4-10-1.gif)

Sea Surface Height

Sea surface height is important to understand the ocean dynamics of sea current, geoid surveys, etc. Sea surface height (SSH) is usually measured by microwave altimeter. The principle of satellite altimetry is shown in Fig. 34.20. The ground height or the sea surface height is measured from the reference ellipsoid. If the altitude of the satellite, *Hs*, is given as the height from the reference ellipsoid, the sea surface height *HSSH* is calculated as follows.

$$HSSH = Hs - Ha$$

where *Ha*: measured distance between satellite and the sea surface

Ha is measured on the basis of the travel time of the transmitted microwave pulses. From the time ($t = 0$), when the first edge of pulse arrives at the surface, to the time ($t = t$) when the end edge of a pulse with a width of τ arrives at the surface, the received power increases linearly as shown in Fig. 34.21. The received pulses are composed of echoes from various parts of the sea surface. Therefore, the travel time from a satellite to the sea surface can be calculated by averaging the received pulses. Pulse compression techniques will also be applied in order to obtain a high-frequency pulse for the improvement of the resolution.

The accuracy of altimeter itself is around 1 cm, but it measures only distance between the satellite and the sea surface, hence it is most important to know the satellite position. In order to measure the satellite position, several means have been developed. Now, the mostly used instrument is GPS. Another way which can measure satellite position in high accuracy is to use ground-based lasers and corner mirrors onboard the satellite. Both means have similar accuracy, i.e., 2–3 cm. Figure 34.22 shows an example of sea surface height measured by Jason1. Anyway, microwave altimeter data are affected by the water vapor. So, microwave radiometer which can measure total water vapor is usually installed with microwave altimeter for water vapor correction.

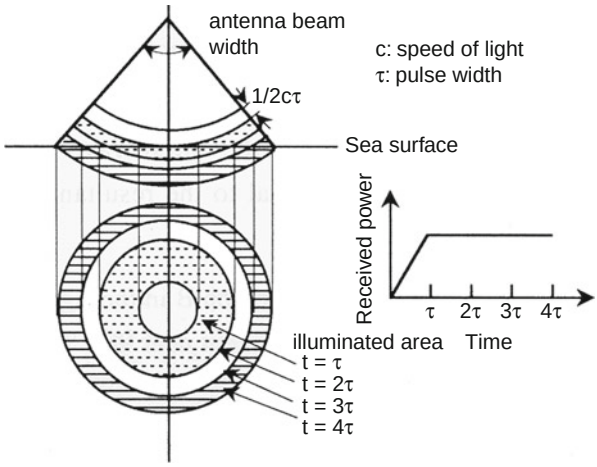


Fig. 34.21 Change of illuminated area and received power (Japan Association of Remote Sensing (eds.), Remote sensing note, Fig. 4.10.2, http://www.dgeo.udec.cl/yuras/tutoriales/rsnote/cp4/4-10-2.gif)

Another way of measuring sea surface height is to use lidar. Icesat measured sea surface height using a lidar. Lidar has much more smaller IFOV compared to microwave altimeter, but as ocean is very dark, signal-to-noise ratio of lidar is not so high.

Ocean Color

Ocean color is one of the key components to characterize the oceanic living matters. Usually, main geophysical parameters retrieved from ocean color sensors are chlorophyll-a concentrations. Chlorophyll-a concentrations are assumed to be proportional to phyto-plankton concentrations. Phyto-planktons are the bottom of oceanic food chain and also play an important role in carbon cycles in ocean. Until now, most of the chlorophyll-a retrieval algorithms use empirical algorithms after atmospheric corrections. The reflected light from open ocean is mostly composed of scattered light from the atmosphere. Only 10% of received light at the satellite comes from the ocean surface. So, the atmospheric corrections are most important in chlorophyll retrievals. Conventional algorithms use near-infrared channels for aerosol estimation. As ocean can be thought to be dark in these channels, near-infrared radiances are thought to be reflected from atmospheric aerosols (Gordon and Clark 1981; Gordon and Wang 1994; Siegel et al. 2000).

Here, one of such algorithms is presented according to Gordon (Gordon and Voss 2004). The radiance received by a sensor at the top of the atmosphere (TOA) at a wavelength λ_i, $L_t(\lambda_i)$, can be written as follows.

$$L_t(\lambda_i) = L_{path}(\lambda_i) + T(\lambda_i)L_g(\lambda_i) + t(\lambda_i)L_{wc}(\lambda_i) + t(\lambda_i)L_w(\lambda_i)$$

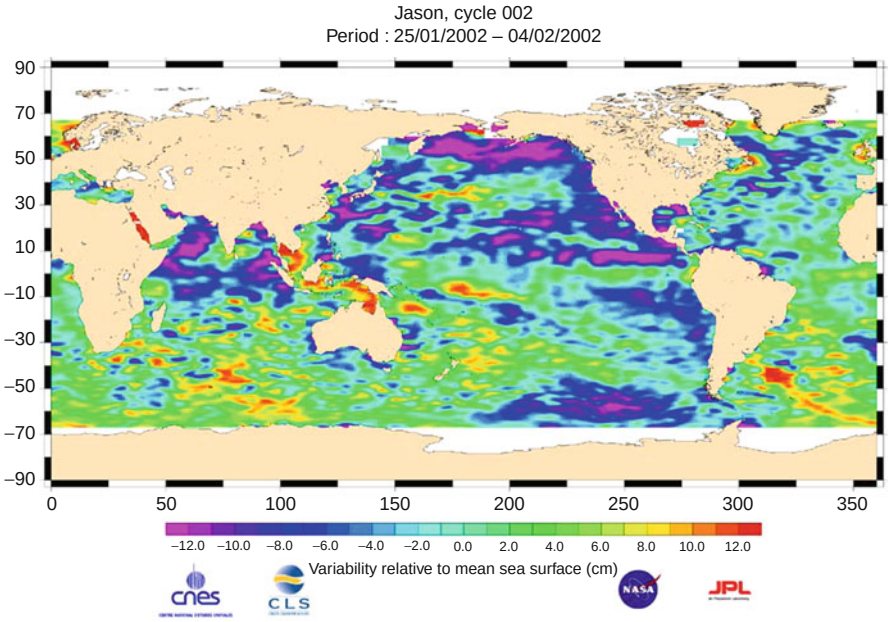


Fig. 34.22 Global sea surface height measured by Jason1 (Jason, cycle002, Period:25/01/2002–04/02/2002 http://www.jpl.nasa.gov/images/earth/jason_mer.jpg) (Courtesy of NASA and CNES)

Here

$L_{path}(\lambda_i)$: The radiance generated along the optical path by scattering in the atmosphere and by specular reflection of atmospherically scattered light (skylight) from the sea surface

$L_g(\lambda_i)$: The contribution arising from specular reflection of direct sunlight from the sea surface (sun glitter)

$L_{wc}(\lambda_i)$: The contribution arising from sunlight and skylight reflecting from individual white caps on the sea surface

$L_w(\lambda_i)$: The desired water leaving radiance

T: The direct transmittance of the atmosphere

t: The diffuse transmittance of the atmosphere

The above equation can be converted to the reflectance as follows.

$$\rho_t(\lambda_i) = \rho_{path}(\lambda_i) + T(\lambda_i)\rho_g(\lambda_i) + t(\lambda_i)\rho_{wc}(\lambda_i) + t(\lambda_i)\rho_w(\lambda_i)$$

In the first approximation, the sun glitter term and the white cap term can be ignored. So, the most important part is to estimate $\rho_{path}(\lambda_i)$. It can be written as follows.

$$\rho_{path} = \rho_r(\lambda) + \rho_a(\lambda) + \rho_{ra}(\lambda)$$

Here, ρ_r is the reflectance resulting from multiple scattering by air molecules (Rayleigh scattering) in the absence of aerosols, ρ_a is the reflectance resulting from multiple scattering by aerosols in the absence of the air, and ρ_{ra} is the interaction term between molecular and aerosol scattering (Antoine and Morel 1998; Deschamps et al. 1983). In the single scattering approximation (called CZCS algorithm), ρ_{ra} can be neglected. As ρ_r can be calculated rather accurately, the problem is how to estimate ρ_a.

The next approximation is that ocean totally absorbs near-infrared light, i.e., there is no reflected light from the ocean in near-infrared channels. Thus, the ρ_a in near-infrared channels can be computed by subtracting ρ_r from ρ_t. Using two near-infrared channels, i.e., λ_s (shorter wavelength) and λ_l (longer wavelength), the ratio can be calculated as follows.

$$\varepsilon(\lambda_s, \lambda_l) = \frac{\rho_a(\lambda_s)}{\rho_a(\lambda_l)}$$

For other ith channels, $\varepsilon(\lambda_i, \lambda_l)$ should be known. For this purpose, radiative transfer calculations are done for many kinds of aerosols. The proper aerosol model is selected from the $\varepsilon(\lambda_s, \lambda_l)$ value. For real MODIS ocean color retrievals, more sophisticated algorithms including multiple scatterings are used, but the principle is the same.

These algorithms have rather higher accuracy in open ocean (called case 1 water), but not so good accuracy in turbid coastal waters (called case 2 water). Analytical algorithms are now developed for case 2 waters (Chomkoa et al. 2003; Park and Ruddick 2005; Goa et al. 2007).

Chlorophyll-a data are used to estimate the primary production in the ocean. The most popular algorithm to estimate ocean primary production was proposed by Behrenfeld and Falkowski (Berenfeld and Falkowski 1997). This algorithm can be described as follows.

$$IPP = 0.66125 \times P^B_{opt} \times \frac{E_0}{E_0 + 4.1} \times DL \times Z_{eu} \times Chl.a$$

Here

IPP: Integrated primary production of 1 day ($mgCm^{-2}\ day^{-1}$)

P^B_{opt}: Maximum carbon fixation quantity within euphotic zone per unit chlorophyll-a(mgC/mgChl/hour)

E_0: Photosynthetic active radiation (PAR) (mol quanta m^{-2})

DL: Day length (hour)

Z_{eu}: Depth of euphotic zone (m)

Chl.a: Chlorophyll-a concentration at the sea surface

In this algorithm, P^B_{opt} is expressed as seventh order polynomials of sea surface temperature. From satellite data, SST and PAR can be retrieved. Figure 34.23 shows SST, PAR, Chlorophyll-a concentration, and primary production estimated

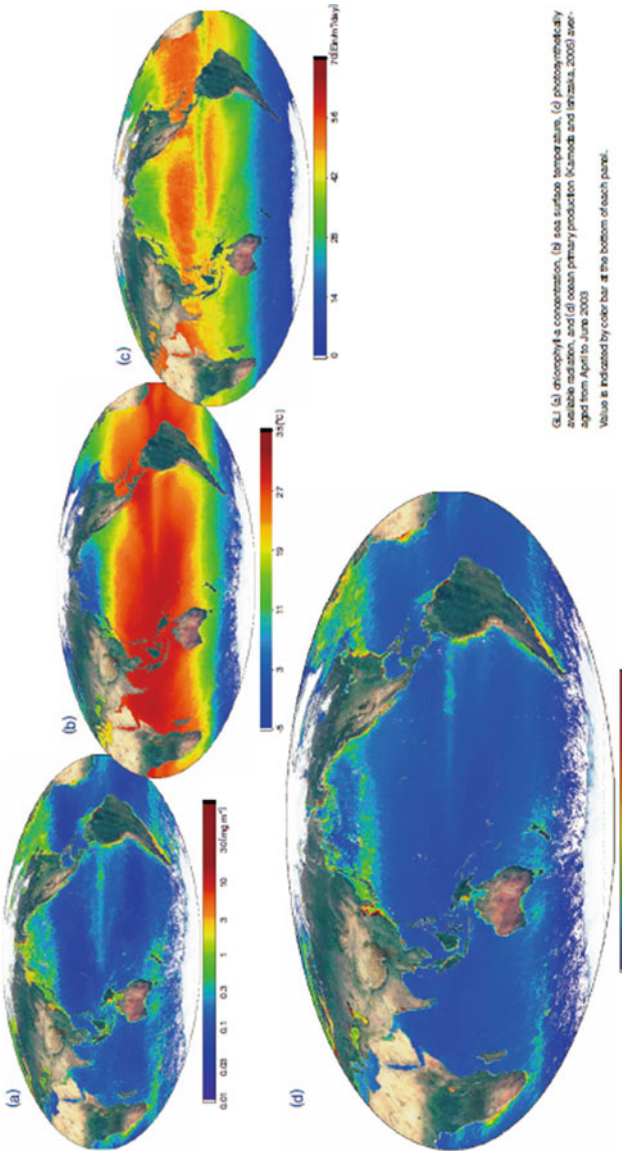


Fig. 34.23 SST, PAR, Chlorophyll-a, and primary production estimated from GLI on ADEOS2 (http://suzaku.eorc.jaxa.jp/GLI/doc/GLI_BOOK_CD/PDF/CHAP_6.PDF) (Courtesy of JAXA)

from GLI on ADEOS2. Many algorithms which modify this algorithm are proposed (Behrenfeld et al. 2005; Carr et al. 2006; Asanuma 2006; Ishizaka et al. 2007; Westberry et al. 2008).

Land Applications

Topography

Topography is one of the most important parameters of land and used in most of the maps. Before the satellite era, topography or altitudes of land were mainly measured by stereo aerial photographs. However, it is rather difficult to measure global topography by aerial photographs. Satellite-borne sensors have provided the means to generate global topographic data. There are several kinds of sensors which can be used to measure land topography. One is the stereo images which are very similar with aerial photographs. The main part of retrieving DSM from stereo pairs is image matching algorithms. Many kinds of image matching algorithms are proposed, but the least squares matching algorithm is the most powerful (Gruen 1985). There still remain problems, especially, occlusion problems and mismatching caused by very homogeneous targets. Second is the synthetic aperture radar (SAR) (Zebker and Goldstein 1986; Goldstein et al. 1988; Madsen and Zebker 1998). The third is to use lidar from satellites.

The stereo imaging from satellites has started from SPOT (▶ Chap. 1) in 1986. However, SPOT made stereo pairs from different orbits. So the stereo pair images were taken on different days. It is rather difficult to collect the stereo pair images with the same conditions. After SPOT, several satellites were installed with multiple sensors, which made possible to get stereo images almost simultaneously. These multiple sensors were OPS on JERS-1 (▶ Chap. 1), ASTER (▶ Chap. 1) on Terra (▶ Chap. 1) and PRISM (▶ Chap. 1) on ALOS (▶ Chap. 1). Only from ASTER, global digital surface model (DSM) was generated and distributed (ASTER 2002). It has 30 m resolution and covers almost all over the world. Figure 34.24 shows an example of ASTER DEM. Higher resolution DSMs can be generated from high-resolution sensors like PRISM and many commercial high-resolution sensors, but the DSM generations are limited and cover very few areas.

The principle of DSM generation from stereo pairs can be described as follows. Figure 34.25 shows the configuration of stereo pair images taken from parallel directions.

Here, object coordinate space is expressed by (x, y, z) and image coordinates are expressed by (u, v). The origin of image coordinates is the crossing point of image projection plane and the optical axis, and u and v axes are parallel with x and y axes, respectively. The distance between two cameras, L and R, is c, and the projection center of each camera, O_L and O_R, is $(-b/2, 0, 0)$ and $(b/2, 0, 0)$ both on the x axis. The optical axes of cameras L and R are parallel with z axis.

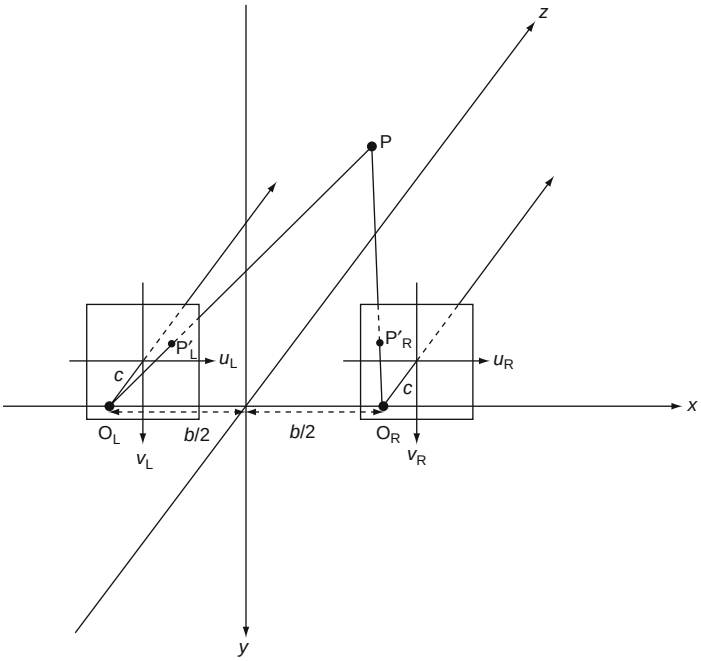


Fig. 34.24 Configuration of stereo pair images taken from parallel directions (Takagi and Shimoda 2004)

The next equation stands for the image $P'_L(u_L, v_L)$ of target P (x, y, z) in image L.

$$\frac{u_L}{c} = \frac{x + b/2}{z} \tag{34.1a}$$

$$\frac{v_L}{c} = \frac{y}{z} \tag{34.1b}$$

Similarly, the next equation stands for the image $P'_R(u_R, v_R)$ of target P (x, y, z) in image R.

$$\frac{u_R}{c} = \frac{x - b/2}{z} \tag{34.2a}$$

$$\frac{v_R}{c} = \frac{y}{z} \tag{34.2b}$$

From Eqs. 34.1a, 34.1b, 34.2a, and 34.2b, the next equation can be deduced.

$$x = \frac{u_L z}{c} - \frac{b}{2} = \frac{u_R z}{c} + \frac{b}{2} \tag{34.3a}$$

$$y = \frac{v_L z}{c} = \frac{v_R z}{c} \tag{34.3b}$$

From the above equations, z can be calculated as follows.

$$z = c\frac{b}{u_L - u_R} \tag{34.4}$$

x and y can be obtained by substituting z in Eqs. 34.3a and 34.3b.

Another way to measure topography is to use interferometric SAR. The most important project to generate global DSM using interferometric SAR was Shuttle Radar Topography Mission (SRTM).[2] SRTM was conducted in 2000 in STS99 mission. Two antennas with distance of 60 m were used, and C and X band SAR were used. It covered $\pm 60^\circ$ latitudinal areas and the original resolution is 30 m. Global 90 m resolution data are distributed to the public. Figure 34.26 shows the global SRTM DSM.

Lidar also can be used to generate land topography. The lidar on Icesat (▶ Chap. 1) was used to retrieve land topography,[3] but its main mission was to measure the height of ice sheet and the changes of ice sheets. There are land topography products, but as the lidar can measure only the nadir of the satellite, the coverage is rather sparse.

Geometric Corrections and Map Projection

Images obtained from satellite include many kinds of geometric distortions. In the remote sensing applications, especially in the land applications, it is very important to correct these distortions. These distortions include inner distortions (lens distortion, distortions within the sensor, etc.) and outer distortions (Earth curvature, spacecraft position and attitude error, etc.).

There are two kinds of correction algorithms. One is to correct these distortions systematically based on known parameters. Another algorithm is to correct the image distortions and map to existing maps using ground control points (GCP). There are also some combinations of the above algorithms. Until sometime ago, the parameters obtained from the satellite were not accurate enough to achieve accurate corrections. Especially, the accuracy of satellite position and attitude were not enough. However, recent satellites have GPS receivers and star trackers, providing almost sufficient accurate position and attitudes data.

GCPs are corresponding points in images and in corresponding maps. It is a rather tedious operation to select accurate GCPs. However, in the GCP correction algorithm, this process is of course very important.

In the case of systematic corrections, a map projection is finally required. There are many kinds of map projections, but usually several projections are used, i.e., universal transverse Mercator (UTM), Mercator, polar stereo, and

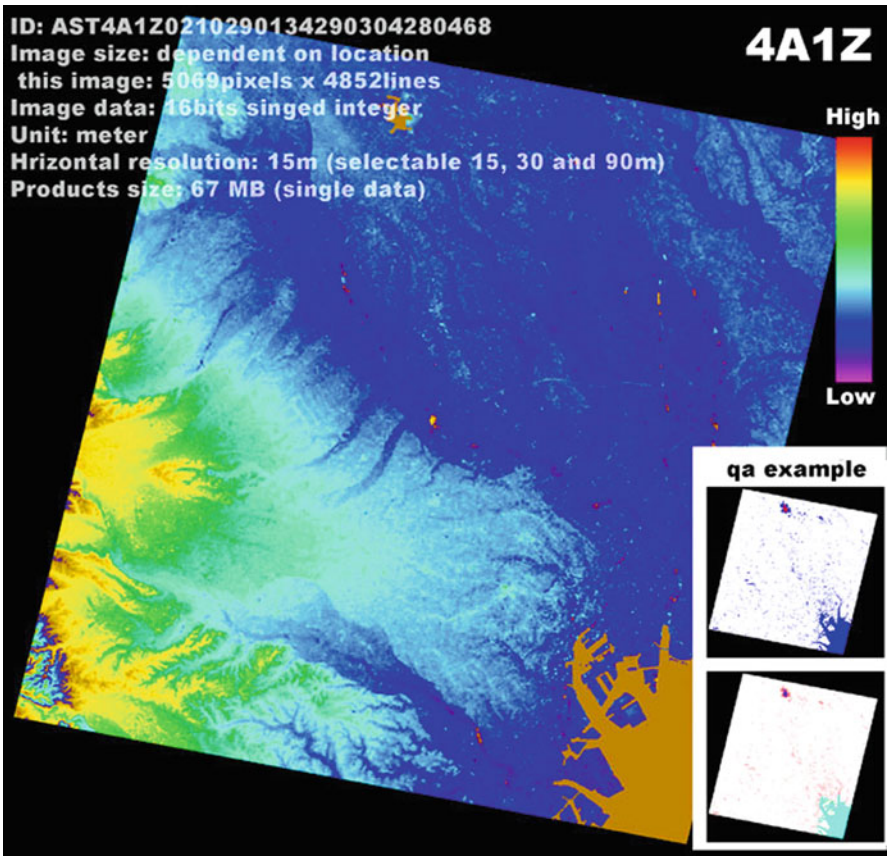


Fig. 34.25 An example of ASTER DEM (ASTER GDS, http://www.gds.aster.ersdac.or.jp/gds_www2002/exhibition_e/a_products_e/4A01Z.jpg) (Courtesy of ERSDAC)

latitude/longitude grid. UTM is usually used for large-scale maps. Mercator is mainly used for oceanic applications, and polar stereo is mainly used for meteorological applications. Latitude/longitude grid is also frequently used because it can be used as starting data for any kinds of projections.

Before map projections, orthographic projection is also sometimes important when the target area is not flat. When there are mountains in the image, these mountains are projected at a slant. Orthographic projection eliminates these distortions, but need DEM of the target areas.

Radiometric Corrections

Radiometric corrections are important when geophysical parameters should be retrieved from satellite data. Radiometric corrections start from converting sensor



Fig. 34.26 Global SRTM DSM (http://photojournal.jpl.nasa.gov/jpegMod/PIA03395_modest.jpg) (Courtesy of NASA)

output values to radiance. This procedure is based upon calibrations. Three kinds of calibrations are usually conducted. The first one is based on the ground calibrations before launch. Ground-based calibrations for optical sensors are done using standard light sources like integrating spheres with standard light sources. However, sensitivities of sensors usually change after launch. After launch, two kinds of calibrations are conducted. One is to use onboard calibration sources. For visible and near-infrared sensors, onboard calibration sources are diffused sunlight, light sources like light bulbs or LEDs, and moon and dark space. For thermal infrared sensors and microwave radiometers, onboard calibration sources are deep space and onboard black body. The third calibration method is vicarious calibrations. In vicarious calibrations, homogeneous ground target and atmospheric measurements are used. The accuracies of calibrations are as follows. For ground-based visible and near-infrared regions, accuracies are 2–3%. The highest accuracy in these regions onboard is to use the Moon, and its accuracy is also around 2–3%. The accuracy of vicarious calibrations is around 5–6%. For thermal infrared, the accuracies depend on the emissivity of the onboard black body, and the highest accuracy is around 0.1 K. For active microwave sensors, different approaches are used. They use ground-based corner reflectors or active radar calibrators, which receive the radar pulse and then send it back to the satellite sensor.

Second step radiometric correction is atmospheric correction. In the visible and near-infrared region, corrections are done to atmospheric scattering by atmospheric molecules and aerosols. Scattering by atmospheric molecules is Rayleigh scattering (Jackson 1962; Craig and Thirunamachandran 1908) and it is rather easy, because the compositions of atmospheric molecules are fairly steady. Aerosols cause Mie scattering (Mie 1908), but its correction is very difficult. It is because aerosol concentration and other properties (radius, species, etc.) are largely dependent on time and place. Many kinds of aerosol correction methods are proposed (Chavez 1988; Gordon and Clark 1981; Kaufman 1989; Vermote et al. 1997), but still their accuracies are not so high. Figure 34.27 shows MODIS (► Chap. 1) images before and after the aerosol correction.

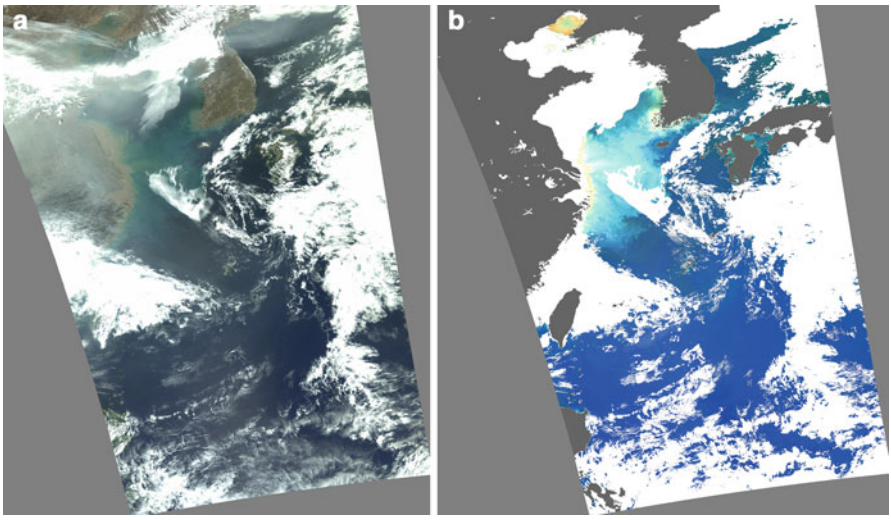


Fig. 34.27 MODIS images before and after the aerosol correction near west Japan and Korea (MODIS near realtime data, http://kuroshio.eorc.jaxa.jp/ADEOS/mod_nrt/) (**a**) before correction (**b**) after correction

Land Cover and Land Use

Land Cover Categories

Land cover and land use are the basic information of land. These data were first used for urban planning, but now they are the fundamental data of land to understand the Earth environment. Land cover and land use are different concepts. Land cover is just what is there. On the contrary, land use is a functional concept. Sometimes they are the same (e.g., forest, agricultural land, lake, etc.). However, for instance, commercial areas and industrial areas are land use categories. From satellite-borne sensors, only land cover can be observed, and land use may be estimated using land cover data. Land cover maps are used, e.g., for urban plannings, vegetation cover estimates, or carbon cycles estimations.

In order to generate land cover maps, usually image classifications are used. Before starting the image classifications, geometric and radiometric corrections are necessary. Geometric corrections are usually necessary, but radiometric corrections are dependent upon the applications. For instance, when the target area is not so large, land cover classifications can be done without radiometric corrections. However, when the target areas are large, e.g., global land cover, radiometric corrections are inevitable. There are effects of sun angles, aerosols, etc., which obscure the accurate classifications. MODIS is distributing land surface reflectance product for these purposes.[4] Also, there is a product called Nadir BRDF adjusted reflectance (NBAR) product which corrects the bidirectional reflectance

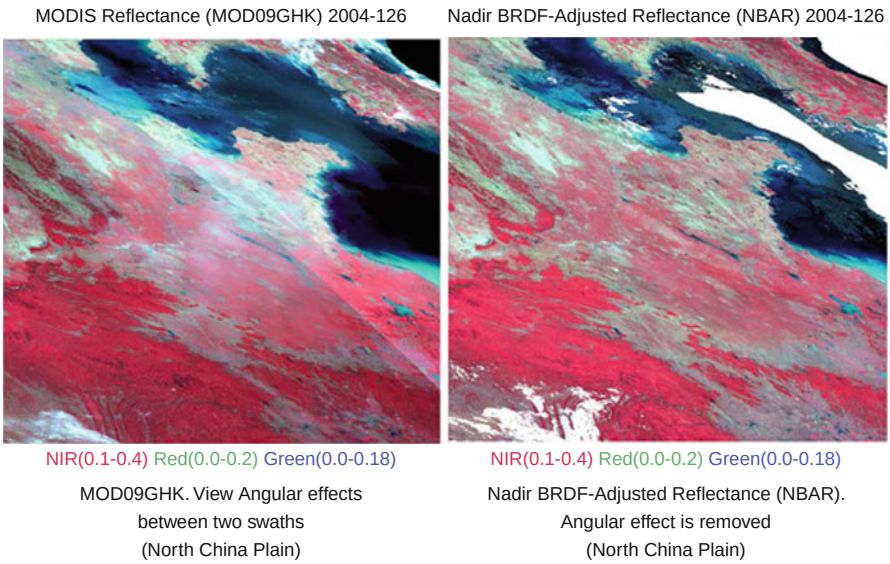


Fig. 34.28 The difference between MODIS surface reflectance product and NBAR product (Schaaf et al. 2007); *left*: MODIS reflectance, *right*: NBAR

distribution function (BRDF) effect.[5] Figure 34.28 shows the difference between MODIS surface reflectance product and NBAR product.

Land cover categories should be defined before classifications. There are several standards on the land cover categories. Tables 34.1, 34.2 and 34.3 show samples of these standard land cover categories for global land cover mapping. These standard classification categories are sometimes useful, but for local land cover mapping, much more specified categories are usually necessary.

Classification Features

Many kinds of features are used for land cover classifications. The most appropriate features depend on the resolutions of images. For high-resolution sensors, i.e., less than 3 m IFOV, spatial features are used in addition to spectral features. Image segmentations (sometimes called an object oriented classifications or object based classifications) are sometimes very useful for these kinds of images (Neubert 2001; Blaschke and Hay 2001; Hofmann 2001). For medium-scale images, i.e., 30–80 m IFOV, spectral features are used in most cases.

For global or regional land cover classifications, other features are used frequently. One of the most popular feature is the time series of normalized difference vegetation indices (NDVI). NDVI is expressed by the following equation.

$$NDVI = \frac{NIR - R}{NIR + R}$$

Here
NDVI: NDVI value
NIR: Value of near-infrared band
R: Value of red band

As NDVI is a kind of ratio between spectral bands, it has some ability to eliminate the radiometric distortions.

Classifiers

Maximum Likelihood Classifier

Many kinds of classifiers are used for land cover classifications. Here, maximum likelihood classifier (MLC), neural net, and support vector machine (SVM), which gives the highest classification accuracy, will be briefly described. The most popular classifier is stochastic classifier. MLC is the most popular within stochastic classifiers. MLC can be described as follows.

The classifier which gives the minimum loss classifies vector $\boldsymbol{x}$ to category C_i which gives the minimum $\sum \lambda_{ij} P(C_i|\boldsymbol{x})$, where $\boldsymbol{x}$ is the classification target data vector, $C = \{C_1, C_2, \cdots, C_n\}$ are categories, and λ_{ij} is the loss when category C_i is misclassified to category C_j. This is called Bayesian decision rule. In the case of image classifications, the losses by misclassifications can be thought to be constant, so the target of the classification is to acquire $P(C_i|\boldsymbol{x})$. From Bayes' theorem, $P(C_i|\boldsymbol{x})$can be expressed as follows.

$$P(C_i|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|C_i)P(C_i)}{\sum p(\boldsymbol{x}|C_i)}$$

Here, probability density function $P(\boldsymbol{x}|C_i)$ is called a likelihood, and $P(C_i)$ is called an a priori probability of category C_i. As the denominator of the above equation is the same to each category, $p(\boldsymbol{x}|C_i)P(C_i)$ should be set up. As $P(C_i)$ is usually difficult to estimate, the vector $\boldsymbol{x}$ is usually classified to the category with the maximum $P(\boldsymbol{x}|C_i)$. This classifier is called a maximum likelihood method. When $p(\boldsymbol{x}|C_i)$ follows a normal distribution, $p(\boldsymbol{x}|C_i)$ is expressed as follows.

$$p(\boldsymbol{x}|C_i) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{V}_i|^{1/2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}}_i)^t \boldsymbol{V}_i^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}_i) \right\}$$

Here
$\boldsymbol{V}_i$: Variance–covariance matrix of category C_i
$\bar{\boldsymbol{x}}_i$: Mean vector of $\boldsymbol{x}$ of category C_i

For the calculation simplicity, the logarithm of the above equation with inverted sign is used, and the x is classified to the category with maximum of the following equation.

$$\log|\boldsymbol{V}_i| + (\boldsymbol{x} - \bar{\boldsymbol{x}}_i)^t \boldsymbol{V}_i^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}_i)$$

The second term of the above equation is called a Mahalanobis distance.

Table 34.1 Land cover categories of IGBP (FAO, FRA 2000 Forest cover mapping and monitoring with NOAA-AVHRR and other coarse spatial resolution sensors, Forest resources assessment programme working paper 29, Rome (2000))

Class	Class name	Description
1	Evergreen needle leaf forests	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 m. Almost all trees remain green all year. Canopy is never without green foliage
2	Evergreen broadleaf forests	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 m. Almost all trees remain green all year. Canopy is never without green foliage
3	Deciduous needle leaf forests	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 m. Consists of seasonal needleleaf tree communities with an annual cycle of leaf-on and leaf-off periods
4	Deciduous broadleaf forests	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 m. Consists of seasonal broadleaf tree communities with an annual cycle of leaf-on and leaf-off periods
5	Mixed forests	Lands dominated by trees with a percent canopy cover >60% and height exceeding 2 m. Consists of tree communities with interspersed mixtures or mosaics of the other four forest cover types. None of the forest types exceeds 60% of landscape
6	Closed shrublands	Lands with woody vegetation less than 2 m tall and with shrub canopy cover is >60%. The shrub foliage can be evergreen or deciduous
7	Open shrublands	Lands with woody vegetation less than 2 m tall and with shrub canopy cover is between 10% and 60%. The shrub foliage can be either evergreen or deciduous
8	Woody Savannas	Lands with herbaceous and other understorey systems, and with forest canopy cover between 30% and 60%. The forest cover height exceeds 2 m
9	Savannas	Lands with herbaceous and other understorey systems, and with forest canopy cover between 10% and 30%. The forest cover height exceeds 2 m
10	Grasslands	Lands with herbaceous types of cover. Tree and shrub cover is less than 10%
11	Permanent wetlands	Lands with a permanent mixture of water and herbaceous or woody vegetation that cover extensive areas. The vegetation can be present in either salt, brackish, or fresh water
12	Cropland	Lands covered with temporary crops followed by harvest and bare soil period (e.g., single and multiple cropping systems. Note that perennial woody crops will be classified as the appropriate forest or shrub land cover type
13	Urban and built-up	Land covered by buildings and other man-made structures. Note that this class will not be mapped from the AVHRR imagery but will be developed from the populated places layer that is part of the Digital Chart of the World
14	Cropland/natural vegetation mosaics	Lands with a mosaic of croplands, forest, shrublands, and grasslands in which no one component comprises more than 60% of the landscape

(*continued*)

Table 34.1 (continued)

Class	Class name	Description
15	Snow and ice	Lands under snow and/or ice cover throughout the year
16	Barren	Lands exposed soil, sand, rocks, or snow and never has more than 10% vegetated cover during any time of the year
17	Water bodies	Oceans, seas, lakes, reservoirs, and rivers. Can be either fresh or salt water bodies

Table 34.2 Land cover categories of FAO (Land Cover Classification System (LCCS), FAO Corporate Document Repository, FAO, www.fao.org/docrep/003/x0596e/X0596e01i.htmlTopOfPage)

11. Cultivated and managed terrestrial areas	Tree crops
	Shrub crops
	Herbaceous crops
	Graminoid crops
	Non-graminoid crops
	Managed lands
12. Natural and semi-natural terrestrial vegetation	Forest
	Woodland
	Thicket
	Shrubland
	Grasslands
	Sparse vegetation
	Lichens/Mosses
23. Cultivated aquatic or regularly flooded areas	Aquatic or regularly flooded graminoid crops
	Aquatic or regularly flooded non-graminoid crops
24. Natural and semi-natural aquatic or regularly flooded vegetation	Forest
	Woodland
	Closed shrubs
	Open shrubs
	Grasslands
	Sparse vegetation
	Lichens/Mosses
15. Artificial surfaces and associated areas	Built-up areas
	Non-built-up areas
16. Bare areas	Consolidated areas
	Unconsolidated areas
27. Artificial surfaces and associated areas	Artificial water bodies
	Artificial snow
	Artificial ice
28. Natural water bodies, snow, and ice	Natural water bodies
	Snow
	Ice

Table 34.3 Land cover categories of USGS (Anderson et al. 1976)

Level I	Level II
1 Urban or built-up land	11 Residential
	12 Commercial and services
	13 Industrial
	14 Transportation, communications, and utilities
	15 Industrial and commercial complexes
	16 Mixed urban or built-up land
	17 Other urban or built-up land
Agricultural land	21 Cropland and pasture
	22 Orchards, groves, vineyards, nurseries, and ornamental horticultural areas
	23 Confined feeding operations
	24 Other agricultural land
Rangeland	31 Herbaceous rangeland
	32 Shrub and brush rangeland
	33 Mixed rangeland
Forest land	41 Deciduous forest land
	42 Evergreen forest land
	43 Mixed forest land
Water	51 Streams and canals
	52 Lakes
	53 Reservoirs
	54 Bays and estuaries
Wetland	61 Forested wetland
	62 Non-forested wetland
Barren land	71 Dry salt flats
	72 Beaches
	73 Sandy areas other than beaches
	74 Bare exposed rock
	75 Strip mines quarries, and gravel pits
	76 Transitional areas
	77 Mixed barren land
Tundra	81 Shrub and brush tundra
	82 Herbaceous tundra
	83 Bare ground tundra
	84 Wet tundra
	85 Mixed tundra
Perennial snow or ice	91 Perennial snowfields
	92 Glaciers

Many other kinds of classifiers are used in the remote sensing applications. They are, for instance, decision tree classifiers, neural net, support vector machine (SVM), etc.

Neural Net

Artificial neural network is one of the strongest classifiers for remote sensing data. The most popular architecture of neural network for classification is the multilayer perceptron as shown in Fig. 34.29. Multilayer perceptron is composed of input layer, output layer, and hidden layers. The number of hidden layers may change according to the problem. In Fig. 34.29, only one hidden layer is shown.

The output signal $\boldsymbol{z} = (z_1, \cdots, z_k)^t$ to the input signal $\boldsymbol{x} = (x_1, \cdots, x_n)^t$ is expressed as follows.

$$\left.\begin{aligned} \zeta_j &= \sum_{i=1}^{I} a_{ij} x_i + a_{0j} \\ y_j &= f_{\text{hidden}}(\zeta_j) \\ \eta_k &= \sum_{j=1}^{J} b_{jk} y_j + b_{0k} \\ z_k &= f_{\text{out}}(\eta_k) \end{aligned}\right\} \tag{34.5}$$

Here

a_{ij}: Weight from ith input to jth hidden layer unit

b_{jk}: Weight from jth hidden layer unit to kth output unit

a_{0j}, b_{0k}: Biases of jth unit of hidden layer and kth unit of output layer, respectively

f_{hidden}, f_{out}: Input–output function of hidden layer unit and output layer unit, respectively

Usually, logistic functions are used for f_{hidden}, and f_{out} is defined according to each problem.

For the learning process of this kind of neural network, error back propagation learning (Rumelhart et al. 1986a, b) is used. This algorithm is shown as follows.

Suppose the combinations of input data and training data as $\{\boldsymbol{x}_p, \boldsymbol{u}_p | \, p = 1, \cdots, P\}$. In this algorithm, the following evaluation criterion is minimized.

$$\varepsilon_{\text{emp}}^2 = \sum_{p=1}^{P} \left\| \boldsymbol{u}_p - \boldsymbol{z}_p \right\|^2 = \sum_{p=1}^{P} \varepsilon_{\text{emp}}^2(p)$$

Using the method of steepest descent, the following iterative equations are deduced.

$$a_{ij}^{l+1} = a_{ij}^{l} + 2\alpha \sum_{p=1}^{P} \gamma_{pj} v_{pj} x_{pi}$$

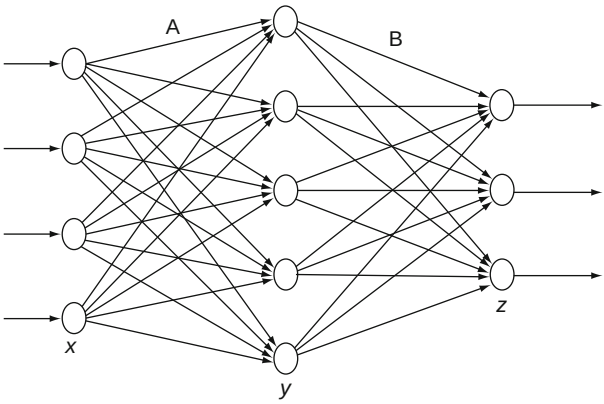


Fig. 34.29 A *block* diagram of multilayer perceptron (Takagi and Shimoda 2004)

$$b_{jk}^{l+1} = b_{jk}^{l} + 2\alpha \sum_{p=1}^{P} \delta_{pk} y_{pj}$$

Here
α: Learning rate (>0)

$$v_{pj} = y_{pj}(1 - y_{pj})$$

$$\gamma_{pj} = \sum_{k=1}^{K} \delta_{pk} b_{jk}$$

$$\delta_{pk} = u_{pk} - z_{pk}$$

Neural network classifier usually gives higher accuracy than MLC. It can deal with nonlinear problem as well as non-normal distribution problems.

Support Vector Machine

Support vector machine (SVM) (Vapnik 1995) started from the pattern classifier of linearly separable two classes. It generates a hyperplane with the largest margin (the minimum distance from training samples to the hyperplane). The learning process uses Lagrange's method of undetermined multipliers, formulated by convex quadratic programming.

However, most problems are not linearly separable and land cover classification needs many-class classifications. In order to deal with nonlinearly separable problems, the original space is nonlinearly projected to higher-order space. In the real calculations, this projection is not actually calculated. Instead, this calculation is replaced by kernel function calculations. This process is called a kernel trick (Schölkopf et al. 1998; Smola 1996).

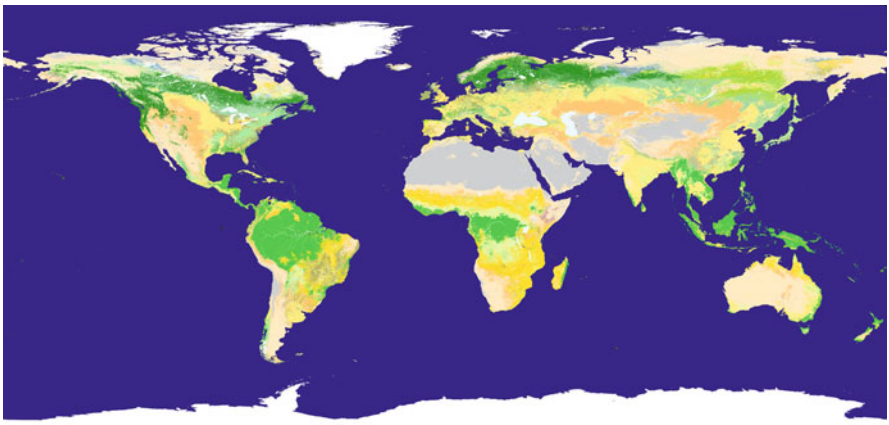


Fig. 34.30 Global land cover map generated by NASA (http://earthobservatory.nasa.gov/Newsroom/view.php?id=22585)

Examples of Land Cover Classifications

Several approaches have been made to generate global land cover maps. Three kinds of approaches are briefly introduced here. The first one is MODIS land cover product generated by NASA using MODIS data. Many kinds of classification features are used in this classification, i.e., land/water mask, Nadir BRDF adjusted Reflectance (NBARs), and spatial texture derived from 250 m red band, directional reflectance information, EVI, snow cover, land surface temperature, and DTM. Classifier is the decision tree classifier.[6] An example of this product is shown in Fig. 34.30.

The second example is generated by ESA using MERIS data. In this approach, also NBAR-like products are used for the classification. The classifier is mainly unsupervised clustering (ISODATA). These products are called GlobCover.[7] An example of this product is shown in Fig. 34.31.

The third example is made by the author's lab. It uses the same MODIS product with NASA product, i.e., NBAR, but the features used are very different from other two examples. Usually, there remain some clouds or snow in mosaicked images. In order to avoid these noises, we have developed a time domain co-occurrence matrix. This matrix is similar to the usual co-occurrence matrix, but the distance is not the space distance, but time difference is used as distance (Fukue et al. 2010). An example of this product is shown in Fig. 34.32. In this figure, upper image shows the result using MODIS surface reflectance and the bottom image shows the result of NBAR.

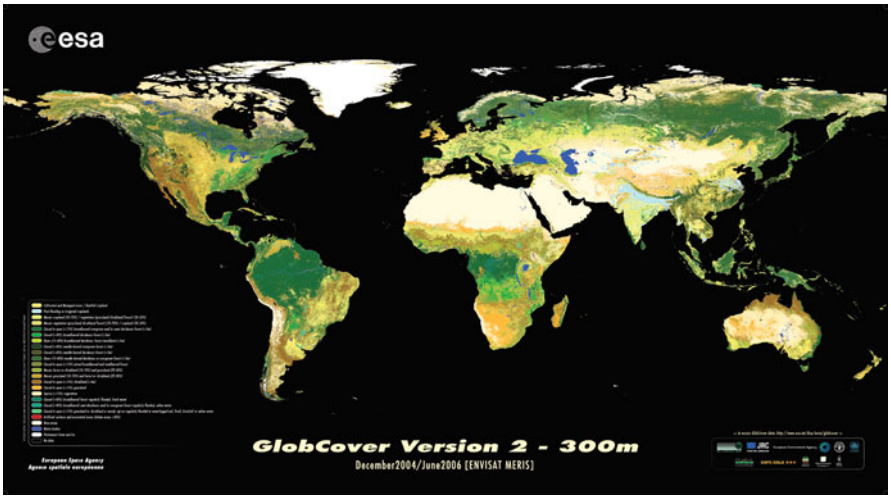


Fig. 34.31 Global land cover map using MERIS data generated by ESA (http://www.esa.int/esaEO/SEMZ16L26DF_planet_0.html) (Courtesy of ESA)

Geological Applications

Geological applications of remote sensing are one of the fastest applications. Four kinds of applications are used in this field.

The first application was to use satellite imagery in logistics. Usually, mineral exploration target areas are very large, and it is common that there are no large-scale base maps. In these circumstances, satellite images can be used as base maps.

The second applications have been the geological structures. Some of the distinct geological structures, like anticlinal structures, circular structures, etc., can be easily interpreted from the satellite images. Petroleum oils can only exist under anticlinal structure and noble metals can be found along circular structures. Figure 34.33 shows an example of anticlinal structures observed by OPS on JERS-1. Other structures which can be interpreted from satellite images are lineaments. Some of these linear lines in the images are considered to express the underground fault structures. Figure 34.34 shows a sample of lineament extraction from a SAR image.

The third application is to directly detect rock types. For bare rock areas, rock types can be classified using their spectral signatures. Especially, metamorphic rocks have discriminative spectral signatures in short wave infrared region. Spectral features of mineral rocks and corresponding ASTER (► Chap. 1) bands are shown in Fig. 34.35. Figure 34.36 shows the rock-extracted results from ASTER data. Figure 34.36a and b shows the visible

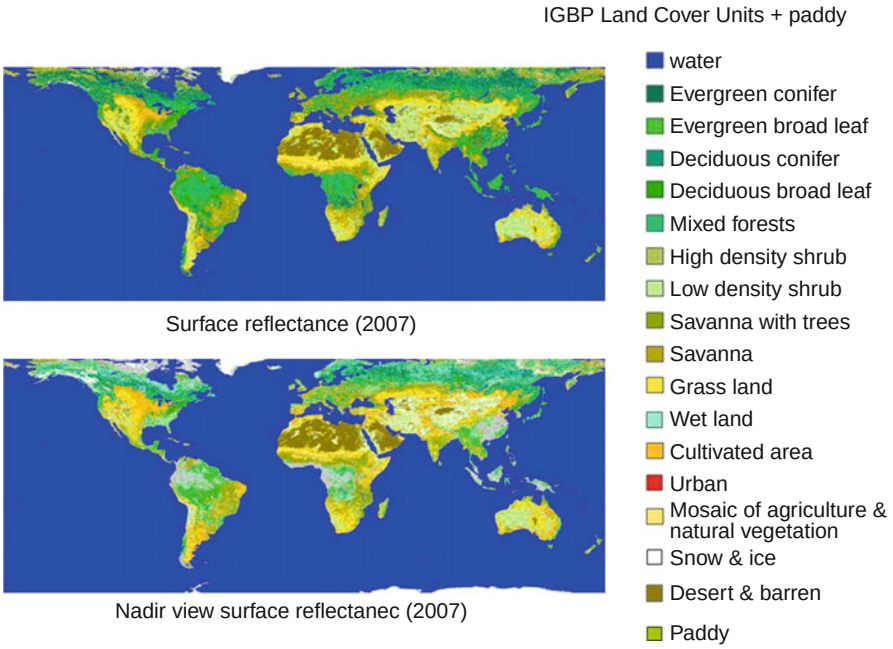


Fig. 34.32 Global land cover map using time domain co-occurrence matrix (Fukue et al. 2010)

and near-infrared image and enhanced short wave infrared image, respectively. Figure 34.36c is the distribution of White Mica and Fig. 34.36d is the distribution of Calcite.

Now, most of the petroleum fields over land are found. So, the efforts to find a new petroleum field are concentrated over jungle and the ocean. For ocean explorations, remote sensing data are sometimes used to find out oil spills, which may be caused by ocean underground petroleum fields.

SAR data are also used for geological applications. It is sometimes effective to find lineaments. Another application areas of SAR data are detection of ground movement. Using differential interferometric SAR, the displacement of ground can be detected with centimeter order accuracy. Detected land displacement is used for earthquake studies as well as volcano monitorings. Figure 34.37 shows the land displacement by the 2011 off the Pacific coast of Tohoku Earthquake observed by PALSAR (▶ Chap. 1) on ALOS (▶ Chap. 1).

Soil Moisture

Soil moistures are not only important parameters to estimate thermal fluxes over land and to estimate evapotranspiration, but also affect crop yields. Soil moistures

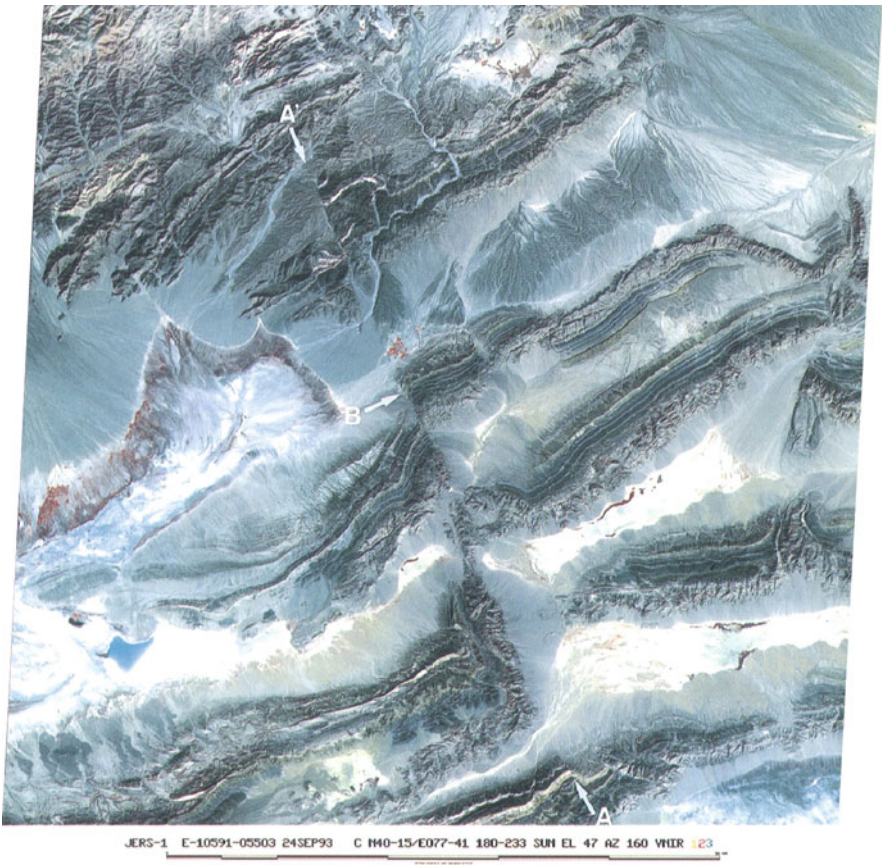


Fig. 34.33 An example of anticlinal structures observed by OPS on JERS-1

are usually retrieved from microwave radiometers. The sensitivity of microwave radiometer increases with the decrease of frequencies. Until recently, the lowest microwave radiometer frequency was C band. AMSR (► Chap. 1) and AMSR-E (► Chap. 1) were the only radiometers which have C band. The problem of C-band radiometer is that the spatial resolution is very low. With 2 m aperture antenna, the spatial resolution is around 50 km. This is a very wide area over land, and its validation is very difficult. Figure 34.38 shows changes of soil moisture of Africa between February and August obtained from AMSR-E.

Very recently, much lower frequency radiometer was launched. It is SMOS (► Chap. 1) and has L-band radiometer. Figure 34.39 shows an example of soil moisture over Australia obtained from SMOS.

Another approach is to use SAR for soil moisture retrievals. SAR has higher spatial resolution compared to microwave radiometers. Several attempts have been made, but the soil moisture retrievals are also difficult because of the sensitivity change associated with incidence angle change.

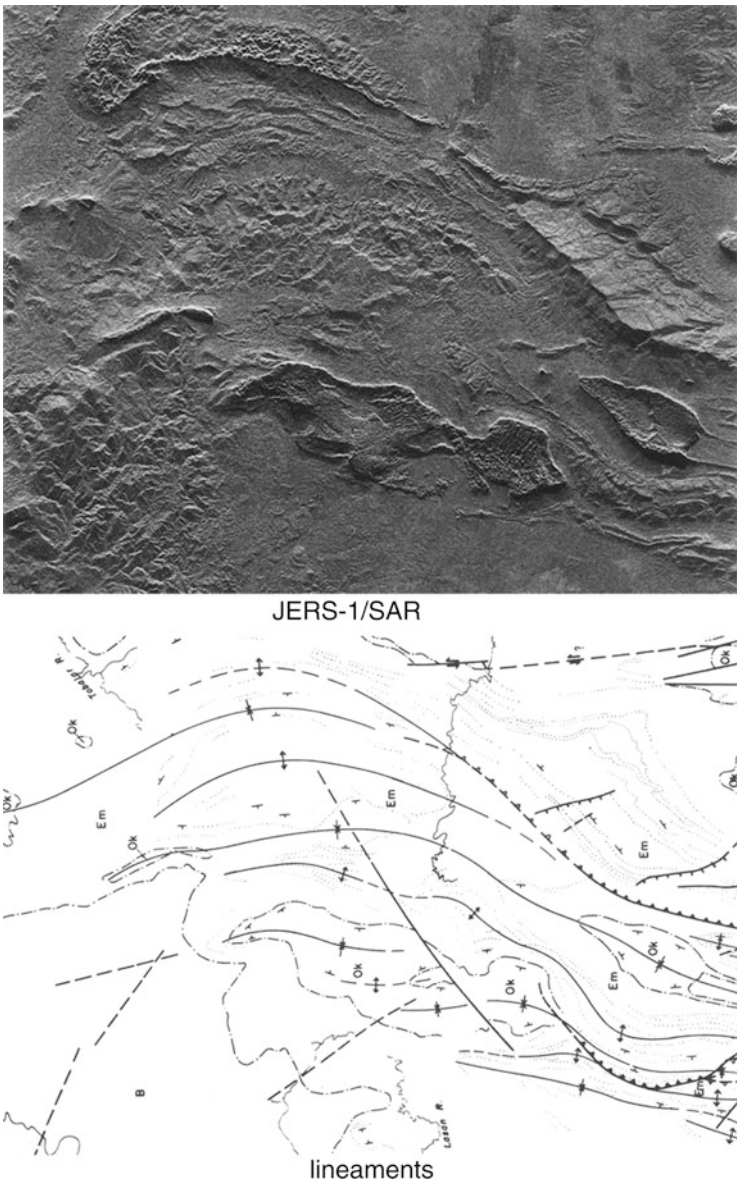


Fig. 34.34 Lineament extraction. (PALSAR project http://www.palsar.ersdac.or.jp/data/kouhou/) (**a**) JERS-1 SAR image of mount Morgan in South Australia. (**b**) Extracted lineaments from the above image (Courtesy of ERSDAC)

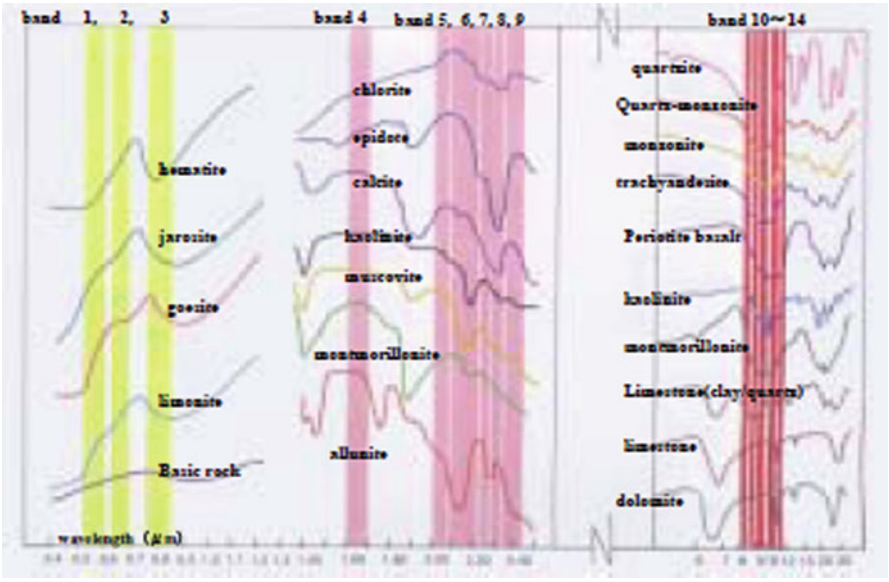


Fig. 34.35 Spectral signatures of mineral rocks and corresponding ASTER bands (ASTER reference guide (version 1.0), ERSDAC, 2003, p.13) (Courtesy of ERSDAC)

Accuracies of the retrieved soil moistures from C-band radiometers and SAR are not sufficiently good. SMOS and Aquarius may provide higher accuracy soil moistures after sufficient validation activities.

Carbon Cycle

In order to accurately project the global warming, understandings of carbon cycle is very important. Land vegetations absorb carbon dioxide, but the quantity of this absorption is not clearly understood. The carbon flux of vegetations should be clarified, but it is rather difficult from satellite data. The first step is to estimate terrestrial primary productions of vegetations from satellite data. The gross primary production (GPP (gCm-2yr-1)), which is the fixed amount of carbon by photosynthesis can be expressed as follows (Monteith 1972; Running et al. 2000, 2004; Nemani et al. 2003).

$$GPP = \varepsilon \times fAPAR \times PAR$$
$$\varepsilon = \varepsilon_{\max} \times T_f \times VPD_f$$

Here
ε: Light use efficiency parameter ($gCMJ^{-1}$)
PAR: Photosynthetically active radiation (MJm-2day-1)

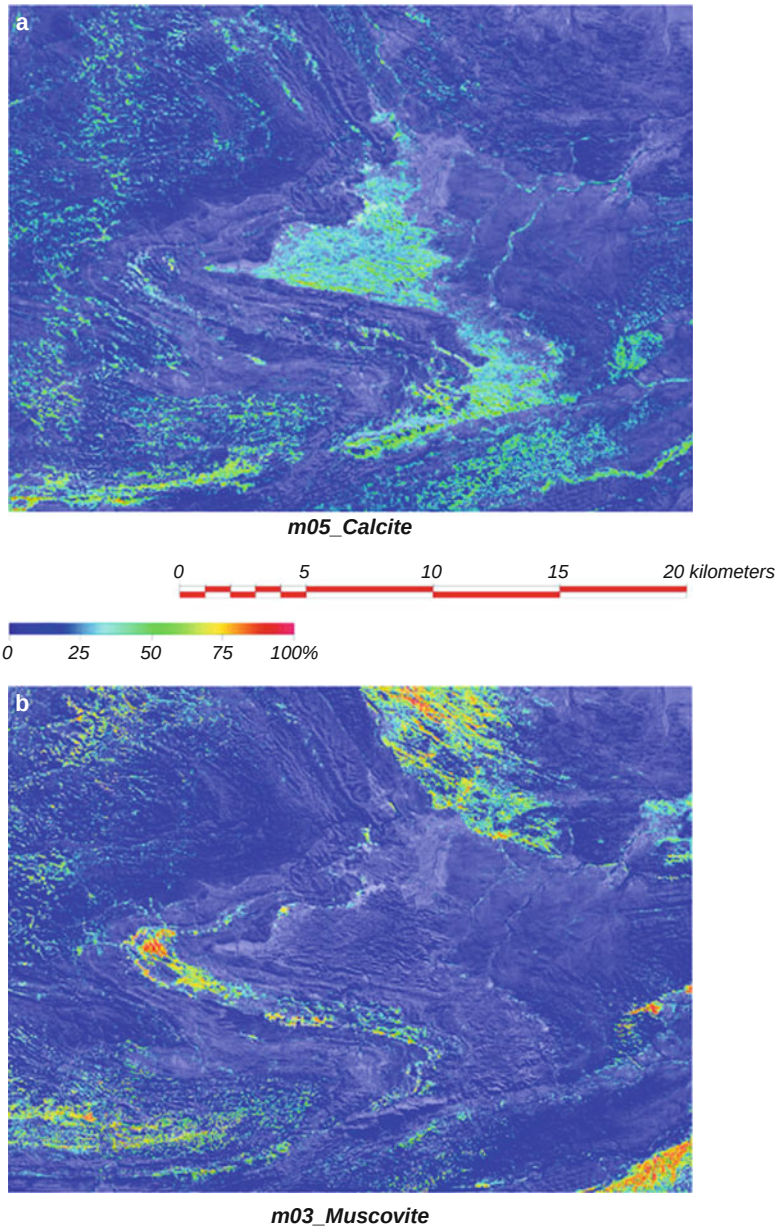


Fig. 34.36 Mineral rocks extracted from ASTER data (The mineral distribution map near Talc Deposit Area of Mt.Fitton (Australia) using full band data of ASTER, http://www.science.aster.ersdac.or.jp/en/topic_image/Geology/001.html): (**a**) calcite, (**b**) mica (Courtesy of ERSDAC)

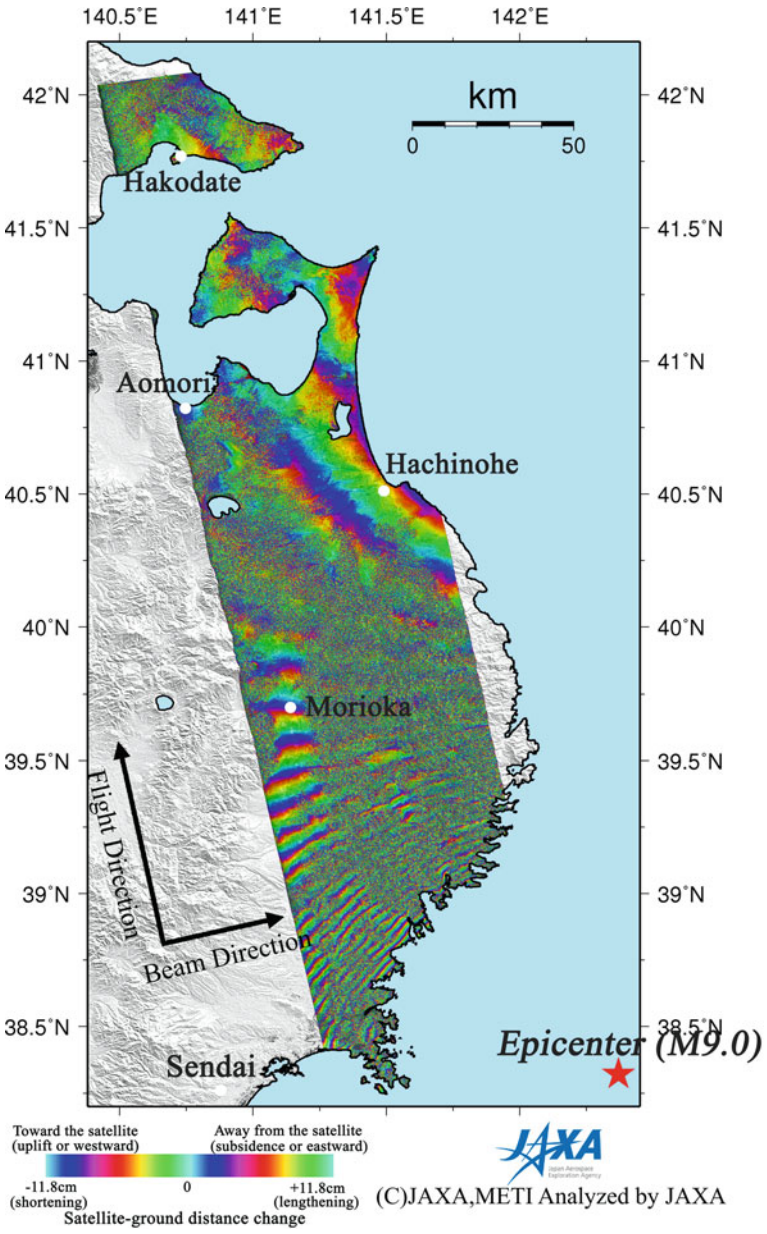


Fig. 34.37 The land displacement by the 2011 off the Pacific coast of Tohoku Earthquake observed by PALSAR on ALOS (http://www.eorc.jaxa.jp/ALOS/en/img_up/l_dis_inf_tohokueq_110315_f2e.htm) (Courtesy of JAXA)

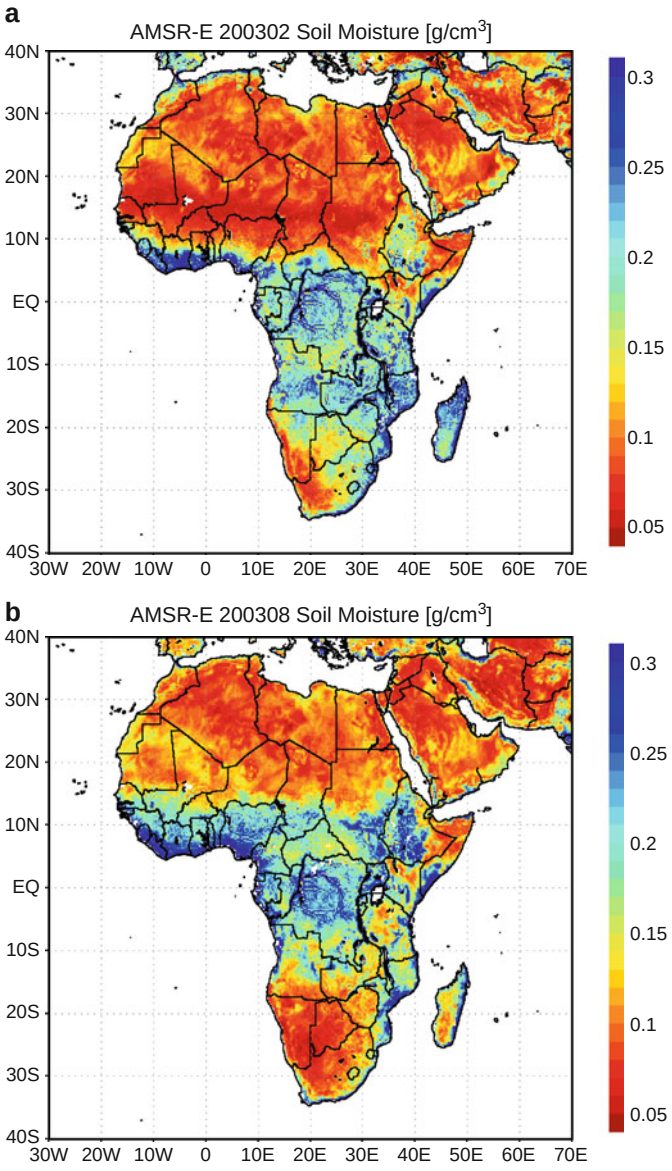


Fig. 34.38 Soil moisture of Africa obtained from AMSR-E: (**a**) February 2003, (AMSR-E 200302 soil moisture, http://www.eorc.jaxa.jp/imgdata/topics/2004/img/tp040723_ 01.gif) (**b**) August 2003 (AMSR-E 200308 soil moisture, http://www.eorc.jaxa.jp/imgdata/topics/2004/img/tp040723_ 02.gif) (Courtesy of JAXA)

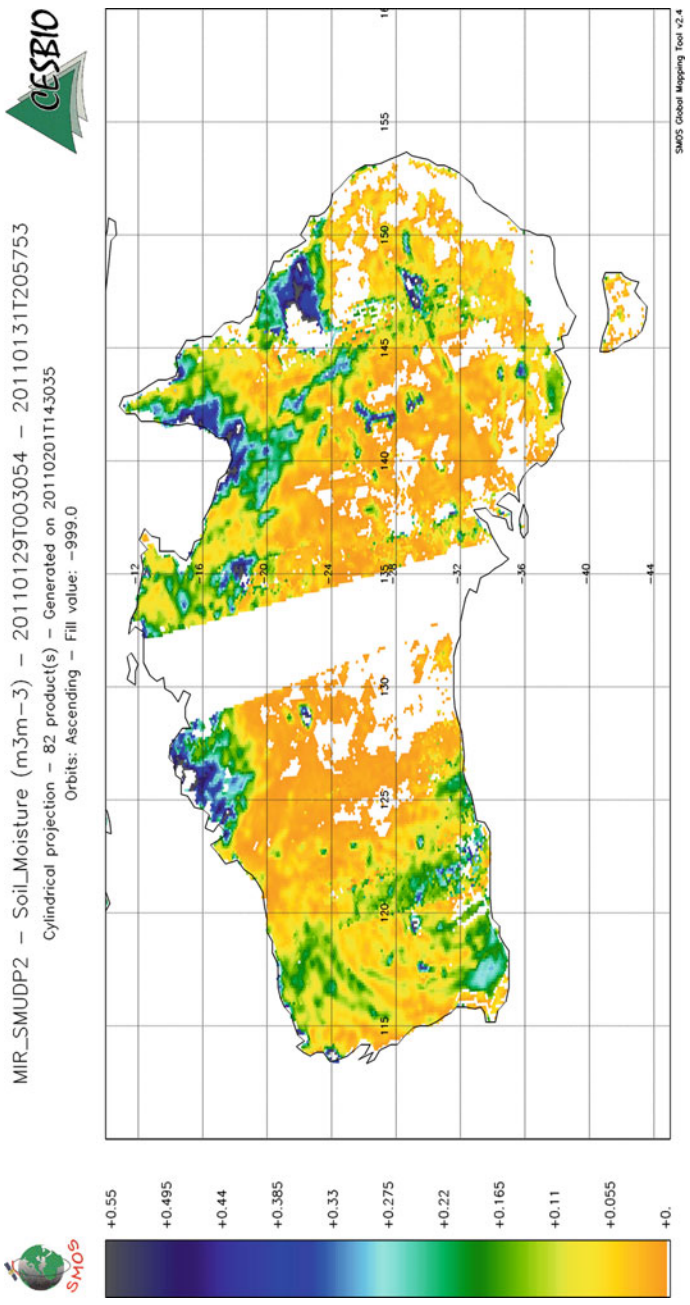


Fig. 34.39 Soil moisture over Australia obtained from SMOS during January 29–31, 2011 (Australia and Yasi. . . New floods? . . . what is SMOS seeing to help forecasts? http://smsc.cnes.fr/SMOS/GP_actualites.htm) (Courtesy of CNES)

fAPAR: Fraction of absorbed *PAR*

$\varepsilon_{\max}$: Potential maximum ε under optimal conditions (no environmental stress)

T_f: Reductions in photosynthesis under low temperature condition

VPD_f: Reductions in photosynthesis under suboptimal surface air vapor pressure deficit

PAR can be derived from satellite data, and *fAPAR* has correlation with satellite-derived *LAI* (leaf area index, calculated from *NDVI* or *EVI*) or *NDVI* or *EVI*. *EVI* is expressed as follows.

$$EVI = G \times \frac{NIR - Red}{NIR + C_1 \times Red - C_2 \times Blue + L}$$

Here, G, C_1, C_2, and L are empirically defined coefficients.

From *GPP*, *NPP* (net primary production) is derived as follows.

$$NPP = GPP - R$$

Here, R is above ground respiration and can be determined from *LAI* and temperature. Figure 34.40 shows an example of global *NPP* derived from MODIS (▶ Chap. 1) data.

The carbon flux is expressed by net ecosystem production (*NEP*). *NEP* is calculated from *NPP* by subtracting under the ground respirations. However, the underground respirations cannot be retrieved from satellite data. In order to estimate *NEP*s, carbon cycle model is necessary. Many kinds of terrestrial carbon cycle models are proposed (Running and Gower 1991; Esser et al. 1994; Foley et al. 1996; Tian et al. 1999; Sitch 2000).

Another approach is to combine ground-based carbon flux estimation with atmosphere-based carbon concentrations. As described in 2.5 of this chapter, satellite-borne sensors now can retrieve atmospheric carbon dioxide and methane in good accuracies. Therefore, models which can describe both atmospheric concentrations as well as ground-based fluxes with assimilation capability may result in better accuracy carbon cycle understandings.

Cryospheric Applications

Sea Ice

Sea ice plays an important role for energy circulations of the Earth. Sensible heat and latent heat over sea ice are very different from those over open ocean. It is also important for ship navigations over high-latitude areas. Sea ice has been monitored using microwave radiometers for a long time. The geophysical parameter which is retrieved from microwave radiometer is ice concentrations. Ice concentration is the

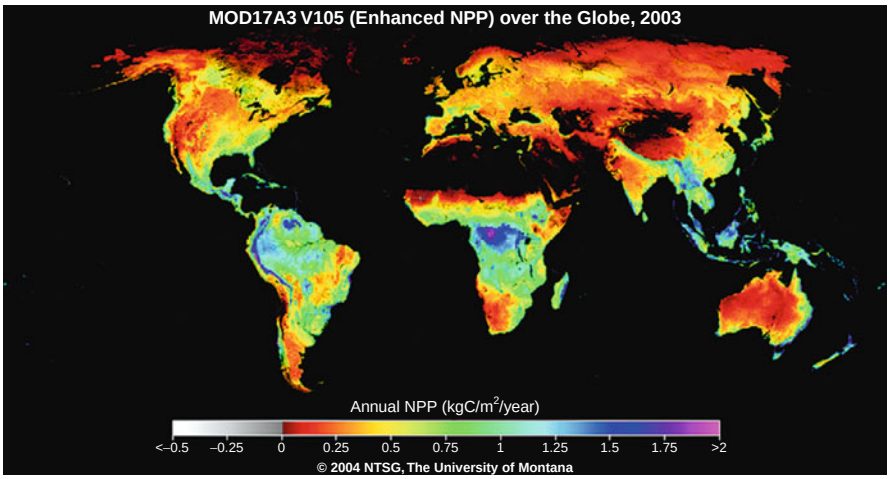


Fig. 34.40 Global NPP derived from MODIS (M. Zhao, S. Running, F.A. Heinsch, R. Nemani, Collection 005 change summary for the MODIS land vegetation primary production (17A2/A3) algorithm, http://landweb.nascom.nasa.gov/QA_WWW/forPage/C005_Change_NPP.pdf) (Courtesy of NASA)

ratio of sea ice–covered area to the total area. Figure 34.41 shows an example of sea ice concentrations retrieved from AMSR-E (► Chap. 1).

Microwave scatterometer also can monitor sea ice. However, parameters which can be retrieved from microwave scatterometer are different from radiometers. From scatterometer, areas of sea ice and the discrimination between multi-year ice and new ice can be obtained.

There are several other parameters which are important for monitoring sea ice. One is the thickness of sea ice, but it is very difficult to retrieve sea ice thickness from satellite data. Another parameter which is important is the thickness of thin ice. When the sea ice is very thin, i.e., less than 1 m, the energy flux between sea and atmosphere changes drastically according to the thickness. Many attempts have been made to retrieve this parameter, but still accurate algorithms are not present.

Another application of sea ice is the monitoring of ice bergs. Microwave scatterometer is now used for this purpose as well.

Snow and Glaciers

Snow also perturbs global climate. Snow is also very important for water supply. Several geophysical parameters are important, i.e., snow cover, snow depth, and snow albedo. Snow cover and albedo can be retrieved from optical sensors, while snow depth can be retrieved from microwave radiometers. Figure 34.42 shows an

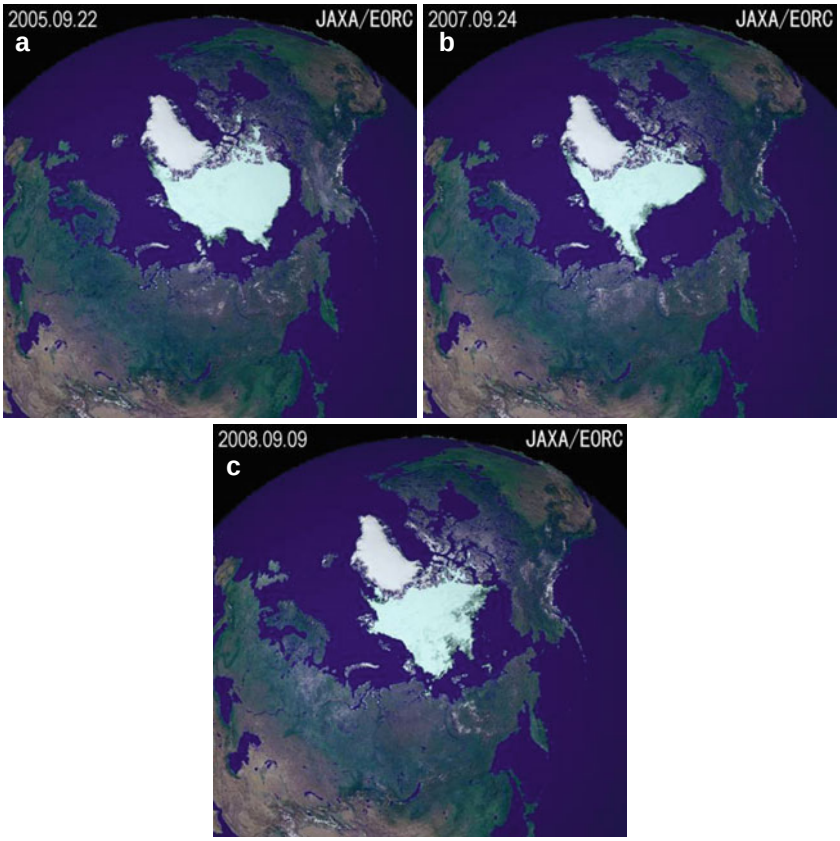


Fig. 34.41 Sea ice concentration of Antarctic sea of 2005, 2007, and 2008 (Arctic sea-ice second smallest on record, http://www.eorc.jaxa.jp/en/imgdata/topics/2008/tp081022.html): (**a**) 2005, (**b**) 2007, (**c**) 2008 (Courtesy of JAXA)

example of global snow cover from MODIS (► Chap. 1) data, and Fig. 34.43 shows the global snow depth retrieved from AMSR-E (► Chap. 1) data.

Glaciers are affected by global warmings. Many of the existing glaciers are retreating. It is not clear that these retreats are caused by global warming or not. Anyway, it is important to monitor the motion of these glaciers. The forefront of glaciers can be monitored using optical sensors. Another way of monitoring glaciers is to use SAR interferometry (Fatland 1998; Mohr et al. 1998; Joughin et al. 1998; Rabus and Fatland 2000). Using the SAR interferometry, the retreating speed of glaciers can be obtained.

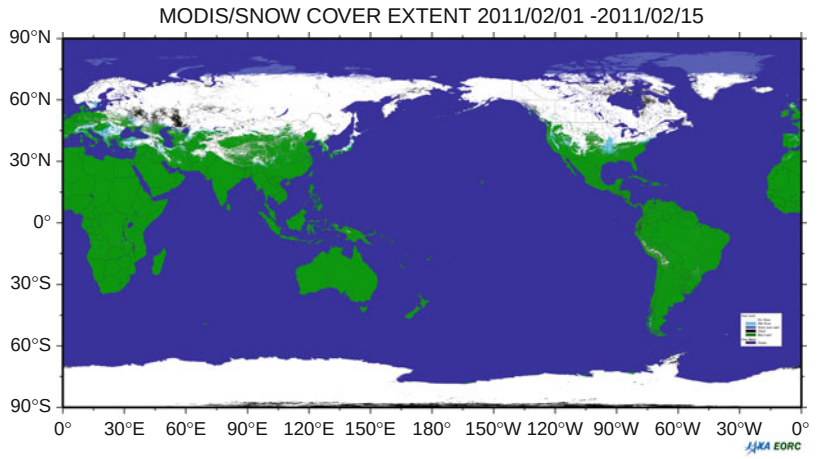


Fig. 34.42 An example of global snow cover from MODIS data (http://kuroshio.eorc.jaxa.jp/JASMES/data/GL/CSF/201102/MDS20110201_20110215_GLBOD0HM_SNWFG_EQ05KM_100.png) (Courtesy of JAXA)

Operational Applications

NWP and Weather Forecasting

The weather forecasting of developed countries is based on the results of numerical weather prediction (NWP) softwares. Until around 15 years ago, these NWPs used only in situ data for the input. However, these NWPs now use many kinds of satellite data in addition to the in situ data. Most popular data used as input by NWPs are microwave sounder data, thermal IR sounder data, microwave radiometer data, microwave scatterometer data, microwave altimeter data, GPS occultation data, etc. Another very important data for NWP is atmospheric winds obtained from geostationary satellite. From visible and thermal infrared images, motion of clouds is extracted and the winds near the clouds are retrieved. For clear areas, water vapor motion extracted from mid-IR is used for winds retrieval. However, geostationary satellite imagers can cover only within $\pm 50^\circ$ latitudinal areas. For higher latitudes, winds extracted from MODIS are now used.

At the first stage of the NWP applications, retrieved geophysical parameters were used for assimilations. However, for thermal IR and microwave sounder and radiometer data, radiances from these sensors are now directly assimilated to the NWP.

The impact of satellite data to the NWP is positive, but to what extent the improvements are made is not so clear, because NWP models themselves have

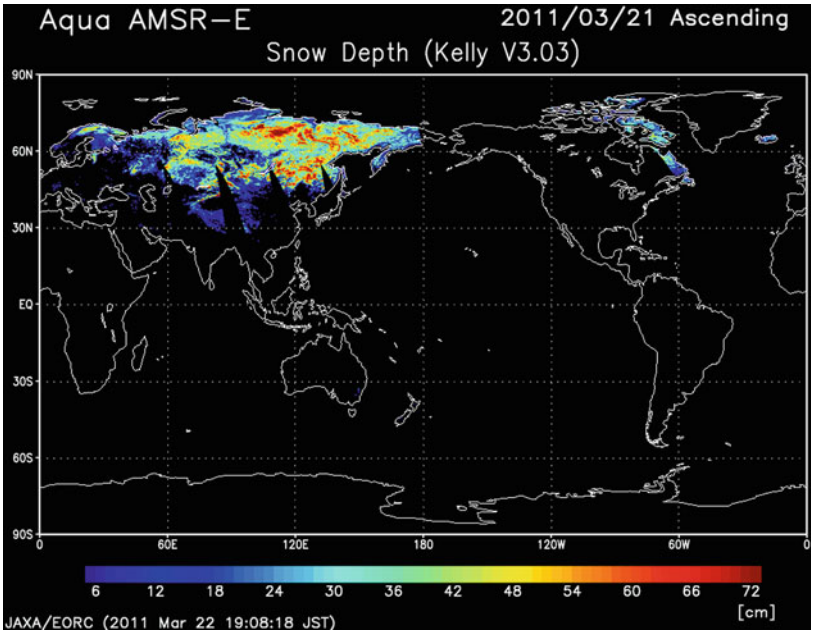


Fig. 34.43 Global snow depth retrieved from AMSR-E data (http://sharaku.eorc.jaxa.jp/cgi-bin/adeos2/amsr/l3brws/l3brws.cgi?lang=eamp;sat=P1amp;ad=Aamp;prd=SWE_Kelamp;ver=newamp;map=E0amp;y=2011amp;m=3amp;d=21amp;ny=2011amp;nm=3amp;nd=21) (Courtesy of JAXA)

been improved. Figures 34.44 and 34.45 show impacts of satellite data to NWP estimated by ECMWF. In Fig. 34.44, baseline is NWP without any satellite data, AMV is NWP with satellite-derived atmospheric winds data, EUCOS is NWP with AMV plus AMSU (▶ Chap. 1) data, and Control is NWP with all satellite data. From these figures, impacts are larger in southern hemisphere than northern hemisphere. In northern hemisphere, satellite data impacts are around 1.6 days at the 6-day forecast, while it is around two and a half days in southern hemisphere. From Fig. 34.45, it is shown that satellite winds, water vapor, optical sounder, and microwave scatterometer have large impacts to NWP.

Fisheries

Fisheries are one of the largest operational application areas of remote sensing. Satellite data applications to fisheries have begun from early 1980s. At the first stage, SST was used. The accuracy of satellite-derived SST is not enough to directly find out fishing grounds, but from the SST imagery, it is very easy to detect SST front and positions of oceanic current. Especially, some of the oceanic fronts are good fishing grounds.

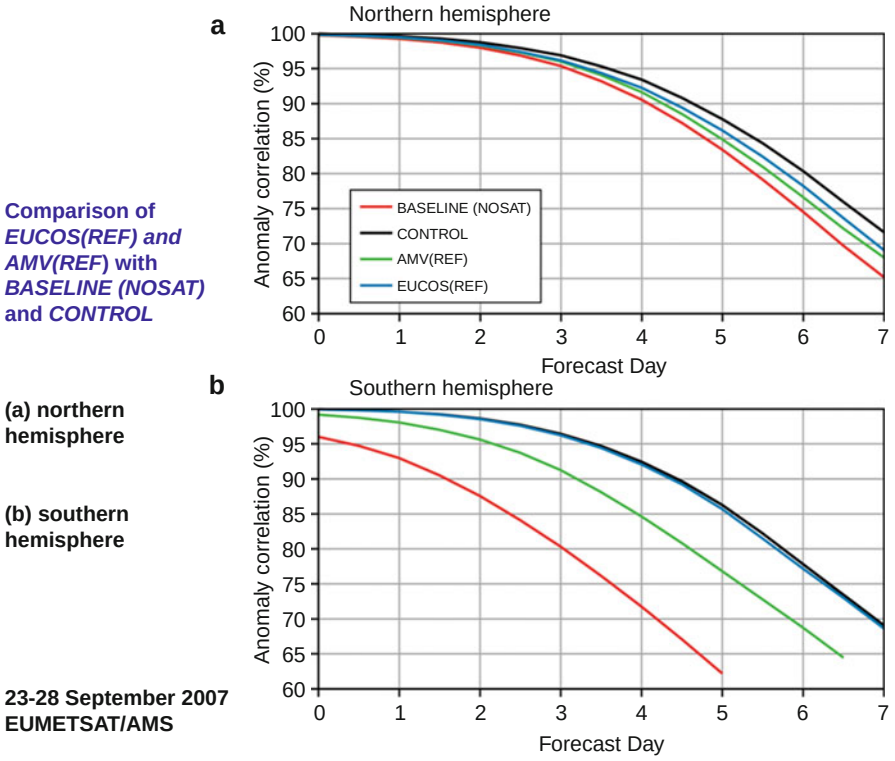


Fig. 34.44 Impacts of satellite data to NWP for several sensor combinations estimated by ECMWF (Kelly 2007)

Another applications have been chlorophyll-a. Chlorophyll-a concentrations correspond to phyto-plankton concentrations, which further correlate to zoo plankton concentrations. Nowadays, fishermen use many other satellite sensor data, e.g., microwave altimeter data, microwave scatterometer data, microwave radiometer data, etc., for finding good fishing grounds.

Figure 34.46a shows SST distributions and fishing grounds. From this figure, fishing grounds lie along the front of SST. Figure 34.46b shows the chlorophyll-a distribution and fishing grounds. From this figure, it is shown that fishing grounds lie near high chlorophyll-a concentration areas.

This figure presents GLI sea surface temperature and chlorophyll-a concentration images overlaid on fisheries of skipjack and tuna. Fisheries of warm-water skipjack were distributed in areas of relatively high sea surface temperature and low chlorophyll-a concentration. Also, saury is in relatively low sea surface temperature and high chlorophyll-a concentration.

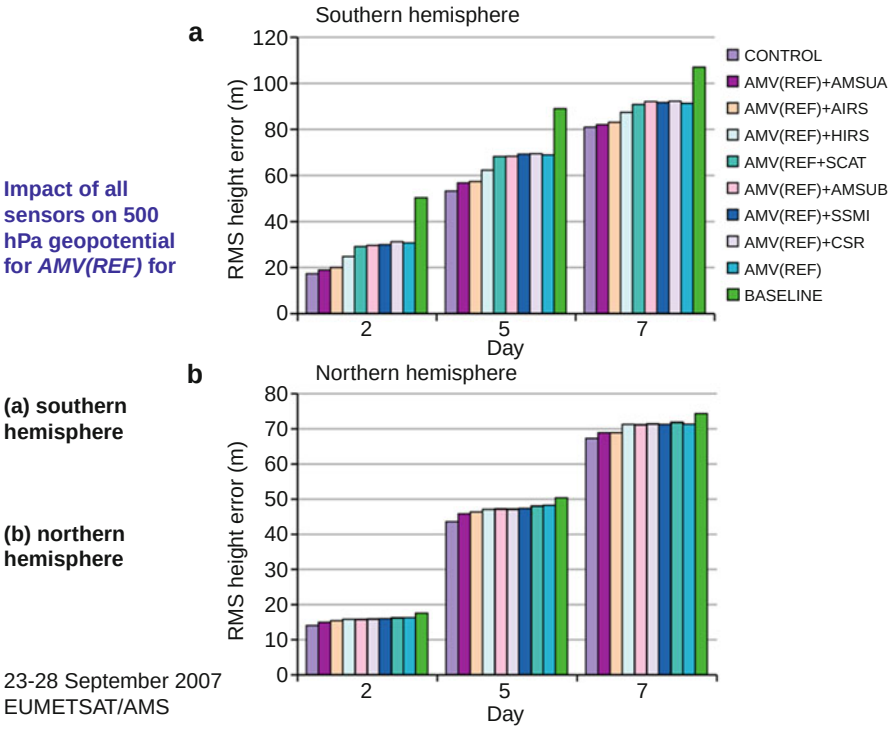


Fig. 34.45 Impacts of satellite data to NWP for each sensor estimated by ECMWF (Kelly 2007)

Disasters

Biomass Burnings

The number of total biomass burnings in the world is around one million times. From these biomass burnings, many kinds of atmospheric constituents are emitted. These gases include, but are not limited to, CO_2, CO, CH_4, NO, NH_3, O_3, etc. The total CO_2 emission amount varies depending upon each year, and is rather difficult to estimate, but may range from 2 Gton carbon to 4 Gton carbon. However, as the regrown vegetation absorbs CO_2, the net emissions will be much smaller (Levine et al. 1995; Jacobson et al. 2004). It is also one of the largest sources of tropospheric ozone, thus degrading the quality of atmosphere. Large-scale biomass burnings also take the lives of people, and burn household articles.

In order to monitor the global biomass burnings, satellite monitorings are the only means. Many kinds of satellite sensors are used for this purpose. The most used sensors are AVHRR (▶ Chap. 1) on NOAA (▶ Chap. 1), MODIS (▶ Chap. 1) on Terra (▶ Chap. 1) and Aqua (▶ Chap. 1), and imagers onboard geostationary meteorological satellites. Infrared channels are used for fire detection. However, 11 and 12 μm regions are saturated quite quickly, so 3.7 μm or shorter wavelength is

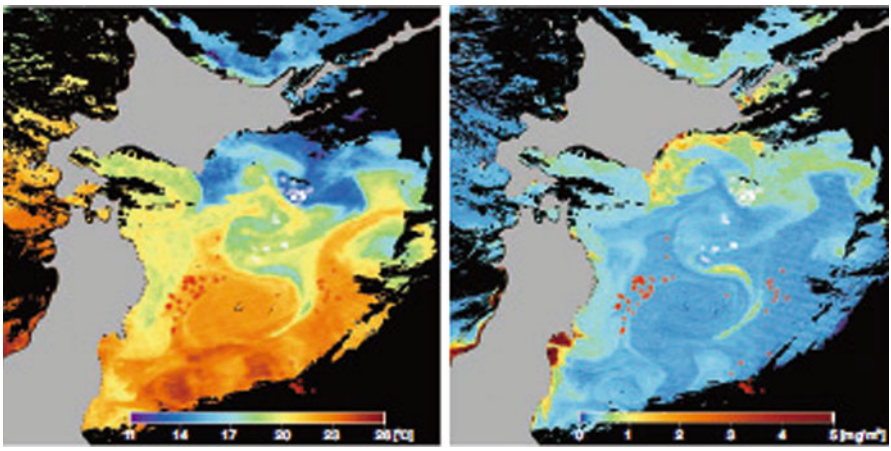


Fig. 34.46 SST (**a**) and chlorophyll-a (**b**) distribution and fishing grounds of Sanriku Coast, Japan, observed by GLI on ADEOS2 and fishing boats information (http://suzaku.eorc.jaxa.jp/GLI/doc/GLI_BOOK_CD/PDF/CHAP_6.PDF) (Courtesy of JAXA)

more appropriate. Global fire maps can be accessed through MODIS Rapid Response System Global Fire Maps[8] of NASA or Current & Archived Significant Global Fire Events and Fire Season Summaries[9] of International Strategy for Disaster Reduction (ISDR), etc. Figure 34.47 shows a global fire map of 03/12/11–03/21/11 distributed by the above NASA site.

Satellite-derived fire monitorings have some disadvantages. One of the problems is that the spatial resolution of infrared sensors is usually not so fine; hence, burnt areas are overestimated. Another disadvantage is that it cannot monitor under thick clouds.

Floods

Flood is one of the most frequent disasters over the world. Figure 34.48 shows the percentages of disaster events by each category between 2000 and 2008 from two international disaster databases, EM-DAT[10] and NatCatSERVICE.[11] There are some differences between these two databases because of the difference of event registrations, but anyway, floods share highest or second highest disaster of natural disasters.

Remote sensing has been used to estimate inundated areas by floods. This is done by comparing two images taken before the flood and after the flood. Both optical sensors and SARs are used for this purpose. Figure 34.49 shows flood areas caused by a cyclone over Myanmar taken by PALSAR (▶ Chap. 1) on ALOS (▶ Chap. 1). Figure 34.50 shows a flood over northeastern China taken by GLI (▶ Chap. 1) on ADEOS2 (▶ Chap. 1).

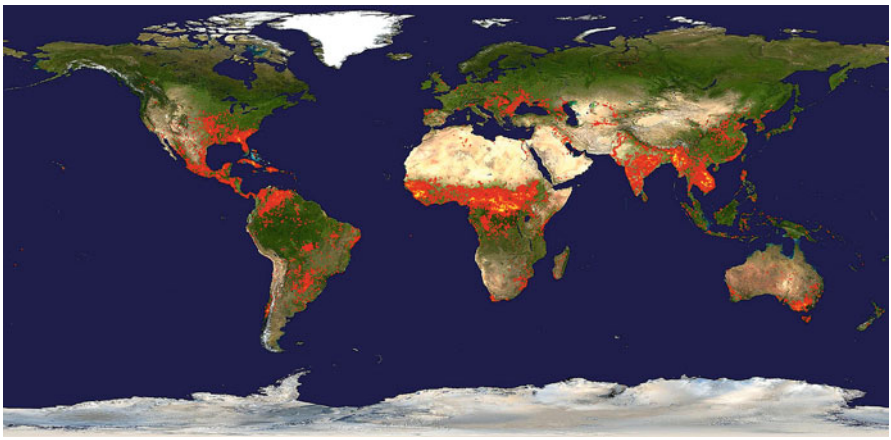


Fig. 34.47 Global fire map of 03/12/11–03/21/11[126] (Courtesy of NASA)

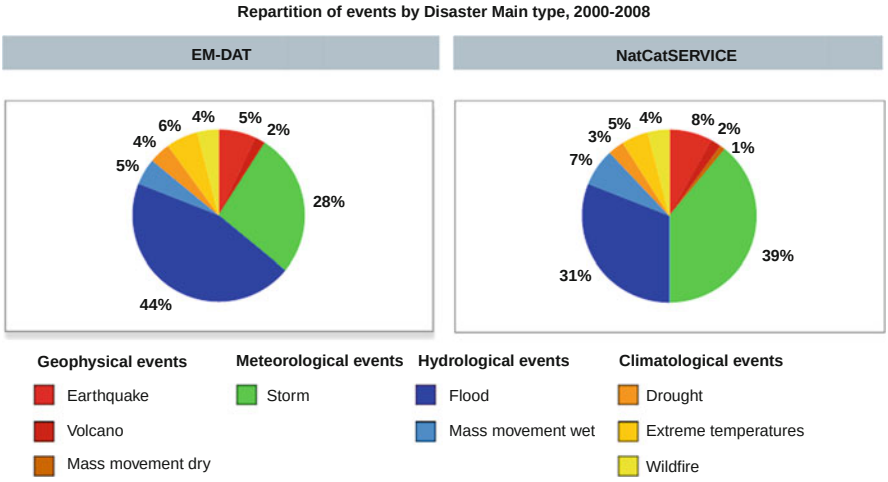


Fig. 34.48 Events by natural disasters main types, 2000–2008 (Regina Below, Angelika Wirtz, Debarati GUHA-SAPIR, 2009, Disaster category classification and peril terminology for operational purposes, p. 9, http://cred.be/sites/default/files/DisCatClass_264.pdf)

Optical sensors can detect inundated areas clearly, but it cannot take images under cloudy conditions. SAR can take images in any conditions, but sometimes it is difficult to extract inundated areas.

Ship Navigation

Ship routing and navigation in Arctic sea areas was one of the earliest operational applications of SAR. One of the disadvantages of SAR for near-real-time

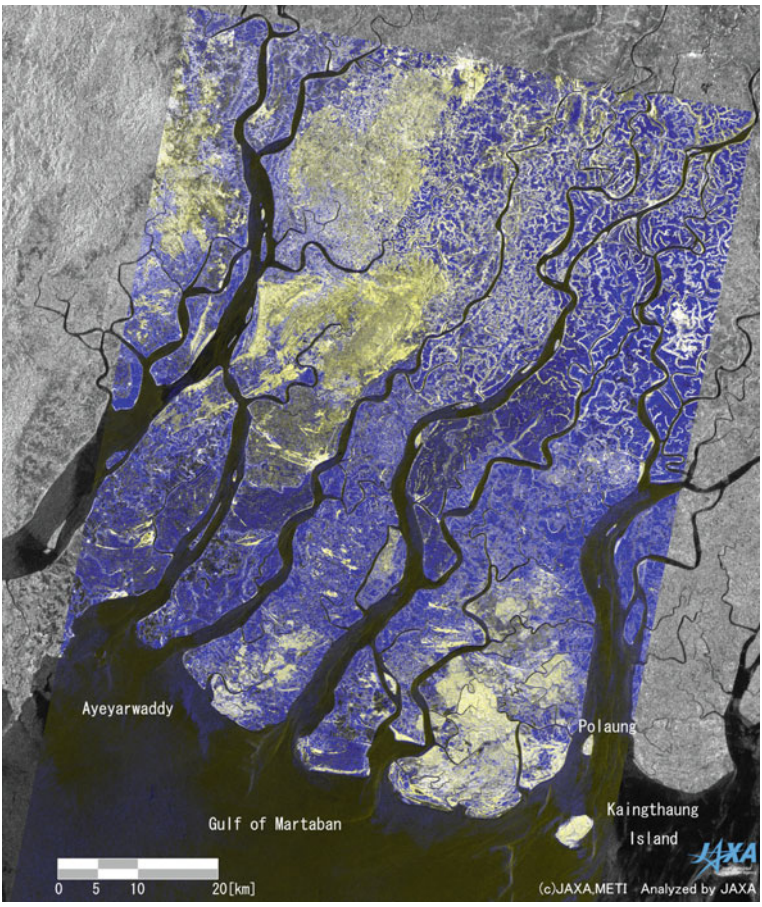


Fig. 34.49 Images of Ayeyarwaddy, Myanmar, observed by PALSAR on April 24 and May 6, 2008 (Myanmar flood water observation by PALSAR, http://www.eorc.jaxa.jp/en/imgdata/topics/2008/tp080509.html). *Blue* color shows the inundated areas and *yellow* color shows high soil moisture areas (Courtesy of JAXA)

applications is the frequency of the observations. However, in high-latitude regions, i.e., higher than 68°, scan mode SAR can observe any areas in this region once a day. C-band SAR is thought to be most useful for this application, and there are now more than four sensors operating. C-band SAR can distinguish multi-year ice, first year ice, landfast ice, thin ice, leads/polynyas, and areas of ridges.

Figures 34.51 and 34.52 show examples of ice classification and ship routing map generated by Canadian Center for Remote Sensing (CCRS) and Canadian Ice Service, respectively.

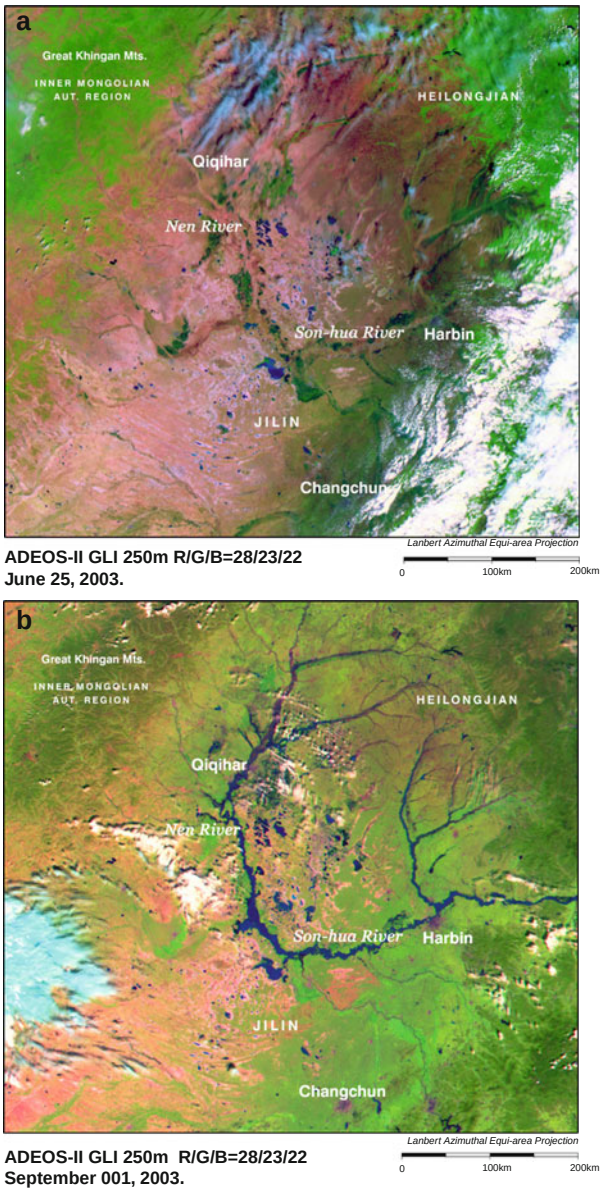


Fig. 34.50 GLI captured the conditions before and during the flooding in Northeastern China that continued from July to October 2003. (Northeastern China suffers large-scale flooding, http://www.eorc.jaxa.jp/en/imgdata/topics/2003/tp031112.html) (**a**) Before the flood, (**b**) after the flood (Courtesy of JAXA)

Agriculture

Agriculture is one of the largest application areas of remote sensing. There are several applications for agriculture, but largest applications are crop acreage estimation and crop yield estimation.

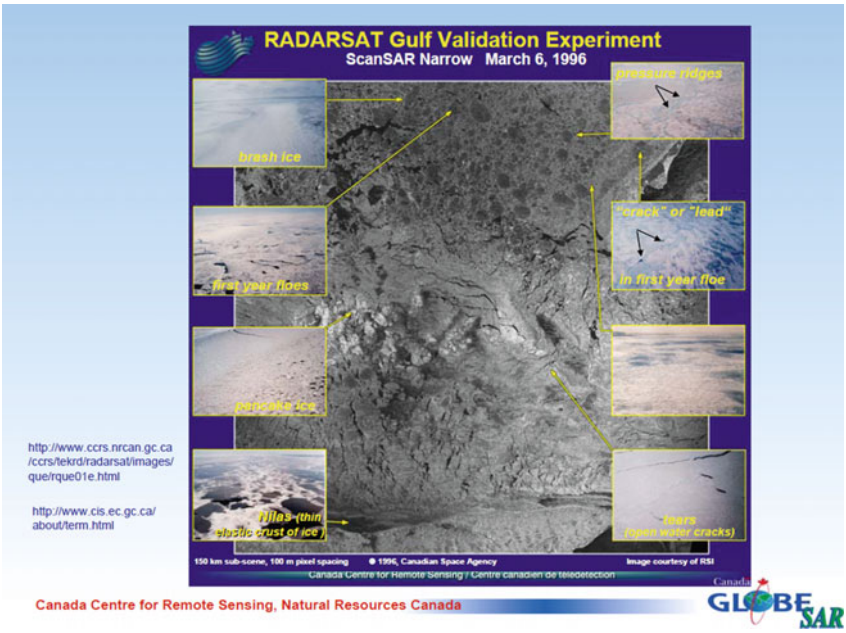


Fig. 34.51 Sea ice classification from Radarsat 1 data made by CCRS (Ice concentration (classification) maps, http://www.ccrs.nrcan.gc.ca/resource/tutor/gsarcd/pdf/ap_ice_e.pdf) (Courtesy of CCRS)

Crop acreage estimation started in USA in the 1970s. National Agricultural Statistics Service (NASS) of USDA has started state-level crop acreage estimation using Landsat (▶ Chap. 1) imagery since 1978 (NASS 2009). Now, many countries are using remote sensing for crop acreage estimation. However, there are several problems of using remote sensing for crop acreage estimation on a global scale. The first problem comes from the spatial resolution. For countries like USA or Canada, the dimensions of each crop field is very large, hence spatial resolution of 30 m of Landsat TM is sufficient for these countries. However, for Japan and most southeast Asian countries, these dimensions are very small, and 30 m resolution cannot discriminate each crop field.

Second problem is the cloud cover. Optical sensors cannot observe under clouds. This problem is most typical for rice fields, most of which resides in Monsoon areas. SAR can penetrate clouds, but its ability to discriminate crop species is very low.

The third problem is the timing which NASS emphasizes (Nass 2009). The discriminability of remote sensing to crop species is highest when crop grows sufficiently, but most statistics need acreage estimation in earlier stages.

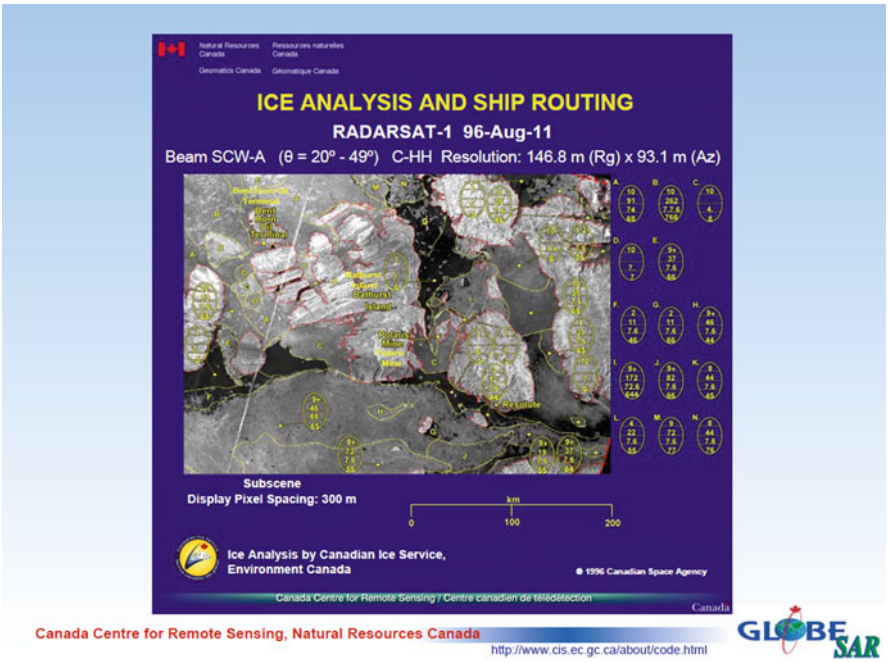


Fig. 34.52 Ship routing map generated by Canadian Ice Service using Radarsat 1 data (Route planning for ships in ice, http://www.ccrs.nrcan.gc.ca/resource/tutor/gsarcd/pdf/ap_ice_e.pdf) (Courtesy of Canadian Ice Service)

The accuracy of crop acreage estimation is rather difficult to estimate. Workshop on Best Practices for Crop Area Estimation with Remote Sensing Data[12] was held in 2008 under GEO, and each country reports the accuracy of their crop acreage estimate by remote sensing. The accuracies range from 60% to 95%, but the real best accuracy will be in the range of around 85%.

Figure 34.53 shows a part of land cover map of State of Illinois using Landsat TM (► Chap. 1) data generated by a NASS project, and Table 34.4 shows the results of validation. From Table 34.4, it is shown that high classification accuracies are obtained for some crops (e.g., corn and soybeans are around 98%), while classification accuracies are low for other crops (e.g., rice, barley, rye, oats, etc., are around 50%).

Use of DMC (► Chap. 1) satellite which can provide 30 m resolution images everyday may improve these problems.

The second application area is the crop yield estimation. Usually, crop yield estimation is done using regression models with climate variables, like temperature, precipitation, etc. In addition to these variables, addition of parameters derived from

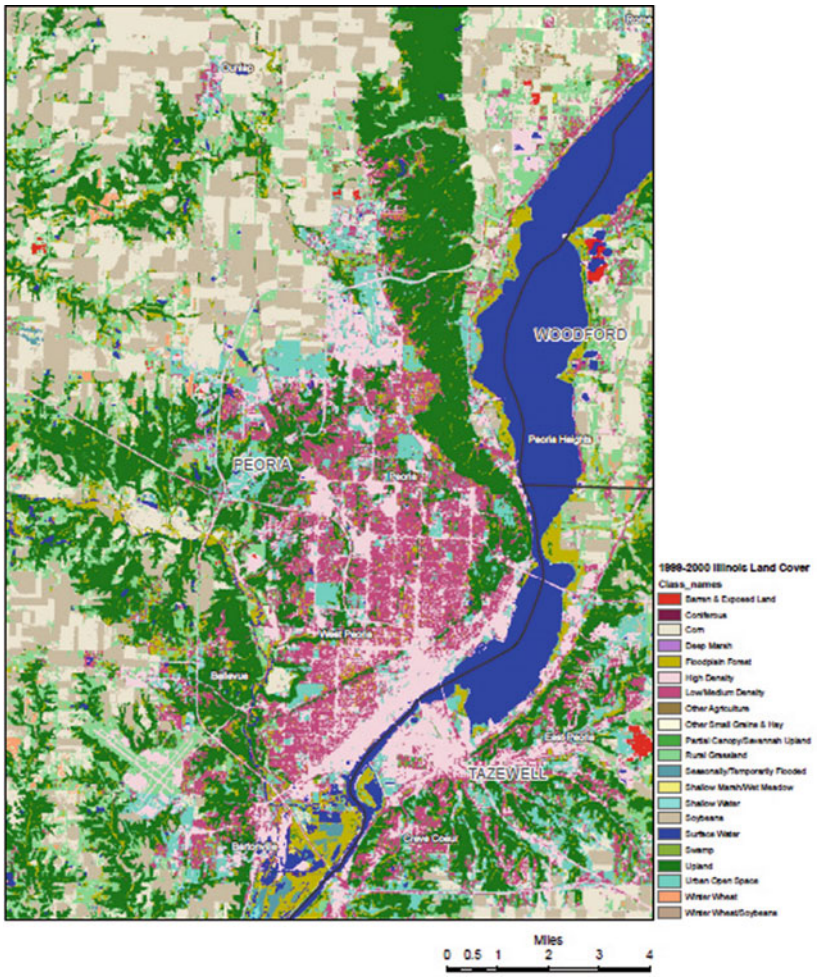


Fig. 34.53 This shows a part of land cover map of State of Illinois using Landsat TM data generated by a NASS project (Luman 2007)

remotely sensed data, e.g., NDVI, EVI, LAI, etc., usually gives better results. Only one problem is the timing of remote sensing data acquisition. For instance, in order to estimate rice yield, there are three timings each of which has only one week duration. These timings also depend on the kinds of rice, and the areas of rice fields. So, it is very difficult to acquire appropriate remote sensing images which can be used for yield estimations.

Table 34.4 Results of validation of the total Illinois land cover (Luman 2007, P. 16)

Individual land cover category	Value	Producer's accuracy	Omission error	Cond'l kappa	User's accuracy	Commission error	Cond'l kappa
Corn	*1*	*98.7%*	1.3%	*0.96*	*99.2%*	0.8%	*0.98*
Rice	3	88.5%	11.5%	0.89	41.4%	58.6%	0.41
Sorghum	4	78.8%	21.2%	0.79	46.8%	53.2%	0.47
Soybeans	*5*	*96.9%*	3.1%	*0.95*	*98.7%*	1.3%	*0.98*
Sunflowers	6	27.3%	72.7%	0.27	2.6%	97.4%	0.03
Tobacco	11	0.0%	100.0%	0.00	0.0%	100.0%	0.00
Barley	21	60.7	39.3%	0.61	21.3%	78.8%	0.21
Spring wheat	23	97.1%	2.9%	0.97	50.8%	49.2%	0.51
Winter wheat	24	81.0%	19.0%	0.81	79.6%	20.4%	0.79
Other grains	25	0.0%	100.0%	0.00	0.0%	100.0%	0.00
Winter wheat/ soybeans	*26*	*87.6%*	*12.4%*	*0.87*	*89.8%*	*10.2%*	*0.90*
Rye	27	74.0%	26.0%	0.74	30.3%	69.7%	0.30
Oats	28	46.6%	53.4%	0.47	52.4%	47.6%	0.52
Alfafa	36	64.4%	35.6%	0.64	84.6%	15.4%	0.85
Dry beans	42	87.0%	13.0%	0.87	62.2%	37.8%	0.62
Potatoes	43	65.9%	34.1%	0.66	50.3%	49.7%	0.50
Other crops	44	52.5%	47.5%	0.53	62.2%	37.8%	0.62
Misc. vegetables	47	89.4%	10.6%	0.89	64.5%	35.5%	0.64
Peas	53	69.1%	30.9%	0.69	50.4%	49.6%	0.50
Clover/ wildflowers	58	60.3%	39.7%	0.60	62.5%	37.5%	0.62
Idle/fallow	61	41.4%	58.6%	0.41	29.8%	70.2%	0.30
Open water	*111*	*89.0%*	*11.0%*	*0.89*	*20.8%*	*79.2%*	*0.21*
Developed/open space	*121*	*83.3%*	*16.7%*	*0.83*	*1.7%*	*98.3%*	*0.02*
Developed/low intensity	*122*	*93.5%*	*6.5%*	*0.94*	*24.9%*	*75.1%*	*0.25*
Developed/ medium intensity	*123*	*90.9%*	*9.1%*	*0.91*	*81.4%*	*18.6%*	*0.81*
Developed/high intensity	124	83.5%	16.5%	0.84	93.0%	7.0%	0.93
Barren	131	86.2%	13.8%	0.86	48.7%	51.3%	0.49
Deciduous forest	*141 (63, 143)*	*97.5%*	*2.5%*	*0.97*	*6.3%*	*93.7%*	*0.06*
Evergreen forest	142	98.6%	1.4%	0.99	100.0%	0.0%	1.00
Grassland herbaceous	*171 (62, 152, 181)*	*86.7%*	*13.3%*	*0.87*	*0.2%*	*99.8%*	*0.00*

(*continued*)

Table 34.4 (continued)

Individual land cover category	Value	Producer's accuracy	Omission error	Cond'l kappa	User's accuracy	Commission error	Cond'l kappa
Woody wetlands	*190*	*82.5%*	*17.5%*	*0.82*	*61.1%*	*38.9%*	*0.61*
Herbaceous wetlands	195 (87)	82.6%	17.4%	0.83	4.1%	95.9%	0.04

Table 5a of the combined summary accuracy assessment report of the 2007 Illinois CDL agricultural and nonagricultural land cover categories. Land cover categories accounting for 1% or more of the state's surface area are italicized

Conclusions

There are many other application areas, e.g., urban planning, archeology, water resources, etc., which are not described in this chapter. However, the application areas of remote sensing are spreading rapidly thanks to the new sensors as well as many remote sensing satellites. For global change monitorings, a long-term record is necessary. There are some long-term records starting from 1960s (NOAA satellites) and 1970s (microwave radiometer), but most of sensors for this purpose started from the end of 1990s, and still need further continuous monitorings. For local applications, high-spatial-resolution sensors made new applications. Problems in this field are cost of image acquisition and frequency of observations.

Cross-References

► Introduction and History of Space Remote Sensing

Notes

1. AMSR/AMSR-E SST algorithm, http://sharaku.eorc.jaxa.jp/AMSR/doc/alg/9_alg.pdf
2. Shuttle radar topography mission, http://www2.jpl.nasa.gov/srtm/
3. GLAS/ICESat L1 and L2 global altimetry data, http://nsidc.org/data/docs/daac/glas_icesat_l1_l2_global_altimetry.gd.html
4. MOD 09 – surface reflectance; Atmospheric correction algorithm products, http://www.atmos-chem-phys.net/9/9619/2009/acp-9-9619-2009.html
5. C. Schaaf, Recent developments in the MODIS albedo, Nadir BRDFAdjusted reflectance (NBAR) and reflectance anisotropy products (MCD43), http://modis.gsfc.nasa.gov/sci_team/meetings/201001/presentations/land/schaaf.pdf
6. A. Strahler, MODIS land cover product algorithm theoretical basis document (ATBD) version 5.0, http://modis.gsfc.nasa.gov/data/atbd/atbd_mod12.pdf
7. GLOBCOVER, Products description and validation report, http://due.esrin.esa.int/files/p68/GLOBCOVER_Products_Description_Validation_Report_I2.1.pdf
8. MODIS rapid response system global fire maps, http://rapidfire.sci.gsfc.nasa.gov/firemaps/#FireLocationData
9. Global fire map, http://www.fire.uni-freiburg.de/current/globalfire.htm
10. EM-DAT, The international disaster database, http://www.emdat.be/

11. Munich RE, NETCATSERVICE, http://www.munichre.com/en/reinsurance/business/non-life/georisks/natcatservice/default.aspx
12. http://www.earthobservations.org/documents/cop/ag_gams/200707_01/Summary_countries_AG.pdf

References

J.R. Anderson, E.E. Hardy, J.T. Roach, R.E. Witmer, *A land use and land cover classification system for use with remote sensor data*, Geological survey professional paper 964 (1976)

D. Antoine, A. Morel, Ralative importance of multiple scattering by air molecules and aerosols in forming the atmospheric path radiance in the visible and near-infrared parts of the spectrum. Appl. Opt. **37**, 2245–2259 (1998)

I. Asanuma, Depth and time resolved primary productivity model examined for optical properties of water, in *Global Climate Change and Response of Carbon Cycle in the Equatorial Pacific and Indian Oeeans and Adjacent Landmasses*, Elsevier Oceanogrhy Series, **73**, 89–106 (2006)

ASTER Level 4A01 data products specification (GDS version) version 1.1, ERSDAC (2002)

M.J. Behrenfeld, E. Boss, D.A. Siegel, D.M. Shea, Carbon-based ocean productivity and phytoplankton physiology from space. Global Biogeochem. Cy. **19**, GB1006 (2005)

M.J. Berenfeld, P.G. Falkowski, Photosynthetic rates derived from satellite-based chlorophyll concentration. Limnol. Oceanogra **42**, 1–20 (1997)

A. Berk, S.L. Bernstein, G.P. Anderson et al., MODTRAN cloud and multiple scattering upgrades with application to AVIRIS. Remote Sens. Environ. (Elsevier) **65**(3), 367–375 (1998)

T. Blaschke, G.J. Hay, Object-oriented image analysis and scale-space: theory and methods for modeling and evaluating multi-scale landscape structures. Int. Arch. Photogramm. Remote Sens. **34**(4/W5), 22–29 (2001)

S.A. Buehler, P. Eriksson, T. Kuhn, A. von Engeln, C. Verdes, ARTS, the Atmospheric Radiative Transfer Simulator. J. Quant. Spectrosc. Radiat. Transf. **91**(1), 65–93 (2005)

M.-E. Carr, M.A. Friedrichs et al., A comparison of global estimates of marine primary production from ocean color. Deep Sea Res. **53**, 741–770 (2006)

S.P. Chavez Jr., An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. Remote Sens. Environ. **24**, 459–479 (1988)

R.M. Chomkoa, H.R. Gordon, S. Maritorenab, D.A. Siegel, Simultaneous retrieval of oceanic and atmospheric parameters for ocean color imagery by spectral optimization: a validation. Remote Sens.Environ. **84**, 208–220 (2003)

S.A. Clough, M.W. Shephard, E.J. Mlawer, J.S. Delamere, M.J. Iacono, K. Cady-Pereira, S. Boukabara, P.D. Brown, Atmospheric radiative transfer modeling: a summary of the AER codes. J. Quant. Spectrosc. Radiat. Transfer. **91**, 233–244 (2005)

D.P. Craig, T. Thirunamachandran, Third-body mediation of resonance coupling between identical molecules. Chem. Phys. **135**(1), 37–48 (1989)

P.Y. Deschamps, M. Herman, D. Tanre, Modeling of the atmospheric effects and its application to the remote sensing of ocean color. Appl. Opt. **22**, 3751–3758 (1983)

G. Esser, J. Hoffstadt, F. Mack, U. Wittenberg, *High-Resolution Biosphere Model (HRBM)-Documentation Model Version 3.00.00*, vol. 2, ed. by G. Esser (Mitteilungen aus dem Institut fur Pflanzenokolgie der Justus-Liebig-Universitat Giessen, Giessen, 1994), p. 70

D.R. Fatland, C.S. Lingle, Analysis of the 1993–95 Bering Glacier (Alaska) surge using differential SAR interferometry. J. Glaciol. **44**, 532–546 (1998)

J.A. Foley, I.C. Prentice, N. Ramankutty, S. Levis, D. Pollard, S. Sitch, A. Haxeltine, An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics. Global Biogeochem. Cy. **10**, 603–628 (1996)

K. Fukue, M. Maeda, H. Shimoda, Continental scalse land cover classification using modis surface reflectance products. Int. Arch. Photogram. Remote Sens. Spatial Inform. Sci. **XXXVIII**(Part 8), 953–957 (2010)

B.C. Gao, M.J. Montes, R.-R. Li, H.M. Dierssen, C.O. Davis, An atmospheric correction algorithm for remote sensing of bright coastal waters using MODIS Land and Ocean Channels in the solar spectral region. TGARS **45**, 1835–1843 (2007)

R.M. Goldstein, H.A. Zebker, C.L. Werner, Satellite radar interferometry: two-dimensional phase unwrapping. Radio Sci **23**(4), 713–720 (1988)

H.R. Gordon, D.K. Clark, Clear water radiances for atmospheric correction of coastal zone color scanner imagery. Appl. Opt. **20**, 4175–4180 (1981)

H.R. Gordon, K.J. Voss, MODIS normalized water-leaving radiance. Algorithm theoretical basis document (MOD18), NASA (2004)

H.R. Gordon, M. Wang, Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm. Appl. Opt. **33**, 443–452 (1994)

A.W. Gruen, Adaptive least squares correlation: a powerful image matching technique. S Afr J Photogramm. Remote Sens. Cartogr. **14**, 175–187 (1985)

History of remote sensing for crop acreage, NASS (2009), http://www.nass.usda.gov/Surveys/Remotely_Sensed_Data_Crop_Acreage/index.asp

P. Hofmann, Detecting urban features from IKONOS data using an object-oriented approach, in *RSPS2001 Proceedings*, pp. 79–91

J. Ishizaka, E. Siswanto, T. Itoh, H. Murakami, Y. Yamaguchi, N. Horimoto, T. Ishimaru, S. Hashimoto, T. Saino, Verification of vertically generalized production model and estimation of primary production in the Sagami Bay, Japan. J. Oceanogr. **63**(3), 517–524 (2007)

J.D. Jackson, *Classical electrodynamics*, 3rd edn. (Wiley, New York, 1962). American association of physics teachers

M.Z. Jacobson, The short-term cooling but long-term global warming due to biomass burning. J. Clim. **17**, 2909–2925 (2004)

N. Jacquinet-Husson, V. Capelle, L. Crépeau, R. Armante, N.A. Scott, A. Chédin, The GEISA/IASI spectroscopic database in its 2008 edition. ISSWG2-2 IASI Sounding Science Working Group, Darmstadt, Allemagne, 21–22 Apr 2009

Z. Jelenak, T. Mavor, L. Connor, N.-Y. Wang, P.S. Chang, P. Gaiser, Validation of ocean wind vector retrievals from WindSat polarimetric measurements. Presented at the 4th international Asian-Pacific environmental remote sensing conference, Honolulu (2004)

W.L. Jones, L.C. Schroeder, F.J. Wentz, Microwave scatterometer measurements of oceanic wind vector, Oceanography from space, in *Proceedings of the symposium*, Venice, Italy; United States, 26–30 May 1980. (1981), pp. 553–562

I.R. Joughin, R. Kwok, M.A. Fahnestock, Interferometric estimation of the threedimensional ice-flow velocity vector using ascending and descending passes. IEEE Trans. Geosci. Remote Sens. **36**, 25–37 (1998)

Y.J. Kaufman, The atmospheric effect on remote sensing and its correction, in *Theory and applications of optical remote sensing*ed. by G. Asrar (Wiley, New York, 1989), pp. 336–428

G. Kelly, The relative contributions of the various space observing system, in *Proceedings of 2007 EUMETSAT Meteoroligical Satellite Conference* (2007)

Y.S. Kotchenova, F.E. Vermote, R. Matarrese, F.J. Klemm Jr., Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part I: path radiance. Appl. Opt. **45**(26), 6762–6774 (2006)

E.R. Kursinski, G.A. Hajj, J.T. Schofield, R.P. Linfield, K.R. Hardy, Observing the earth's atmosphere with radio occultation measurements using the global positioning system. J. Geophys. Res. **102**(D19), 23429–23465 (1997)

G.S.E. Lagerloef, Introduction to the special section: The role of surface salinity on upper ocean dynamics, air sea interaction and climate. J. Geophys. Res. **107**(C12), 8000 (2002). doi:10.1029/2002JC001669

G. Lagerloef, R. Schmitt, Role of ocean salinity in climate and near-future satellite measurements: Meeting Report. EOS, Trans. Am. Geophys. Un. **87**(43), 466–467 (2006)

J.S. Levine, W.S. Cofer, D.R. Cahoon Jr., E.L. Winstead, The global impact of biomass burning. Environ Sci Technol, 1995. http://asd-www.larc.nasa.gov/biomass_burn/globe_impact.html

D. Luman, T. Tweddale, Assessment and potential of the 2007 USDA-NASS Cropland data layer for statewide annual land cover applications, INHS technical report 2008 (49), p. 21

S.N. Madsen, H.A. Zebker, *Imaging Radar Interferometry. Principles & Applications of Imaging Radar*, 3rd edn. Manual of Remote Sensing, vol. 2 (Wiley, New York, 1998)

E.P. McClain, W.G. Pichel, C.C. Walton, Comparative performance of AVHRR-based multichannel sea surface temperatures. J. Geophys. Res. **90**, 11587–11601 (1985)

T. Meissner, F. Wentz, An updated analysis of the ocean surface wind direction signal in passive microwave brightness temperature. IEEE Trans. Geosci. Remote Sens. **40**(6), 1230–1240 (2002)

W.G. Melbourne, et al., *The application of spaceborne GPS to atmospheric limb sounding and global change monitoring*. (JPL Publication 94–18, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, 1994), p. 147

G. Mie, Ann. Physik **25**(4), 377 (1908)

G. Mie, Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. Leipzig, Ann. Phys. **330**, 377–445 (1908)

J.J. Mohr, N. Reeh, S.N. Madsen, Three dimensional glacial flow and surface elevation measured with radar interferometry. Nature **391**, 273–276 (1998)

J.L. Monteith, Solar radiation and productivity in tropical ecosystems. J. Appl. Ecol. **9**, 747–766 (1972)

F. Naderi, M.H. Freilich, D.G. Long, Spaceborne radar measurement of wind velocity over the ocean – an overview of the NSCAT scatterometer system. Proc. IEEE, **9**(6) 850–866 (1991)

T. Nakajima, et al., EORC bulletin, Technical report No.15, 360 p., ISSN 1346–7913, JAXA/EORC, GSS reference handbook (RSTAR reference handbook) (2004)

T.Y. Nakajima, A. Higurashi, T. Nakajima, S. Fukuda, S. Katagiri, Development of the cloud and aerosol retrieval algorithms for ADEOS-II/GLI mission. J. Remote Sens. Soc. Jpn **29**, 60–69 (2009)

R.R. Nemani, C.D. Keeling, H. Hashimoto, W.M. Jolly, S.C. Piper, C.J. Tucker, R.B. Myneni, S.W. Running, Climate-driven increases in global terrestrial net primary production from 1982 to 1999. Science **300**, 1560–1563 (2003)

M. Neubert, Segment-based analysis of high resolution satellite and laser scanning data, in sustainability in the information society, in *Proceedings of the 15th International Symposium Informatics for Environmental Protection*, Zurich, 10–12 Oct 2001, by eds. L.M. Hilty, P.W. Gilgen (Mtropolis, Marburg, 2001), pp. 379–386

Y.-J. Park, K. Ruddick, Model of remote-sensing reflectance including bidirectional effects for case 1 and case 2 waters. Appl. Opt. **44**, 1236–1249 (2005)

R.T. Rabus, D.R. Fatland, Comparison of SAR-interferometric and surveyed velocities on a mountain glacier: Black Rapids Glacier Alaska U.S.A. J. Glaciol. **46**, 119–128 (2000)

L. Rayleigh (J.W. Strutt), On the light from the sky, its polarization and colour. Philos. Mag. **41**, 107–120, 274–279 (1871)

C.D. Rodgers, Inverse method for atmospheric sounding (World Scientific publishing, New York, 2000)

L.S. Rothman, I.E. Gordon, A. Barbe et al., The HITRAN 2008 molecular spectroscopic database. J. Quant. Spectrosc. Radiat. Transf. **110**, 533–572 (2009)

D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning internal representations by error propagation, in *Parallel distributed processing*, ed. by J.L. McCleland, D.E. Rumelhart, The PDP Research group, vol. 1 (MIT Press, Cambridge, MA, 1986a)

D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors. Nature **323–9**, 633–536 (1986b)

S.W. Running, S.T. Gower, FOREST-BGC, a general model of forest ecosystem processes for regional applications. II. Dynamic carbon allocation and nitrogen budgets. Tree Physiol. **9**, 147–160 (1991)

S.W. Running, P.E. Thornton, R. Nemani, J.M. Glassy, Global terrestrial gross and net primary productivity from the Earth Observing System, in Methods in ecosystem science, ed. by O. Sala, R. Jackson, H. Mooney (Springer, New York, 2000), pp. 44–57

S.W. Running, R.R. Nemani, F.A. Heinsch, M. Zhao, M. Reeves, H. Hashimoto, A continuous satellite–derived measure of global terrestrial primary production. Bioscience **54**, 547–560 (2004)

C. Schaaf, A. Strahler, et al., MODIS reflectance albedo and reflectance anisotropy, in Proceedings of land and vegetation direct readout workshop (2007)

B. Schölkopf, A. Smola, K.-R. Mülller, Nonlinear component analysis as a Kernel Eigenvalue Problem. Technical Report No. 44. Max-Planck-Institut für biologische Kybernetik, Tübingen (18 p, 239 K). The final version of this article is published in Neural Comput. **10**(5), 1299–1319 (1996), (MIT Press, 1998)

D.A. Siegel, M. Wang, S. Maritorena, W. Robinson, Atmospheric correction of satellite ocean color imagery: the balack pixel assumption. Appl. Opt. **39**, 3582–3590 (2000)

S. Sitch, The role of vegetation dynamics in the control of atmospheric CO_2 content. Ph. D. thesis, Lund University, Sweden, (2000) p. 213

A. Smola, Regression estimation with support vector learning machines. Master's thesis, Technische Universität München (1996)

M. Takagi, H. Shimoda (eds.), *Handbook of image analysis*(in Japanese) (University of Tokyo Press, Tokyo, 2004), p. 1327

M. Takagi, H. Shimoda (eds.) *Handbook of Image Analysis*(revised ed.) (in Japanese), (University of Tokyo Press, Tokyo, 2004), p.198

H. Tian, J.M. Melillo, D.W. Kicklighter, A.D. McGuire, J. Helfrich, The sensitivity of terrestrial carbon storage to historical climate variability and atmospheric CO_2 in the United States. Tellus **51B**, 414–452 (1999)

US CLIVAR Office, Report of the US CLIVAR Salinity Working Group. US CLIVAR Report 2007–1, Washington, DC, 46 p (2007), http://www.usclivar.org/Pubs/Salinity_final_report. Accessed 27 Jan 2008

V. Vapnik, The nature of statistical learning theory (Springer, New York, 1995)

E.F. Vermote, N. El Saleoql et al., Atmospheric correction of visible to middle-infrared EOS-MODIS data over land surfaces: Background, operational algorithm and validation. JGR **102** (D14), 17131–17141 (1997)

T. Westberry, M.J. Behrenfeld, D.A. Siegel, E. Boss, Carbon-based primary productivity modeling with vertically resolved photoacclimation. Global Biogeochem Cy. **22**, GB2024 (2008)

H.A. Zebker, R.M. Goldstein, Topographic mapping from interferometric SAR observations. J. Geophys. Res **91**, 4993–4999 (1986)

Geographic Information Systems and Geomatics

35

Jesus A. Gonzalez

Contents

Introduction 936
GIS Conceptual Framework 937
GIS Sources of Data 938
Vector and Raster Models 939
Georeferencing 940
The Ellipsoid 941
Mean Sea Level (Geoid) 941
Map Projection 944
Metadata 945
Interactions of GIS with Satellite Systems 946
Geographic Information Systems and Remote Sensing 946
Geographic Information Systems and Intelligent Positioning 947
Spatial Data Analysis 948
Applications 949
Examples of Current Trends 952
Cross-References 953
References 953

Abstract

The role of spatial data for decision making has increased the need for geographic information systems. This chapter starts by briefly describing the theory of geographic information systems. After that we present the interactions of geographic information systems with remote sensing and global navigation,

J.A. Gonzalez
National Institute of Astrophysics, Optics, and Electronics (INAOE)/Regional Center for Space Science and Technology Education for Latin America and the Caribbean (CRECTEALC), Campus Mexico, Calle Luis Enrique Erro No. 1, Tonantzintla, Puebla, 72840, Mexico
e-mail: jagonzalez@inaoep.mx

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_48, © Springer Science+Business Media New York 2013

positioning, and timing satellite systems. This is done with the idea to illustrate how geographic information systems (GIS), remote sensing, and global navigation satellite systems work together in order to generate final products. Then, we focus on the capability of GIS to analyze spatial data. We finally present examples of applications of GIS and current trends of research in the area. The examples presented through this chapter capture how GIS can be used for decision-making tasks in different areas of knowledge.

Keywords

Coordinate system • Data mining • Datum • Ellipsoid • Geographic information systems • Geoid • Geomatics • Georeferencing • Global Positioning System • Location based service • Map projection • Metadata • Pattern • Remote sensing • Spatial data analysis • Spatial database • Vector and raster models

Introduction

A great part of data that we generate can be related to a location on Earth. The concept of an object's location could be as simple as the place where it was generated or the place where an object can be found. The purpose for keeping track of the location of data varies depending on the application and may change in time. For instance, the location of an object might not be static; it could be dynamic, as in the case of an automobile in motion. This issue can be as complicated as the scope of the functionality of our application.

The inherent relation of geographic information to spatial dimensions attaches it to a location on Earth, with reference to a coordinate system. For this purpose, we require a representation of Earth in order to assign a specific location to an object. An option for this representation is a sphere with a radius of approximately 6,400 km and coordinates that are measured as latitude and longitude. Latitude is measured as the distance from the equator to a point to the north or the south of it. Longitude is measured as the distance from the (commonly) Greenwich meridian to the east or west. Height is another variable to take into account although more difficult to measure with high precision. A variation of this representation is the use of an ellipsoid instead of a sphere in order to obtain a closer representation of Earth for our geographic coordinate system.

Geographic data is produced from different sources. It can be identified by its spatial component. It may come from a source located in space, such as a remote sensing satellite or an airplane taking an aerial image. It could be data coming from a GPS satellite which we use to calculate our location. It could even be data from a sensor on Earth such as an electronic total station, as the tool of a civil engineer. An electronic total station is an electronic distance measurement (EDM) instrument that can be seen as the modern theodolite used to measure angles and directions, among others. No matter what the source of the data is, if it is being acquired, it is

important to store it. Database management systems have evolved to make room for spatial data. Nowadays, these systems are able not only to store this type of data but also allow users to formulate queries that evaluate objects on the basis of spatial relations. In this way, we can find important information that is useful for decision making in different applications.

The evolution of spatial databases did not stop with these facilities. It is here where Geographic Information Systems (GIS) were developed to create, visualize, and manipulate spatial data. A GIS is composed of a spatial database, a graphic user interface, and a set of tools to manipulate spatial data. Furthermore, many GIS are created to work in the web environment so that multiple users are able to obtain the benefits of a web application.

Once the GIS has been built and data has been collected, there is a treasure to be exploited, resulting in valuable information that can be used for decision making. There are different ways to analyze the data and information stored on GIS. One of them consists of overlaying of layers of data or information, as a way of organizing different types of data (as we will see in the following sections). Another way could be with the use of spatial queries. One more could be the specific processing of the data, as in the case of the creation of a network flow model. One more could be a data mining analysis.

Geomatics is defined as the field of study related to the gathering, storing, processing, management of spatially georeferenced data and delivering of geographic information. These topics, among others, will be covered in the following sections. The rest of the chapter describes with more detail the conceptual framework of a GIS, its interactions with remote sensing data, some of its applications, the role of GIS in decision making, future trends of GIS, and conclusions.

GIS Conceptual Framework

In the previous section, we were introduced to the broad picture of GIS. Now we will concentrate in more detail on important concepts that support the theory of this multidisciplinary area. Let us start with by defining a GIS as a collection of components necessary to store spatial data to be manipulated in order to create spatial products (see Fig. 35.1). Then, the components of a GIS can be classified in three main categories. The first component is computer hardware to store data and software. The second component is computer software, required to manipulate data and create valuable products. Finally, the last and the most valuable component is geographic data. In order to manage geographic data, database management systems (DBMS) have been extended to deal with the spatial component. There are both commercial and open source tools with such an extension. An example of an open source DBMS with such an extension is PostgreSQL, with its spatial component called postGIS. For more information about postGIS please refer to http://postgis.refractions.net/.

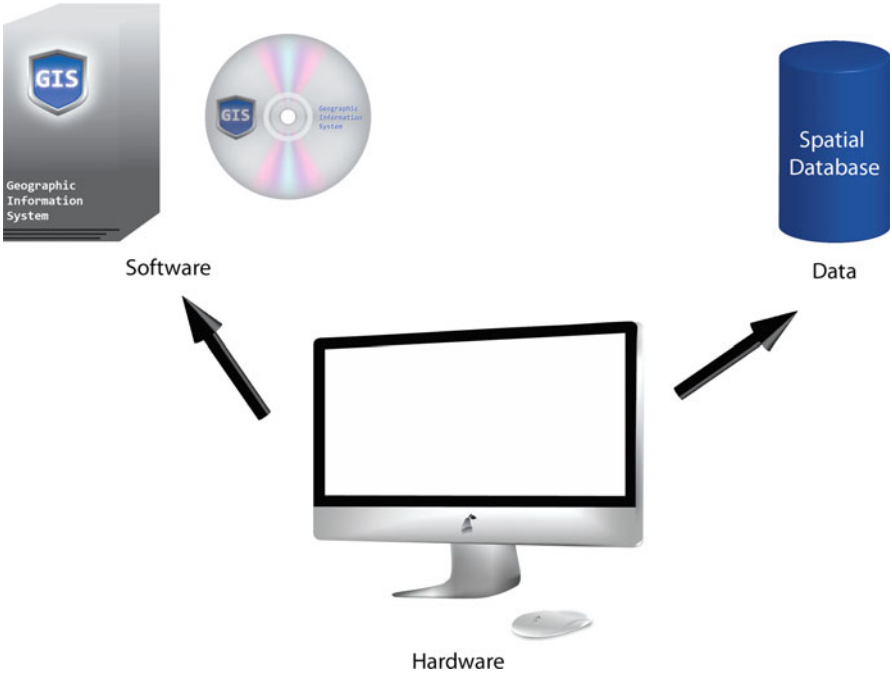


Fig. 35.1 Main components of a geographic information system: computer hardware, software, and geographic data. Computer hardware is the data and software repository. Software is used to manipulate geographic data and to create valuable products used for decision making. Data is the heart of a GIS, its prime matter

GIS Sources of Data

There are different sources of data that we can store in a GIS, see Fig. 35.2. Much of the information was already available when GIS came into the market. This was the case of paper maps that were digitized and introduced to them. Data obtained from field work is another important source of data because although it is not an efficient way to obtain it, sometimes it is the only way to get data on a specific variable. Satellite and aerial images are yet another source of data for GIS. This type of data requires different preprocessing steps, depending on the purpose of the GIS. As we discussed before, information coming from Global Navigation Satellite Systems, such as GPS, can also be stored in a GIS (Lee 2001a). These are some of the sources of data for a GIS but there might be more. Which other data depends on the spatial components of our GIS. Once we have identified these spatial components, we ask ourselves the question, what is the source of any other information that we require? The answer is directly related to the problem that our GIS application is intended to solve.

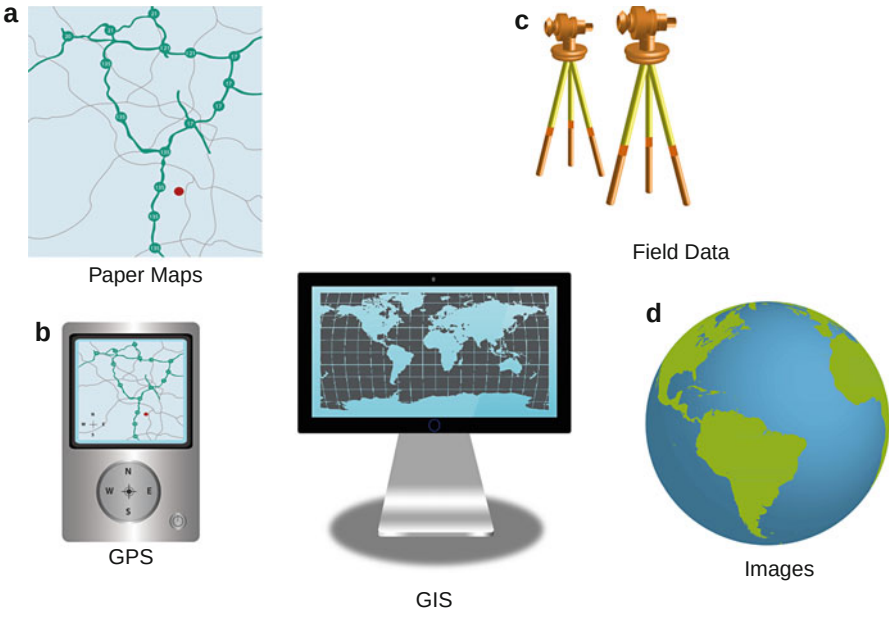


Fig. 35.2 Some of the sources of data of a geographic information system. (**a**) Paper maps, (**b**) data coming from GPS satellites, (**c**) data coming from field work, and (**d**) data coming from satellite and aerial images

Vector and Raster Models

In order to introduce data into a GIS, we need to know how it is organized. There are two models to store data in a GIS. They are known as the vector and raster models. The goal is to model the real world, to represent it in a level of abstraction with the adequate level of detail to be mapped to the GIS. We might not completely store every detail of reality but only as much as we require. Let us assume for now that we are capturing this data from satellite images, but as we know, we could also obtain it from measures in the field or even from other ways.

In the vector model, each object in the real world is represented with one of three possibilities. These are a point, a line, or a polygon. Figure 35.3 shows an example of a road, a school, and a parcel represented with a line, a point, and a polygon, respectively. An important characteristic of this model is that the spatial relations between different objects can be captured. These are known as topological, metric, and direction relations. More information about the topic can be found in Kopersky and Han (1999).

In the case of the raster model, data is represented with a grid of data. Each cell in the grid corresponds to a pixel in the image and is classified as a type of object. In the previous example, the line representing the road fills the points in the grid that

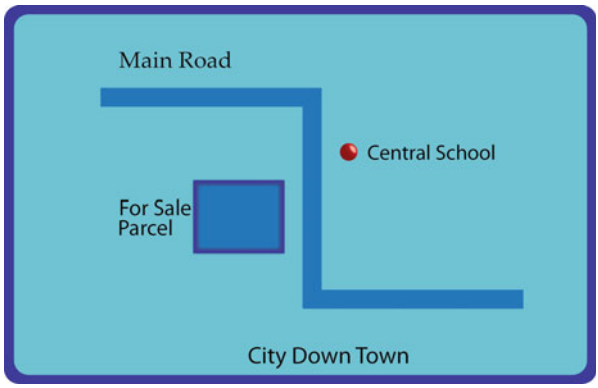


Fig. 35.3 GIS Vector Model. Partial representation of the downtown of a city. A parcel for sale is represented with a polygon, main road with a poly-line, and central school with a point

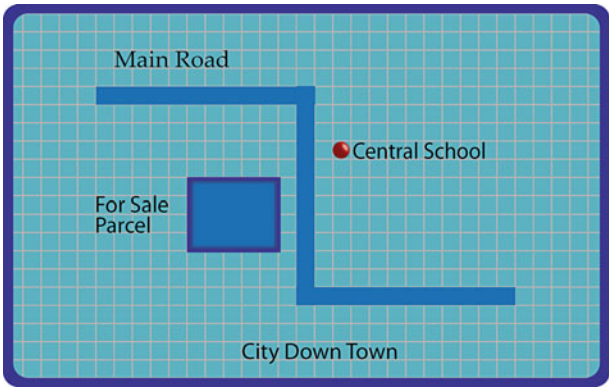


Fig. 35.4 GIS Raster Model. Partial representation of the downtown of the city modeled with the GIS Vector Model of Fig. 35.3 with the Raster Model. In this case, the parcel for sale can be identified by the cells filled in blue. Main road can be identified by the cells filled as well in blue. Finally, central school corresponds to the cell filled in *red*

overlap with the road. The school is represented with one pixel and the parcel for sale is represented by a set of pixels (an area) as can be seen in Fig. 35.4.

Up to now we know how data is structured to be stored in a GIS. However, we are missing one of the fundamental concepts that make a GIS so useful: its capability to manage locations. We need to be able to attach location to the objects that we store. For this, we require the use of a georeference system as we describe in the following section.

Georeferencing

When we load an image in a GIS, we want to identify the location of the objects in it. In order to do it, we need to georeference the image. We do this by finding a correspondence of the locations and a map projection using a coordinate system.

For this, we introduce some concepts about Earth. We are first concerned with the measurement of Earth (Gelati 2006). There is a division of science known as geodesy or geodetics that does this. The important pieces of geodesy that concern GIS are the reference Earth shape or *ellipsoid* geodetic positioning or *geodetic datum* and *coordinates* the true Earth shape or *geoid* and *vertical datum*, and the practical representation of the Earth or *map projections*.

The Ellipsoid

We use an ellipsoid as a good approximation to represent the shape of the Earth. This ellipse rotates around one of its axis. There have been different proposals of ellipsoids. One of the most common is the WGS84, which parameters are an equatorial axis of 6,378,137.00 m, a polar axis of 6,356,752.3142 m, and an inverse flattening of 298.257223563. As we know, the ellipsoid does not accurately represent the shape of the Earth. There are mountains that are higher than the line of the ellipse and there also exist places below sea level, which are below the line of the ellipse. This proves that this is not an accurate representation of Earth.

Mean Sea Level (Geoid)

As we can note, the shape of the Earth is very irregular. There is not any geometric body that has the exact shape of the Earth. This is the reason why the Earth's shape received the name of the *Geoid*. The Geoid is an irregular equipotential surface that coincides with mean sea level over the oceans. It has also an imaginary continuity across the continents, which have undulations on the surface (the topography) because of the irregular distribution of the gravitational mass forces of the planet. The geoid is used as a reference surface for leveling, that is, we measure elevation relative to the geoid (Li and Götze 2001). A more detailed description about the geoid concept can be found in (Mok and Chao 2001). Figure 35.5 illustrates how the geoid differentiates from the ellipsoid.

A *geodetic datum* is defined as a reference model that associates a geodetic reference ellipsoid (the ellipsoid parameters: equatorial axis, polar axis, and inverse flattening) to a coordinate system (defined by a geodetic space through orientation, position, and scale). A geodetic datum is a mathematical model of Earth (Gelati 2006). There are two types of datums. These can be either geocentric or local. A *geocentric datum* is globally centered and is a good approximation of the whole Earth. In this case, the center of the reference ellipsoid coincides with the Earth's center of mass (Gelati 2006). In Fig. 35.6 we can see how the center of the ellipsoid and the center of mass of the Earth coincide. We can also see that the geoid (or mean sea level) and the reference ellipsoid are in general a good approximation.

A geocentric datum is best suited for global applications, just as GPS uses the WGS84 geocentric datum. In contrast, a *local geodetic datum* better suits a particular region where the reference ellipsoid has better adjustment with the

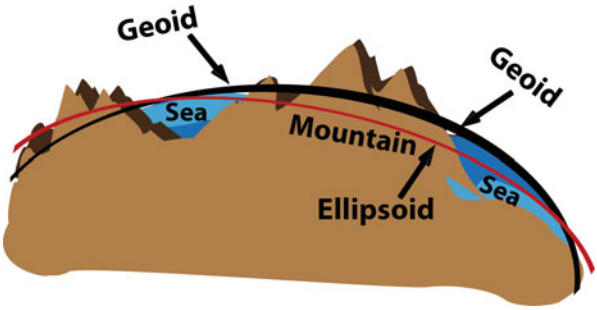


Fig. 35.5 Difference between the geoid and the ellipsoid. The ellipsoid is a smooth geometric shape. The geoid passes over or under the ellipsoid depending on the irregular distribution of the gravitational mass forces of Earth. We can also appreciate how the topography of Earth differs from that of the geoid and the ellipsoid

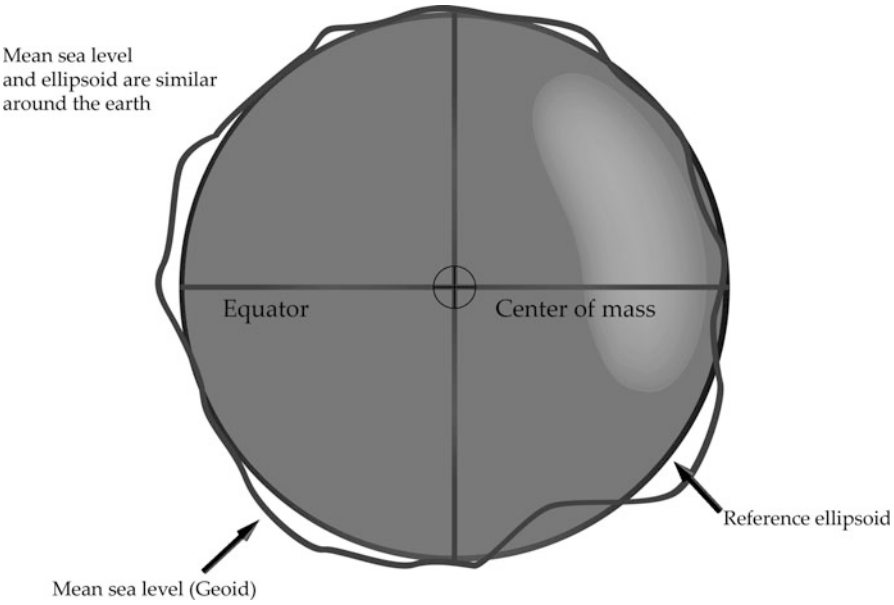


Fig. 35.6 The Geocentric Datum. In a Geocentric Datum, the center of mass coincides with the center of the reference ellipsoid. It is a good approximation of the whole Earth but there can be more accurate approximations for particular regions. This figure was adapted from Fig. 7.1 in Gelati (2006)

Earth's shape. In this case, the center of the ellipsoid does not always coincide with the Earth's center of mass. Because of this, a local geodetic datum does not provide a good global representation of the Earth. Figure 35.7 shows how the ellipsoid's adjustment to the Earth is better in a local region than in the rest of it.

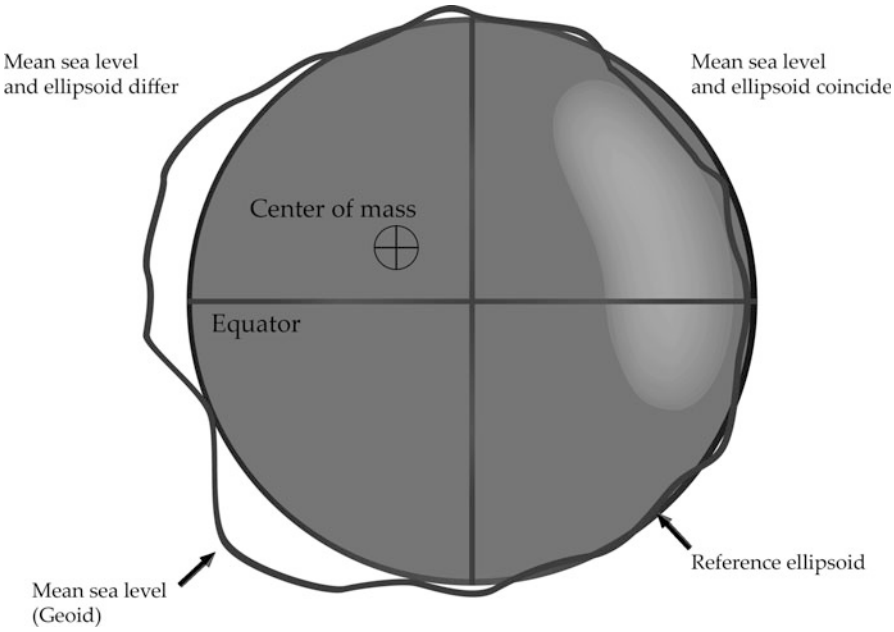


Fig. 35.7 The Local Geodetic Datum. In a Local Geodetic Datum, the center of mass does not coincide with the center of the reference ellipsoid. It is a good approximation of a region of the Earth, but it is not for the whole Earth. We can see this in the figure because the geoid and the ellipsoid have better adjustment in a particular region of Earth than the rest of it. This figure was adapted from Fig. 7.2 in Gelati (2006)

Once we described the ellipsoid, the geoid, and the datum we will describe the geographic coordinate system, which is based on the Earth's rotation around its center of mass. We can determine the geographical coordinates of any point on the Earth's surface based on its latitude and longitude. The Earth's center of mass is on its rotation axis. The plane that passes through the center of mass and perpendicular to the axis defines the *Equator*. Latitude is defined as the angle from the meridian between the equator and the reference parallel. This will always be north (N) or south (S). Then, the maximum latitude will be of 90°. Longitude is a geographic coordinate that defines the east (E) or west (W) position of a point on the Earth's surface. It is the angle measured east or west between the plane containing the Prime Meridian (Greenwich) and a plane containing the North Pole, the South Pole, and the location in question (Araujo 2006; Longley et al. 2005) (Fig. 35.8).

Even when an ellipsoidal representation of Earth has the advantage of being realistic, it has some disadvantages. For example, it is impossible to observe the entire terrestrial surface at the same time; we can only see one of its faces but not the other. This does not happen with a map. In a map, we can show the whole world. The ellipsoidal representation is not easy to manage as a map is. We cannot change the scale of the terrestrial globe in a practical way as we do with our paper map. Of course,

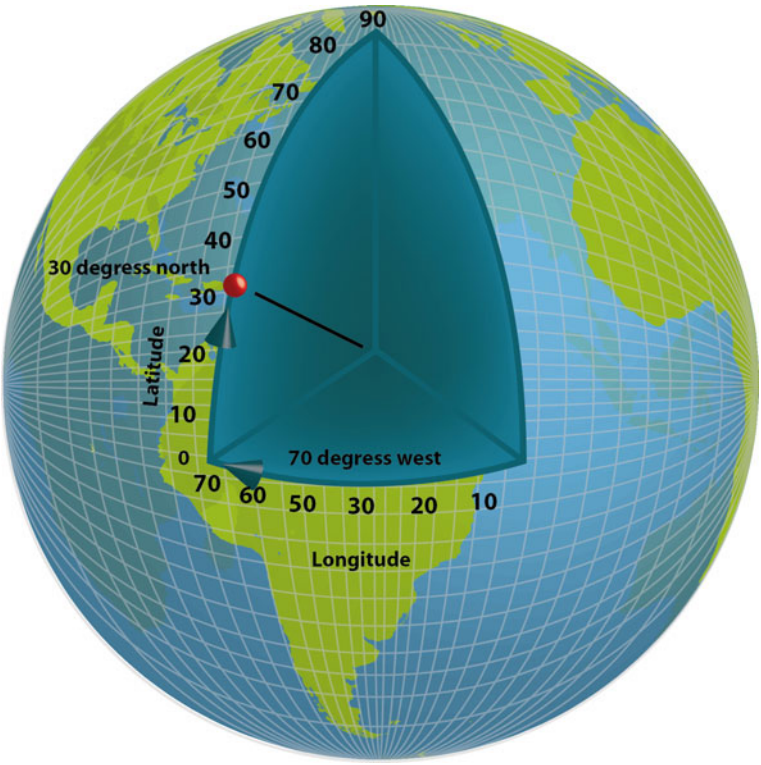


Fig. 35.8 Geographic coordinates system. Latitude is measured north or south from the equator while longitude is measured east or west from the Greenwich meridian. Figure adapted from (http://services.arcgisonline.com/arcgisexplorer500/help/latlong_from_globe_center.png)

there are many disadvantages of paper maps that we do not have with a terrestrial globe such as the geometric deformations suffered during the *map projections* that were used to create the maps. Now we will describe what a map projection is.

Map Projection

A map projection is defined as a mathematical transformation between the geographic coordinate system (in latitude and longitude) and a system on the plane surface. There are two common planar coordinate systems. One of them is the *Cartesian system* (with X, Y coordinates) and the other is the *polar system* (with range and angle coordinates). The problem with this transformation is that three main types of distortions are introduced. These are length, area, and angular distortions. This means that length, area, and angles cannot be preserved by a single map projection at the same time. As an example, a length distortion

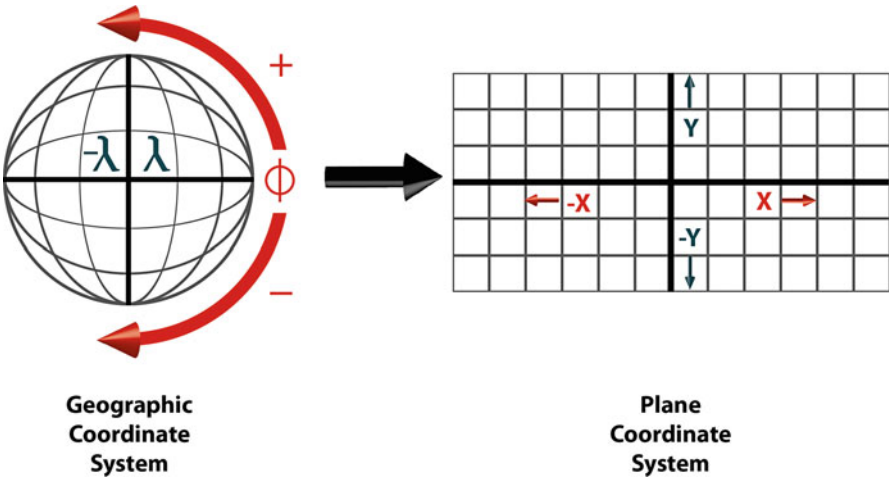


Fig. 35.9 Map projection. The geographic coordinate system using latitude (ϕ) and longitude (λ) is projected into a plane coordinate system using X and Y coordinates; in this case, a Cartesian coordinate system. This figure was adapted from the online course of geography, lesson 7: A deeper understanding of coordinate systems and projections (https://www.e-education.psu.edu/geog486/l7_p9.html)

means that length measured on a map does not correspond to the length of the same feature measured in the real world. This is the distortion introduced by the map projection. That is, the use of plane geometry and trigonometry involving Cartesian coordinates to perform the calculations does not lead to correct results after the map projection. There is also an error when we measure angles in the map. For more information about map projections, please refer to (Lee 2001b) (Fig. 35.9).

Until now, we have studied the basics of GIS. We know what a GIS is, its sources of data, and the vector and raster models. We are able to recognize spatial data, and we learnt the difficulties involved with approximating the shape of the Earth. We can also find the coordinates of a location on Earth. Now, we will learn about the interactions between remote sensing and geographic information systems as a way to improve each other's capabilities.

Metadata

Metadata is commonly referred as data about the data. This is the information that we use to document our data. In this way, metadata describes all the parameters necessary to work with spatial data: the data owner, source, resolution, and scale. A metadata framework can be described in different formats such as ASCII, HTML, Extensible Markup Language (XML), Standard Generalized Markup Language (SGML), and Resource Description Framework (RDF). In order to create compatibility among geospatial products and tools, a great effort to create standards

over geospatial data has been done. Some of the available standards for geospatial metadata are (Gelati 2006; ISO 2011):

- ISO 19139 Geographic Information Metadata XML Schema Implementation
- ISO 19115 Geographic Information Metadata
- Content Standard for Digital Geospatial Metadata
- Dublin Core Metadata Element Set
- Australian Government Locator Service
- UK GEMINI Discovery Metadata Standard

Interactions of GIS with Satellite Systems

In this section, we describe in more detail how GIS interacts with remote sensing and the global positioning system platforms. The integration of these interactions has made possible what we have today and most of what we are creating for the future.

Geographic Information Systems and Remote Sensing

Remote Sensing (RS) and GIS are two areas that interact with each other. There are three main ways in which these interactions can be combined to enhance each other (G. Wilkinson 1996). In the first one, RS is used as a tool to obtain data to be used in a GIS. Second, GIS data can be used as auxiliary information to improve products created from RS sources. Finally, RS and GIS are usually used together for modeling and analysis processes (Weng 2010).

RS contributes to the information that is stored in a GIS in different ways. One of the most important contributions is the extraction of thematic information from satellite images to create GIS layers. RS images are used to extract cartographic information to be the input to GIS, as in the case of the production of base maps. A very important application that requires the use of RS is the update of GIS databases. In this case, RS images are used to detect changes in thematic information to update GIS databases. RS images have also been used as background for GIS representations. This is the case of visualization tools for digital elevation models, which are very important for different applications (Weng 2010).

On the other hand, GIS data is used to improve some of the processes used in RS. These processes are, among others, the selection of the area of interest, its preprocessing, or its classification (Weng 2010). It is of great interest how GIS context information can be used to post-process the classification results of a statistical RS classification algorithm to improve its accuracy (Gonzalez et al. 2008). Another interesting approach is the use of a structural data representation (i.e., a graph-based representation) in order to use both types of information at the same time (nonspatial and spatial). In this way, the classification algorithm takes advantage of all the available information at the moment that it is performing the classification task (Pech et al. 2004). Figure 35.10 shows two patterns found through the Subdue system (Cook and Holder 1994), a graph-based spatial data

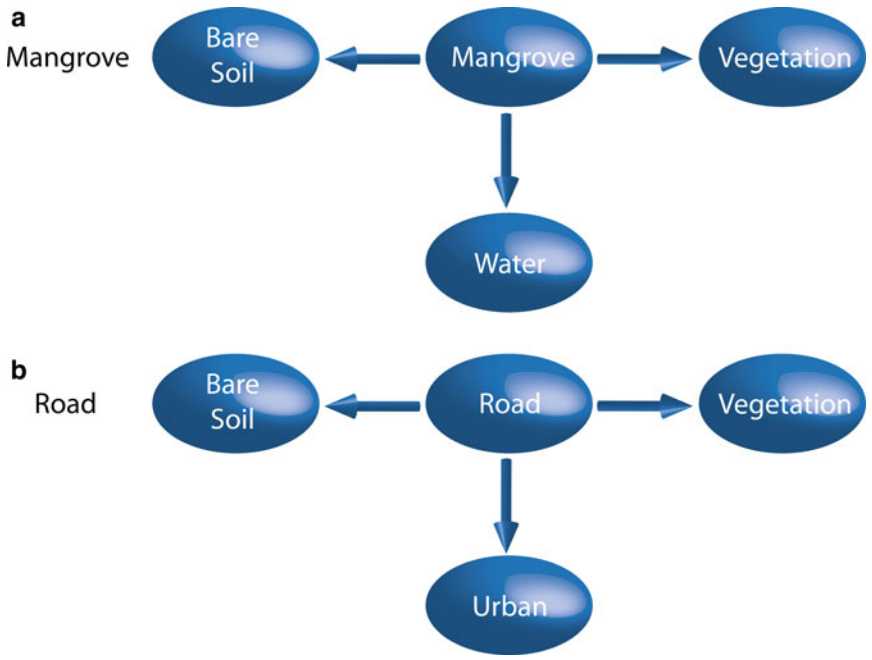


Fig. 35.10 Two patterns found by a graph-based spatial mining algorithm. (**a**) A graph pattern describing the Mangrove class. This pattern says that a mangrove is usually found adjacent to regions of the classes bare soil, vegetation, and water. (**b**) A pattern describing the road class. This pattern says that a road is usually found adjacent to regions of the classes bare soil, vegetation, and urban

mining process. Pattern (a) corresponds to the description of the class "Mangrove" and tells us that, in general, a region of interest (ROI) that belongs to this class is adjacent to other regions of the classes "Bare Soil," "Vegetation," and "Water." Pattern (b) describes the class "Road" and tells us that, in general, a ROI that belongs to this class is adjacent to other regions of the classes "Bare Soil," "Vegetation," and "Urban." This information is used in a post-processing step to validate the class assigned by a statistical classification algorithm in order to improve its classification accuracy.

Geographic Information Systems and Intelligent Positioning

GIS has also a mature level of integration with positioning systems, as in the case of the Global Positioning System (GPS) of the USA. In this case, we can say that there are four main levels of integration. In the first one, a GIS only takes the information reported by GPS and displays it in a map. In a second level, there are more functions. The GIS can manage WGS84 coordinates and different layers of the map (i.e., boundaries, counties, roads, rivers, and more). It is also possible to zoom

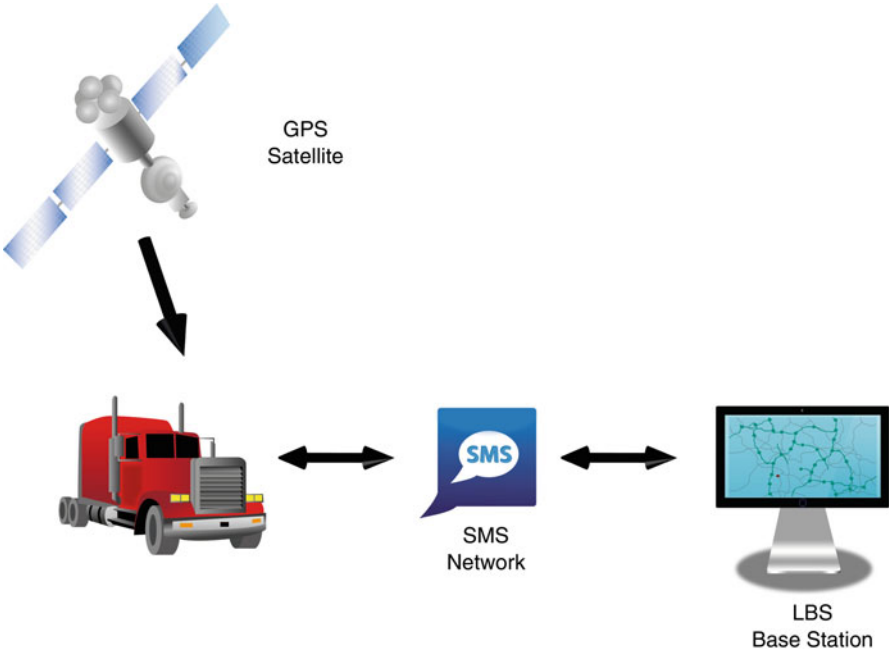


Fig. 35.11 GPS satellites updating a location-based service application. The satellites update the position of vehicles being monitored by the base station. The base station receives the position via SMS messages. The base station sends useful information (depending on their location: addresses of gas stations, restaurants, hotels, etc.) to the vehicles via SMS messages

in and out to take a look to a specific location. In a third level of integration, it allows entering waypoints (the coordinates of important reference points) describing interesting features. This allows the GPS–GIS system to create a GIS database that can be used to make a map. In the last level of integration, the map, an intelligent map, is associated to a set of logical rules that are used to improve the accuracy of the reported position (Taylor and Blewitt 2006). Figure 35.11 depicts a set of satellites sending positioning information to vehicles.

Spatial Data Analysis

The organization of spatial data in spatial databases is a plus that does not only facilitate the way a GIS accesses information but also provides users powerful analysis tools. The extension of a relational database into a relational spatial database requires adding geometry information to the spatial objects stored in it. This includes the coordinates that define both the shape of the objects (as points, lines, or polygons) and their coordinates in space. This information is commonly stored in a table related to another table that stores the nonspatial information

describing the spatial object. Indexes over these tables are created in order to access the data in an efficient way.

A GIS connected to a spatial database provides different tools to analyze the data stored in such a database. It allows organizing data of the same type (i.e., roads are represented by lines, trees by points, parcels by polygons, county divisions with polygons, a satellite image of each county with polygons) in layers. Because of this, a form of visual analysis consists of the overlaying of several layers to allow the user to identify how different features of distinct layers interact in space. We can give transparency to any of the layers (i.e., a satellite image) so that we can appreciate important features with more detail. In this level of spatial data analysis the user interacts with the GIS to create a useful product, name it a map, that can be used for decision making.

Another level of spatial data analysis is known as Spatial Data Mining (SDM). In this case, data is usually extracted from the GIS or spatial database and transformed into a data representation that can be managed by the spatial data mining system. There are different spatial data mining tasks: clustering (Ng and Han 1994), spatial association rules (Koperski et al. 1996), colocation patterns (Xiong et al. 2004), and outliers detection (Shekhar et al. 2002), among others. Some of the more interesting data representations (and useful for spatial data) are those able to deal with structural data, such as Inductive Logic Programming (Muggleton 1995) and Graph-based Learning (Cook and Holder 1994). Spatial clustering methods find patterns that share a spatial component. Spatial association rules try to associate spatial objects to neighboring objects. An approach known as co-location patterns states that we usually find in a nearby region instances of a set of spatial features. That is, when a subset of such spatial features are commonly located together (in a nearby region), it can be considered a co-location pattern.

In Fig. 35.12 we show the integration of a GIS with a graph-based data mining tool. The GIS loads the spatial data stored in a postGIS spatial database and presents the base map located at the center of the interface. The GIS allows the user to analyze the data by presenting the spatial layers contained in the spatial database (the option to perform this function is located in the upper left of the graphic user interface, see Fig. 35.12). It is also possible to perform spatial queries using topological, distance, and direction relations (as we can see to the right of Fig. 35.12). The interface has an option to transform the queried data into its graph-based representation in order to send it as input to a graph-based data mining system, for instance, the option called Subdue. Subdue performs the data mining task and finds spatial patterns. The instances of the patterns found can then be visualized in the main map so that the domain expert can interpret the mined results. This interface integrates a GIS system with a set of tools for decision making.

Applications

The high capability of GISs to store spatial data, process it, analyze it, and create final products such as thematic maps makes them a powerful tool to apply to any

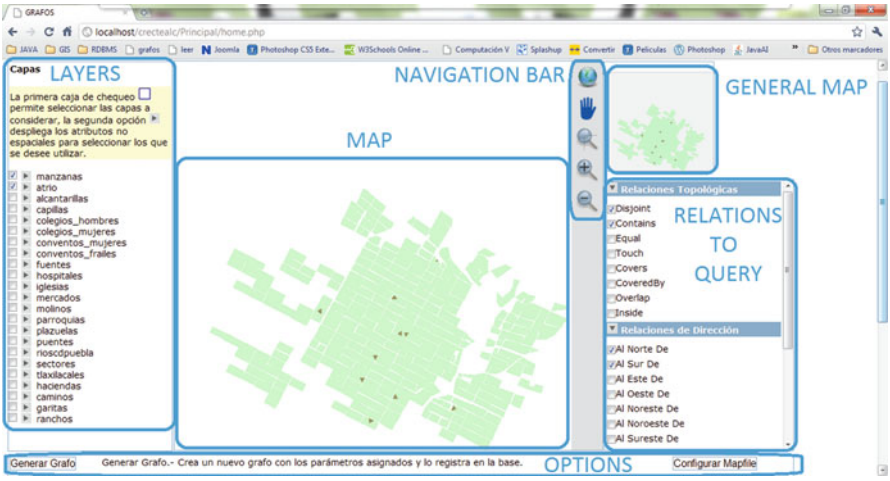


Fig. 35.12 A GIS-data mining graphical user interface. This is a GIS graphical user interface created to analyze data from the City of Puebla, in Mexico (This is the reason why the labels are written in Spanish). The GIS has three main components. The first component allows the user to analyze data overlaying layers. In the second level, the user can perform spatial queries. In a third level, the user can perform the graph-based spatial data mining task, having the opportunity to visualize the resulting patterns (subgraphs) in the map

field where spatial data plays a role. GIS are used for applications in industry, in government, in health care, in environment protection, and many other areas. In the rest of this section, we briefly describe a couple of applications as examples of GIS applications.

In the area of medicine, GISs are very useful for epidemiology studies. This type of application allows physicians to keep track of how a disease expands geographically. If a different type of treatment is being applied in different counties or states, and the statistics are shown in real time in the GIS, the efficiency of each treatment can be appreciated in real time in the GIS graphical user interface. This application could be used for any type of disease. It could be the swine influenza A (H1) in humans, malaria, aids, cancer, or any other disease.

More dynamical GISs are those that receive signals from different sensors such as GPS, as in the case of navigation consoles for automobiles or Electronic Chart Display and Information System (ECDIS) for vessels. An ECDIS is commonly connected to a GPS, a radar system, a meteorological station, a gyroscope, and other sensors of the vessel. The GIS presents the navigation charts and with the help of the GPS, it plots the position of the vessel. The GIS allows drawing the path that the vessel should follow in the navigation chart. In addition, the radar system communicates with the navigation console (the ECDIS) and transmits the objects that it detects so that they can be plotted in the navigation chart and can be considered as dangerous or being in the middle of the path of the vessel. In this case, the ECDIS should play an alarm so that the vessel's Captain considers a maneuver to avoid the blocking object (perhaps another vessel) (Fig. 35.13).



Fig. 35.13 An electronic chart display and information system. This system is composed of a touch screen monitor (*bottom*) to control the GIS functionality. The top monitor displays the information received through the internal network from all the sensors of the vessel connected to the navigation system

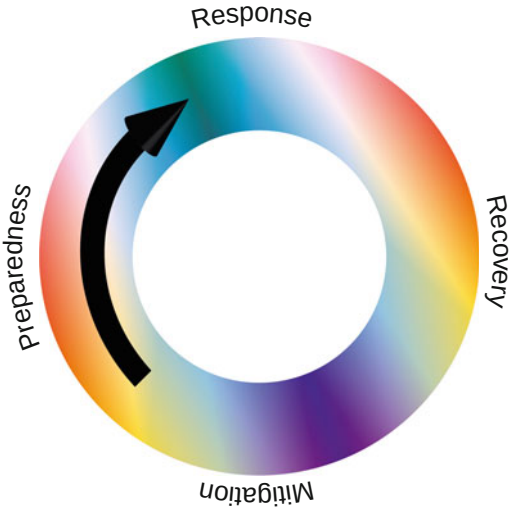


Fig. 35.14 The emergency management cycle. Adapted from Wikipedia

Another important area of application for GIS is that dedicated to disaster management. These GISs tools are created for any of the four phases of disasters: mitigation, preparedness, response, or recovery (UN-SPIDER 2011). Examples of information in GISs in this application may include floods, earthquakes, oil spills, storms, fires, tsunamis, volcano eruptions, epidemics, and droughts, among others. This, being an important area for any government, is an area of opportunity for GIS. For more information about the area of disaster management, please refer to the United Nations Portal of Knowledge at http://www.un-spider.org/knowledge-base. Figure 35.14 shows the emergency response cycle that considers its four phases: preparedness, response, recovery, and mitigation.

Examples of Current Trends

GISs are tools that can be used in any field of study. They are being used to make more efficient processes as part of any industry or government. This multidisciplinary work demands different research areas to meet and innovate. Some of these current topics are the following.

- Augmented reality and GIS are being combined in different applications. The goal is to simulate how the real world would look like if we added artificial objects to it. Examples of augmented reality applications are its integration to landscape visualization (Ghadirian and Bishop 2008). Another example is the use of augmented reality for underground infrastructure visualization (Schall et al. 2009).
- Another important application is the integration of semantic information to the 3D reconstruction of city models as we can see in recent research (Kolbe 2008). In this work, the author uses GML3 to represent the shape, the graphical appearance of the city models, the semantics, the representation of the thematic properties, and the taxonomies of the objects. GML is the Geography Markup Language, an eXtensible Markup Language (XML) grammar created to express geographical features. In (Wolf and Asche 2010) the authors create a 3D tactical intelligence surveillance map for a group of crime experts who study spatiotemporal patterns of residential burglary crimes.
- The integration of artificial intelligence techniques, such as fuzzy theory, is not new but is being more and more useful. An example of such a case can be seen in Kanjilal et al. (2010), in which the authors find an appropriate implementation approach to fuzzy regions.
- These are some examples of both current research areas and applications of GIS that are used to solve real-world problems.

Conclusion

Geographic information systems are an advanced technology that allows developing applications in any area of study. Their power to analyze data enables them to create tools ideal for decision making. The advances in the development of satellite technology as sources of data for GIS enhance the quality of data as well

as its analysis capability. In the current and future years, more research in this area will contribute to the development of more technology to solve more real-world problems, either in the industry, government, academia, and social areas.

Cross-References

- ▶ Digital Image Acquisition: Preprocessing and Data Reduction
- ▶ Digital Image Processing: Post-processing and Data Integration
- ▶ Global Navigation Satellite Systems: Orbital Parameters, Time and Space Reference Systems and Signal Structures
- ▶ Remote Sensing Data Applications

References

D.J. Cook, L.B. Holder, Substructure discovery using minimum description length and background knowledge. J. Artifi. Intell. Res. **1**, 231–255 (1994)

S.R. Gelati, *Geographic Information Systems Demystified* (Artech House, Boston, 2006)

P. Ghadirian, I.D. Bishop, Integration of augmented reality and GIS: a new approach to realistic landscape visualisation. Landsc. Urban Plan. **86**(3–4), 226–232 (2008)

J.A. Gonzalez, L. Altamirano, J.F. Robles, Data mining with context information for satellite image classification. Ambiencia **4**, 147–158 (2008)

ISO, International Organization for Standardization (2011). http://www.iso.org/iso/home.htm. Last visited, 7 Apr 2011

V. Kanjilal, H. Liu, M. Schneider, Plateau regions: an implementation concept for fuzzy regions in spatial databases and GIS, in *Proceedings of the Computational Intelligence for Knowledge-Based Systems Design, and 13th International Conference on Information Processing and Management of Uncertainty* (IPMU'10). (Springer, Berlin/Heidelberg, 2010), pp. 624–633

T.H. Kolbe, *Representing and Enhancing 3D City Models with CityGML*. Lecture Notes in Geoinformation and Cartography (LNGC) (Springer, Berlin/Heidelberg, 2008)

K. Koperski, J. Adhikary, J. Han, Spatial data mining: progress and challenges, in *SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD96)*, Montreal, 1996

A. Koperski, J. Han, Spatial data mining: progress and challenges. Research issues on data mining and knowledge discovery, Montreal, 1999

Y.-C. Lee, Geographical data and its acquisition, in *Geographic Data Acquisition*, ed. by Y.-Q. Chen, Y.-C. Lee (Springer, Wien, 2001a), pp. 1–8

Y.-C. Lee, Map projections, in *Geographic Data Acquisition*, ed. by Y.-Q. Chen, Y.-C. Lee (Springer, Wien, 2001b), pp. 43–63

X. Li, H.-J. Götze, Ellipsoid, geoid, gravity, geodesy, and geophysics. Geophysics **66**, 1660–1668 (2001)

P.A. Longley, M.F. Goodchild, D.J. Maguire, D.W. Rhind, *Geographical Information Systems and Science*, 2nd edn. (Wiley, Chichester, 2005)

E. Mok, J.C.H. Chao, Coordinate systems and datum, in *Geographic data acquisition*, ed. by Y.-Q. Chen, Y.-C. Lee (Springer, Wien, 2001), pp. 11–24

S.H. Muggleton, Inverse entailment and Progol. New Gener. Comput. **13**, 245–286 (1995)

R. T. Ng, J. Han, Efficient and effective clustering methods for spatial data mining, in *Proceedings of the 20th International Conference on Very Large DataBases*, Santiago, 1994, pp. 144–155

A.D. Paulo, *Fundamentals of Cartography*, 3rd edn. UFSC Editorial, (Federal University of Saint Catarina, Brasil, 2006).

M. Pech, A. Tchounikine, R. Laurini, J. Gonzalez, D. Sol, Graph-based representation for spatial data mining: a proposal, in *Proceedings of the CASSINI-SIGMA Conference*, Grenoble, 2004

G. Schall, E. Mendez, E. Kruijff, E. Veas, S. Junghanns, B. Reitinger, D. Schmalstieg, Handheld augmented reality for underground infrastructure visualization. Pers. Ubiquitous Comput. **13**(4), 281–291 (2009)

S. Shekhar, Lu Chang-Tien, P. Zhang, Detecting graph-based spatial outliers, in *Intelligent Data Analysis*, vol. 6 (IOS Press, Amsterdam, 2002), pp. 451–468

G. Taylor, G. Blewitt, *Intelligent Positioning: GIS-GPS Unification* (Wiley, Hoboken, 2006)

UN-SPIDER, United Nations platform for space-based information for disaster management and emergency response, http://www.un-spider.org/. Last visited, 10 Apr 2011

Q. Weng, *Remote Sensing and GIS Integration, Theories, Methods, and Applications* (McGraw-Hill, New York, 2010)

G.G. Wilkinson, A review of current issues in the integration of GIS and remote sensing data. Int. J. Geogr. Inf. Syst. **10**, 85–101 (1996)

M. Wolf, H. Asche, Towards 3D tactical intelligence assessments for crime scene analysis, in *Computational Science and Its Applications: ICCSA 2010: International Conference, Fukuoka, Japan March 23–26, 2010: Proceedings Part I*, ed. by D. Taniar, O. Gervasi, B. Murgante, E. Pardede, B.O. Apduhan. Springer Lecture Notes in Computer Science (LNCS), vol. 6016 (Springer, Berlin, 2010), pp. 346–360

H. Xiong, S. Shashi, H. Yan, K. Vipin, M. Xiaobin, S.Y. Jin, A framework for discovering co-location patterns in data sets with extended spatial objects, in *Proceedings of the 2004 SIAM International Conference on Data Mining (SIAM2004)*, Florida, 2004, pp. 78–89

Section 4

Space Systems for Meteorology

Introduction to Space Systems for Meteorology

36

Joseph N. Pelton, Scott Madry, and Sergio Camacho-Lara

Contents

Introduction 960
Advances in Meteorological Satellite Technology 960
International Cooperation in the Field of Meteorological Satellite Services 961
Development of the Meteorological Satellite Systems 962
Present and Future Coordination of Meteorological Satellite Systems 963
Cross-References 966
Notes 966

Abstract

The world's meteorological satellite systems are today vital to every nation in the world not only for reliable weather forecasts, but also for key storm warnings and potential disaster alerts in the case of hurricanes, tornadoes, typhoons, monsoons, floods, and other violent and potentially lethal meteorological events. Since the advent of polar-orbiting and geosynchronous meteorological satellites, the ability to predict weather accurately and reliably, for ever longer periods of time, has increased to a remarkable extent. With a diverse suite of sophisticated instruments, meteorological satellites also gather essential data for climate change studies. The first systems were pioneered by the United States and by NASA experimental

J.N. Pelton (✉)
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: peltonjoe@gmail.com

S. Madry
International Space University, Chapel Hill, North Carolina, USA
e-mail: Scottmadry@mindspring.com

S. Camacho-Lara
Centro Regional de Enseñanza de Ciencia y Tecnología del Espacio para América Latina y el Caribe (CRECTEALC), Luis Enrique Erro 1, Tonantzintla, Puebla, Mexico
e-mail: sergio.camacho@inaoep.mx

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_52, © Springer Science+Business Media New York 2013

satellites, but over time Europe, Russia, Japan, China, and India have evolved increased capabilities to monitor weather systems using increasingly sophisticated imaging systems and allow effective sharing of meteorological data on a global scale. The combination of polar orbiting and geosynchronous satellites has allowed higher resolution images to be combined into accurate regional and global displays to see broad patterns of weather formations. New capabilities such as lightning intensity mapping have allowed greater capability to predict storm patterns and where the most violent and more energetic storm fronts are headed.

The global sharing of meteorological data and the combined imaging of international meteorological satellites have not only greatly contributed to effective long range weather forecasting on land and in the oceans, but have also allowed the collection of data to monitor longer-term conditions associated with climate change, global warming, and increases in aridity in some regions and increased rainfall in others. In short, meteorological satellites have evolved, particularly in the last decade, to serve an important role in not only short- to medium-term weather forecasting but also to provide important data with regard to national, regional, and global environmental assessment and analysis.

The value of meteorological satellite systems and other Earth observation satellite systems for measuring the internationally recognized Essential Climate Variables (ECVs) and for monitoring changes on land, oceans, and atmosphere is greatly increased when the acquired data is made available to national and international user organizations. A study by the International Academy of Astronautics recommends, among other things, that space agencies, companies, universities and nongovernmental organizations, and other international bodies already acting for the coordination of space agencies in the area of climate monitoring should work together to guarantee over time the continuous operational availability of the space sensors and datasets that are necessary for the monitoring of the space-observable ECVs (International Academy of Astronautics, in *Study on Space Applications in Climate Change and Green Systems: The Need for International Cooperation*, November 2010, eds. J.C. Mankins, M. Grimard, Y. Horikawa, ISBN EAN 9782917761113, pp. 15–21). Such cooperation is aimed at reinforcing the programmatic coordination of the Earth Science programs worldwide, in the frame of institutions such as the Group on Earth Observations (GEO) and the Committee on Earth Observation Satellites (CEOS), with the goal of guaranteeing the continuous long-term availability for all nations of all space-observable ECV, as defined by the Global Climate Observing System (GCOS); and contribute to the elaboration and implementation of GEO data sharing principles (http://www.earthobservations.org/art_015_002.shtml).

Despite patterns of data sharing and international cooperation with regard to data collection by meteorological satellite systems, there are limits to full disclosure of all satellite data. In particular, there remain certain areas of strategic concern in the context of possible military or defense-related use of meteorological and remote sensing data in the case of hostilities. It is partially because of these strategic and national defense-related concerns that so many

different meteorological satellite systems have been deployed and why some parts of the imaging might be encrypted in a manner so that all data that is collected may not be in all instances fully shared. Despite these strategic concerns and interests, most meteorological satellite imaging data is today widely shared and global cooperation is quite universal.

The areas of satellite communications, remote sensing, and satellite navigation have all – to some degree – evolved toward more commercialized economic models and thus have become more oriented to competitive markets. This is not to say that these other space applications services are fully commercialized since there are some well-defined military, defense, and governmental services for satellite communications, remote sensing, and satellite navigation that remain as "publicly provided services." In the case of meteorological satellites, however, these space applications remain entirely as governmental services.

Despite discussions and analysis of how meteorological services might transition to commercial service providers within the United States, the provision of space-based meteorological services seems likely to remain as essentially a "public good" and not commercialized in any space faring nation. Although many countries, private businesses, and individuals derive major benefits from meteorological satellite images, no viable economic model has yet evolved whereby these services might be commercialized.

In the chapters that follow specific information about various national and regional meteorological satellite systems will be presented. This introductory chapter provides a quick overview of the various systems that have evolved over time and some information perspective on how these systems are coordinated and work together through the World Meteorological Organization (WMO) and the World Weather Watch (WWW) program (World Meteorological Organization http://www.wmo.int/pages/index_en.html also see "Lessons Learned about the Integrated Global Observing Strategy through the World Weather Watch" http://www.un.org/earthwatch/about/docs/igusland.htm).

Keywords

Chinese Feng-Yang system • Committee on Earth observation satellites (CEOS) • Data and information service (NESDIS) • Essential climate variables (ECVs) • EumetsatMetOps polar orbiting system • European organisation for the exploitation of meteorological satellites(Eumetsat) • Geostationary lightning mapper (GLM) • Geostationary operational environmental satellites (GOES) • Indian INSAT • Japanese GMS or Himawari system • Japanese MTSAT satellites • Joint polar satellite system (JPSS) • Landsat • METEOR satellites • Meteosat • National Aeronautics and Space Administration (NASA) • National Atmospheric and Oceanic Administration • National environmental satellite • NIMBUS • Polar orbiting operational environmental satellites (POES) • Radiometer • Russian geostationary operational meteorological satellite (GOMS) system • TIROS system • U.S. Defense Weather Satellite Service • World Meteorological Organization (WMO) • World weather watch (WWW)

Introduction

The first application satellites were communication satellites that quickly expanded to fulfill a strong demand – especially for international and overseas telecommunications. The advent of meteorological satellites quickly followed. The TIROS satellites, developed and launched by NASA, quickly demonstrated that meteorological satellites could be used to detect weather patterns and to forecast weather much more accurately. This was followed by the Nimbus Landsat and GOES satellite systems deployed by the United States to capture more and more detailed and up-to-date weather information.[1]

The success of these first meteorological satellite systems led to the design and deployment of both polar-orbiting (i.e., close to Earth and sun-synchronous) satellites and geostationary satellites that provide synoptic overviews of broad aspects of the Earth's surface by a number of the space faring nations. The increase in the number of meteorological satellites in both of these orbits plus the advancing of sensor technology – to obtain higher resolution images in the infrared and in multispectral frequency ranges – have allowed meteorologists to develop more sophisticated modeling of weather patterns. Over time the advent of meteorological satellites has allowed the development of more accurate short-, medium-, and even longer-term weather forecasts. Most recently, the evolution of this technology and its interpretation has allowed meteorological satellites to be applied to not only weather forecasting, but to an array of environmental purposes that include monitoring of atmospheric and oceanic pollution, changes to the polar ice caps and glacial coverage, changes to the protective ozone layer, and broad patterns of climatic changes and global warming. Thus today's meteorological satellites are in many cases environmental monitoring satellites that can provide crucial information as to methane release from the frozen peat fields in Siberia, pollution in the wetlands of the U.S. Atlantic coast, increases in desertification around the world, or changes to the "ozone holes" in the polar regions of the planet.

Advances in Meteorological Satellite Technology

The initial polar-orbiting satellites that revolved around the Earth on a sun-synchronous basis provided updated information on weather conditions approximately every 90 min, but the coverage was for only a narrow strip of the Earth's surface and data in a particular location was updated only once every 24 h. The advantage was much higher resolution than the geostationary satellites that are about 40 times further away from the Earth than the polar-orbiting satellites in 800-km-high orbits. Modern computer processing techniques were increasingly able to process the data from these two types of meteorological satellites to create useful composite images.

In time, the multispectral and infrared cameras were able to produce higher and higher resolution. Over time the technology has continued to develop and improve. The evolving sensor technology has seen the development of Advanced Baseline Imagers (ABIs), Advanced Microwave Sounding Units (AMSUs), Advanced Very High

Resolution Radiometer (AVHRRs), and High Resolution Infrared Radiation Sounders (HIRS). The very latest technology is the development of a Geostationary Lightning Mapper (GLM) which can monitor the intensity of lightning strikes. This mapping allows meteorologist to see where the most intense part of a storm actually is and to "see" the directionality of the moving storm front on a near-real-time basis.

The combination of imagers, radiometers, infrared radiation sounders, and microwave sensors allows for both environmental and meteorological monitoring. The complex suite of sensors on board meteorological satellites combines to follow not only short-term weather phenomena but also enables the observation of longer-term environmental changes, including the monitoring of pollution. Military and civilian technology in this area – both in terms of onboard sensors for Geostationary and polar-orbiting systems as well as computer capacity for processing of much, much greater volumes of data – has evolved quickly in the last two decades. In the process, warnings based on meteorological satellites data have allowed for hundreds of thousands of lives to be saved both by being able to provide short-term warnings with regard to dangerous weather systems and through longer-term forecasts that allow the creation of improved emergency response capabilities involving the most destructive hurricanes, typhoons, and tropical storm systems.

International Cooperation in the Field of Meteorological Satellite Services

Just as NASA research efforts in the 1960s and 1970s with the Television InfraRed Observation Satellite (TIROS), NIMBUS, and Landsat-1 gave rise to the operational meteorological satellites of the National Oceanic and Atmospheric Administration(NOAA) known as GEOS and Advanced TIROS (TIROS-N or ATN) in the United States, the European Space Agency with its Meteosat in the 1970s gave rise in the 1980s to the European Organisation for the Exploitation of Meteorological Satellites (Eumetsat) and its operational METEOSAT and MetOps geostationary and polar-orbiting programs for Europe.[2] In 1977, Japan launched its first meteorological satellite (GMS) into geostationary orbit with an orbital location to cover the western Pacific and East Asia. Similarly, from 1982, the Indian Space Research Organization (ISRO) has deployed INSAT geostationary meteorological satellites and in 1994, building on its experience with the METEOR series of polar-orbiting satellites, the Russian Federation launched the first of its Geostationary Operational Meteorological Satellite (GOMS), later renamed Elektra-1. The 1988 launch by China of the FY-1 (Feng-Yun 1), its first polar sun-synchronous orbit meteorological satellite, led in 1997 to the launch of its FY-2A geostationary meteorological satellite. Through a joint development by the Korea Aerospace Research Institute (KARI) and EADS Astrium, the Communication, Ocean and Meteorological Satellite (COMS-1), the latest of the geostationary meteorological satellites was launched in June of 2010.[3]

There has been a close relationship between the U.S. and European meteorological satellite system for many years and this has most recently resulted in the Joint

Polar Satellite System with the United States developing and operating the Polar Orbiting Operational Environment Satellites (POES) and the Europeans operating the polar-orbiting MetOps system on a coordinated basis.

JPSS, in its current form, was reconstituted by the White House signing an Executive Order in February 2010. This presidential executive order set in motion the dissolution of the so-called National Polar-orbiting Environmental Satellite System that was a joint program involving NASA, NOAA, and the US Department of Defense. In the original form of the NPOESS, the National Ocean and Atmosphere Administration (NOAA) was to operate the polar satellite for the afternoon orbit, while the Defense Weather Satellite System (DWSS) would operate the morning orbit. The earlier program had experienced schedule slips and cost overruns and thus the JPSS seeks to consolidate management, still meet civil and strategic meteorological data requirements, and also bring forward the objective of numerical weather projections. Under the restructured program NOAA, with NASA support, would operate the overall program for the morning and afternoon orbiting satellites, but the Defense Weather Satellite System would still be able to access the data.[4] These systems provide meteorological data to meteorological researchers and to defense-related agencies in the United States and Europe. The POES and MetOps systems, known as the Joint Polar Satellite System (JPSS), are not currently a part of the World Weather Watch global WMO Space Programme network (Fig. 36.1).

Development of the Meteorological Satellite Systems

The Japanese government has also had a strong interest in developing and operating meteorological satellites. This interest is driven by the fact that there are an annual series of typhoons and monsoons that threaten the Japanese islands with highly destructive storms. This has led the Japanese government to undertake the development and deployment of the Geostationary Meteorological Satellite system of Japan also known as the MTSAT or Himawari ("Sunflower in English") Satellite System. The MTSAT satellite series that were constructed by the Space Systems Loral company had difficulties with launch failures that occurred with the Japanese II launch vehicle. After two successive launch failures, the MTSAT 1R was successfully launched. During the period 1999–2005, the time of the two MTSAT failed launch attempts and MTSAT 1R, temporary arrangements were made with the United States for the lease of the GEOS 9 satellite to provide meteorological services to Japan and surrounding areas.[5]

Other countries have continued to develop and launch meteorological satellite systems as well. These system have included a modernization of the Russian Geostationary Operational Meteorological Satellite (GOMS) and METEOR satellites with the Elektro-L and METEOR-3 series of satellite systems, the Chinese Feng-Yang FY-2 and FY-3 systems, and the Indian INSAT-2 satellite system. The highly capable Russian and Chinese systems are dedicated geostationary systems developed to provide detailed meteorological monitoring. Under a joint agreement with NASA, a data server to provide GOMS data has been established in cooperation the Russian



Fig. 36.1 One of the three European Space Agency developed MetOps satellite prior to launch into polar orbit (Graphic courtesy of the European Space Agency)

Ground Microprocessing Information Systems SRC "PLANETA" and the Space Monitoring Information Support Laboratory (IKI RAN).[6]

The INSAT-2 satellite system – also a geo satellite – was designed in an unconventional manner in that some of these spacecraft were designed with two different payloads. One payload provides telecommunications and television broadcasting services, while the other payload supports a meteorological package. In one of these designs a long boom was extended to create overall equilibrium to the solar cell arrays that was deployed on only one side of the spacecraft so as to not block the meteorological imaging (see Fig. 36.2 below on the left). In the latest INSAT 3 and INSAT 4 series, however, a common spacecraft bus was used, but the spacecraft were designed either to provide telecommunications in one edition or as meteorological satellites in the alternative design.[7]

Present and Future Coordination of Meteorological Satellite Systems

The World Meteorological Organization (WMO), a specialized agency of the United Nations, has recognized the importance of coordination of the data produced

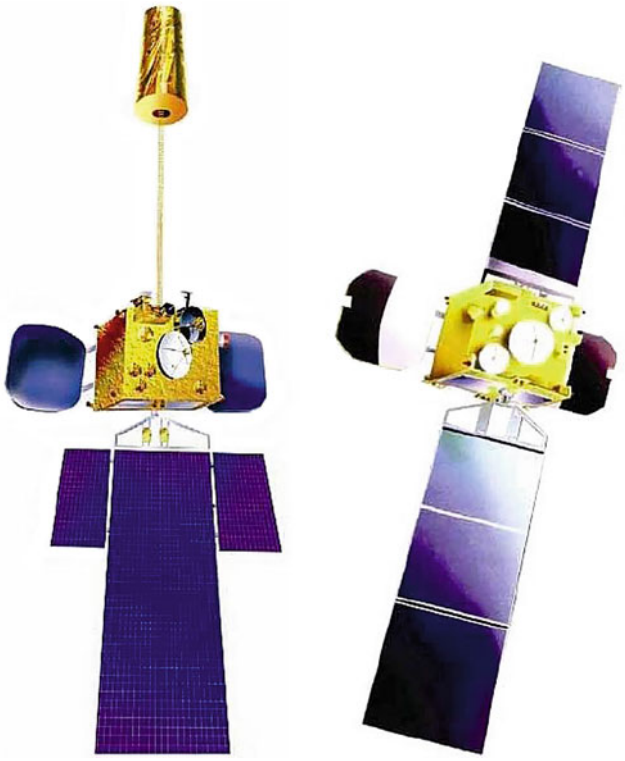


Fig. 36.2 Insat 2 unconventional asymmetrical design along side a more conventional later Insat satellite (Graphics courtesy of ISRO)

by the various national meteorological satellite systems. In May 2003 at the Fourteenth WMO Congress a "WMO Space Programme" was created during the meetings in Geneva, Switzerland. The purpose of the WMO Space Programme was to coordinate all environmental satellite matters around the world – including both meteorological programs but also remote sensing for related environmental purposes such as hydrology, oceanic, and glacial monitoring.[8]

With this WMO Space Programme there is now what is called the World Weather Watch (WWW) Global Observing System (GOS). The GOS includes three different types of satellites. These are operational polar-orbiting satellites, operational geostationary satellites, and research satellites that collect important data but are not considered operational satellites.

The three different types of satellites provide imagery in optical and infrared frequencies, microwave and other types of soundings, data collection, and data distribution. In particular, the present operational meteorological satellites include the following geostationary and polar-orbiting satellites: GOES-10, GOES-12,

NOAA-15, NOAA-16, and NOAA-17 operated by the United States; GMS-5 operated by Japan; GOMS N-1, METEOR 2-20, METEOR 2-21, and METEOR 3-5 operated by the Russian Federation; METEOSAT-5, METEOSAT-6, METEOSAT-7, and METEOSAT-8 (formerly MSG-1) operated by EUMETSAT; and FY-2B, FY-1C, FY-1D operated by China. Additional satellites in orbit include GOES-8, GOES-9, and GOES-11 operated by the United States. It should be noted that most space agencies contributing operational polar-orbiting and geostationary satellites have in place contingency plans for satellite systems that guarantee the continued daily flow of satellite data, products and services WMO Members have come to depend on (Donald Hinsman, *Implementation Activities within the WMO Space Programme* World Meteorological Organization (WMO)).

This has in some instances allowed countries to call into service on a lease or sharing agreement the meteorological satellites of partner countries such as the above noted agreement between Japan and the United States for the interim lease of the GOES-9 satellite by Japan.

Conclusions

The evolution of meteorological satellites in the half century since the 1960s has shown remarkable technological growth and development. Satellites can detect atmospheric temperatures, cloud cover, oceanic current flows and pollution, monitor storm development and energy levels. Remote satellite sensing provides imaging and soundings over a wide range of different frequencies that range from microwave, infrared, multi-spectral imaging (in the optical range) up to the ultraviolet spectra. These diverse imaging platforms provide data from polar and geo orbits and produce an enormous amount of information now measure in petabytes. Modern high speed computers, expert systems, and artificial intelligence allows this data to be processed in near real time to produce valuable data on which aviation, transportation industries, agriculture, forestry, fishing, construction, and virtually every business and individual on the planet to some extent now depend.

Over the past decades several key shifts have occurred with regard to the mission of meteorological satellites. The first shift was to expand the mission from short-term weather forecasts to include medium- and longer-term forecasts as well. The second, more recent, shift has been to expand the mission from weather forecasting to environmental monitoring. This has become increasingly important as the significance of atmospheric, oceanic, and glacial pollution has become better understood and environmental issues such as holes in the ozone layer that protects us from radiation and the potential longer-term implications of climate change, the buildup of green house gases, and global warming have become clearer. The various eyes in the sky that are our meteorological and remote sensing satellites give us the best perspective to "see" the changes that are occurring to our planet and how we can best cope with these shifts in the health and vitality of planetary environment.

Cross-References

- ▶ EUMETSAT Geostationary Meteorological Satellite Programs
- ▶ International Meteorological Satellite Systems
- ▶ United States Meteorological Satellite Program

Notes

1. See Chap. 37, "United States Meteorological Satellite Program" for further details on the early history of satellite communications
2. European Organisation for the Exploitation of Meteorological Satellites(Eumetsat) http://www.eumetsat.int/Home/index.htm
3. See Chap. 39, "International Meteorological Satellite Systems" for more information on these satellite systems
4. Joint Polar Satellite System http://www.nesdis.noaa.gov/pdf/jpss.pdf
5. Japanese Geostationary Meteorological Satellite (GMS) "Himawira" (Sunflower) system by JAXA www.jaxa.jp/projects/sat/gms/index_e.html
6. Present and Future Chinese Meteorological Satellite Systems www.fas.org/spp/guide/china/agency/nsmc.htmData server to provide Russian GOMS satellite data http://gcmd.nasa.gov/KeywordSearch/Metadata.do?Portal=GCMD&KeywordPath=&NumericId=4140&MetadataView=Full&MetadataType=0&lbnode=mdlb2
7. Indian National Satellite System (Insat) ISRO designed Insat E, http://isro-news.blogspot.com/search/label/INSAT%20-%202E
8. Donald Hinsman, *Implementation Activities within the WMO Space Programme* World Meteorological Organization (WMO)

United States Meteorological Satellite Program

37

Sergio Camacho-Lara, Scott Madry, and Joseph N. Pelton

Contents

Introduction 969
Historical Background 970
Nimbus Satellite Program 971
NASA Experimental Programs and the Birth of Geostationary Systems 972
The Geostationary Operational Environmental Satellite (GOES) Network 974
GOES-8 to -12 Satellites 974
GOES 13, 14, and 15 976
GOES-R Series 979
Polar-Orbiting Operational Environmental Satellite (POES) System 981
Initial Joint Polar-Orbiting Operational Satellite (IJPS) System 983
The Defense Meteorological Satellite Program (DMSP): Another Asset 984
New and Future NOAA Satellites: The Joint Polar Satellite System (JPSS) 985
Planned JPSS Satellite Technical Characteristics 986
The NPP Satellite 987
The GOES and POES Ground Systems 988
Conclusion 990
Notes 990
References 990

S. Camacho-Lara (✉)
Centro Regional de Enseñanza de Ciencia y Tecnología del Espacio para América Latina y el Caribe (CRECTEALC), Luis Enrique Erro 1, Tonantzintla, Puebla, Mexico
e-mail: sergio.camacho@inaoep.mx

S. Madry
International Space University, Chapel Hill, North Carolina, USA
e-mail: Scottmadry@mindspring.com

J.N. Pelton
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: peltonjoe@gmail.com

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_54, © Springer Science+Business Media New York 2013

Abstract

Over the past half century, weather and sophisticated environmental imaging satellites have evolved providing an increasing ability to monitor a wide range of conditions on Earth. A long-term and effective partnership between the National Aeronautics and Space Administration, the United States space agency, NASA, and the National Oceanic and Atmospheric Administration, NOAA, has worked to design, launch, and operate a series of environmental monitoring satellites. These environmental monitoring satellites have grown in their technical capabilities to monitor cloud coverage, temperature, and wind velocity over the oceans and seas, lightning intensity, and storm formations. Interactive capabilities have for some time allowed these satellites to assist with search and rescue activities. In short, the expanded technical capabilities of these satellites, and particularly of the Geostationary Operational Environmental Satellite system, have allowed the development of an ever increasing range of applications and functionality.

The initial United States meteorological or weather satellite program that began with TIROS created a specific type of remote sensing satellite that could assist in monitoring weather conditions for the continental US. Today's GOES and Polar-Orbiting Environmental Satellites (POES) have now grown to become global in scope. These US satellites allow the development of an increasingly wide range of knowledge of the oceans and the Polar region, allow for more accurate mathematical models of meteorological conditions, help to monitor "space weather" conditions, assist with rescue of distressed ships and aircraft, aid transportation systems, and help with monitoring atmospheric pollution and conditions associated with climate change.

The National Ocean and Atmospheric Administration, through its National Environmental Satellite, Data, and Information Service (NESDIS), continuously operates a global network of satellites to achieve these goals. NOAA works closely with NASA in the design of environmental satellites and cooperates with the US Department of Defense in obtaining and distributing environmental information. Data obtained from US environmental spacecraft as well as from other spacecraft around the world are used for a wide range of applications. Currently these applications relate to the oceans and seas, coastal regions, agriculture and resource recovery, detection of forest fires, detection of volcanic ash, monitoring the ozone hole over the South Pole, and even the space environment in terms of the so-called space weather such as solar flares.

Each day NOAA's NESDIS processes and then distributes more than 3.5 billion bits of data. The processed images are distributed to weather forecasters in the United States and globally so that various users, for instance, disaster managers, and the general public can see weather patterns via television or on computer or smart phone displays. The timeliness and quality of the combined polar and geostationary satellite data have been greatly improved by enhanced computer installations, upgraded ground facilities, and international data sharing agreements as well as by military weather services.

Keywords
Advanced baseline imager (ABI) • Advanced microwave sounding unit (AMSU) • Advanced very high resolution radiometer (AVHRR) • Data • EUMETSAT • Geostationary lightning mapper (GLM) • Geostationary operational environmental satellites (GOES) • High resolution infrared radiation sounders (HIRS) • Improved TIROS operating system (ITOS) • Information service (NESDIS) • Initial joint polar-orbiting satellite (IJPS) • Joint polar satellites system (JPSS) • MetOp satellites • National oceanic and atmospheric administration (NOAA) • National polar-orbiting operational environmental satellite system (NPOESS) • National weather service (NWS) • Nimbus • Polar-orbiting operational environmental satellites (POES) • TIROS • TIROS-N series

Introduction

For more than 50 years, US environmental satellites have been an integral key to life-saving weather and climate forecasts for the United States and many other countries. NOAA, created in October 1970 to consolidate atmospheric and oceanic related activities and to operate environmental satellites, has been carrying out this mission on behalf of the United States for over 40 years. Its activities include operation of the Geostationary Operational Environmental Satellite (GOES) system and the Polar-Orbiting Environmental satellites (POES). It is also currently involved in developing the next generation Joint Polar Satellite System (JPSS) with EUMETSAT, the European Organisation for the Exploitation of Meteorological Satellites, and the next generation of GOES spacecraft with NASA. Since its establishment, NOAA has cooperated with NASA in the design of new environmental spacecraft that have allowed more sophisticated monitoring capabilities to be deployed with increasing resolution, new types of sensors, and increasing levels of global coverage.

NOAA has, on an on-going basis, coordinated with the US Department of Defense in identifying environmental monitoring requirements and developing increasingly capable environmental spacecraft for monitoring and scientific research. International agreements have allowed NOAA, through its National Environmental Satellite, Data, and Information Service (NESDIS) and its National Weather Service (NWS) to perform its responsibilities more effectively and at an increased level of cost effectiveness. NOAA has, in particular, been able to obtain an augmented amount of information from other satellite systems operated by other countries via various data sharing arrangements. This has not only allowed more accurate weather forecasts, but it has also increased knowledge related to climate change and global patterns of atmospheric, oceanic, and pollution over the polar region.

Today's US capabilities in environmental monitoring and reporting grew out of the International Geophysical Year (IGY) that was carried out during the 1950s during the Eisenhower administration and, in fact, gave birth to the Space Age. The launch of such a first "weather satellite" was committed to by the United States in 1953 as part of the planning process for IGY.

Table 37.1 The TIROS 1 to TIROS 10 satellite network (Gary Davis, NOAA, History of the NOAA Satellite Program, June 2011 http://goes.gsfc.nasa.gov/text/history/History_NOAA_Satellites.pdf)

TIROS satellites	Launch date	Operating life (days)[a]	Orbit	Features
TIROS-1	1 Apr 60	89	Inclined	2 TV cameras
TIROS-2	23 Nov 60	376	Inclined	2 TV cameras, radiometer
TIROS-3	12 Jul 61	230	Inclined	2 TV cameras, radiometer[b]
TIROS-4	8 Feb 62	161	Inclined	1 TV camera, radiometer[b]
TIROS-5	19 Jun 62	321	Inclined	2 TV cameras
TIROS-6	18 Sep 62	389	Inclined	2 TV cameras
TIROS-7	19 Jun 63	1,809	Inclined	2 TV cameras, radiometer[b]
TIROS-8	21 Dec 63	1,287	Inclined	1 TV camera, APT[c]
TIROS-9	22 Jan 65	1,238	Sun Synch[d]	Global coverage, 2 TV cameras
TIROS-10	2 Jul 65	730	Inclined	2 TV cameras

[a]Number of days until satellite was turned off or failed
[b]Radiometer (visible and infrared channels)
[c]Automatic picture transmission for direct readout locally
[d]Sun-synchronous

The TIROS 1 satellite (i.e., Television Infrared Observation Satellite) that inaugurated US environmental space capabilities was actually launched on April 1, 1960, and this became the first of ten such TIROS satellites. A US space-based weather and environmental monitoring capability has continued ever since this event over 50 years ago. Since that time, many dozens of environmental monitoring satellites have been launched. This chapter reviews these past, current, and planned space assets as well as describes cooperative relationships with other environmental spacecraft monitoring systems around the world.

Historical Background

The TIROS series of "weather" satellites initially provided Infrared images of weather conditions in the United States, providing these capabilities through the 1960s. The experimental meteorological spacecraft proved the feasibility of using spacecraft to monitor and predict weather systems, and even to create mathematical models of Earth-based weather conditions. Later satellites in the series provided increased capabilities that included radiometers that operated in the visible as well as the infrared frequencies, that is, in TIROS 3 and 7, while TIROS 8 provided direct read out for local reception and readout (Table 37.1).

After the TIROS series, there was a follow-on TIROS program known as ITOS (Improved TIROS Operational System). These satellites represented a step up from a research and development phase into a fully operational program. ITOS-1 was launched in January 1970 and greatly surpassed the performance of the earlier satellites by providing both direct transmission and storage of television and infrared imagery. Later ITOS spacecraft also supplied vertical profiles of

atmospheric temperature. ITOS satellites remained in service through 1979. The problem of the ITOS and the TIROS series before them was that the near polar orbit required the piecing together of a mosaic of television images to create a unified image of the Earth. This required elaborate and labor intensive ground processing of the satellite images and then redistribution of the processed data.

The initial TIROS system was essentially a "weather" satellite and the ITOS that followed was a much improved system but still essentially a television camera imaging system focused on weather monitoring and forecasting. TIROS and ITOS, however, served as important precursors to the much more complex and capable space-based environmental monitoring system that is operated by the United States today. Another key step in the evolution was the Nimbus satellites designed and launched by NASA as part of its environmental research program from the mid 1960s through much of 1970s.

Nimbus Satellite Program

NASA's Nimbus satellites were flown from 1964 through 1978, as advanced research satellites that tested new sensing instruments and data-gathering techniques rather than as operational weather satellites. These satellites served to test new sensors and to augment the understanding the Earth's environment, including the dynamics of the ozone layer that protects the Earth from solar and cosmic radiation. The Environmental Science Services Administration (former name for the National Weather Service), however, did become a routine user of Nimbus data. This data was valuable for its coverage of conditions over oceans and other areas where few other upper atmospheric measurements were then being made. Instruments on the Nimbus satellites included microwave radiometers, atmospheric sounders, ozone mappers, the Coastal Zone Color Scanner, and infrared radiometers and provided significant global data on sea-ice coverage, atmospheric temperature profiles, atmospheric chemistry (i.e., ozone distribution), the amount of radiation in the Earth's atmosphere, and sea-surface temperature. The Total Ozone Mapping Spectrometer (TOMS) instrument aboard the final Nimbus, Nimbus-7, mapped the extent of the phenomenon known as the "ozone hole." A total of seven Nimbus spacecraft were launched by NASA into near-polar and sun-synchronous orbits. The first of these satellites was Nimbus 1 launched on August 28, 1964. See Fig. 37.1.

The Nimbus series of satellites provided a key capability for satellite remote-sensing of the Earth, atmospheric data collection, and weather forecasting. The seven Nimbus satellites actually provided key data from orbit through 1995 – a remarkably long 30-year period for an experimental program. The technology and lessons learned from the Nimbus missions underlie most of the Earth-observing satellites NASA, NOAA, and other space programs have launched over the past three decades. The basic imaging and sounder systems and the technologies used within the Nimbus are still very much a part of today's systems although current imaging and sounder systems operate at much higher levels of accuracy and frequency of image update. Nimbus operated in near-polar and sun- synchronous orbits, but experiments with geosynchronous orbits that began with the Syncom systems

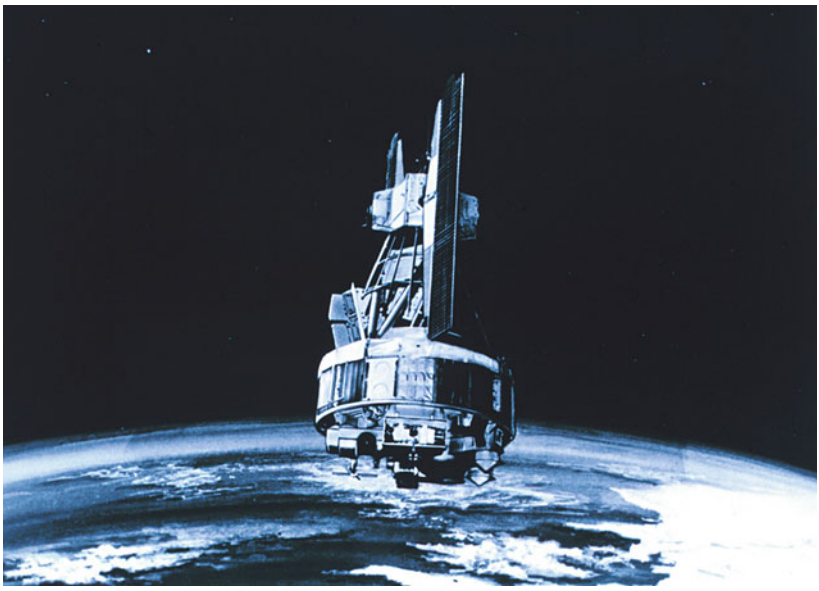


Fig. 37.1 The sun-synchronous near-polar orbiting nimbus meteorological satellite (Graphic courtesy of NASA-GSF and NOAA GOES)

described in earlier chapters related to the history of satellite communications suggested that these orbits could also be effectively used for meteorological and environmental monitoring purposes.

The start of a process that led to a new type of environmental monitoring satellite that would operate from geostationary orbit and provide a near real-time image of the complete "global disk" actually started in 1964, the year the first Nimbus satellite was launched.

In January 30, 1964, a formal agreement was reached between NASA and the United States Weather Bureau to work together to establish a National Operational Meteorological Satellite System. Under this agreement, the Weather Bureau would establish overall requirements and performance characteristics and would also operate command and data acquisition stations. As a consequence of this agreement, NASA was formally tasked with designing the spacecraft and conducting their launch. With the later formation of NOAA, this cooperative agreement continued essentially as initially agreed.

NASA Experimental Programs and the Birth of Geostationary Systems

Part of the answer about how a geostationary satellite might be used for environmental purposes came on December 6, 1966, with NASA's launch of the first Applications Technology Satellite (ATS-1). This experimental satellite

demonstrated the value of using the geostationary orbit for maintaining a continuous watch over one spot on the globe – in this case the continental land mass of the United States. ATS-1's spin-scan cloud camera, invented by Verner Suomi and Robert Parent at the University of Wisconsin, was capable of providing full disk visible images of the Earth, providing an updated cloud cover image every 20 min.

Professor Suomi commented about this totally new capability after the first ATS-1 images were received. He noted, "Now the clouds move and not the satellite." Research by meteorological scientists into tracking clouds and producing models of wind products using image sequences began almost immediately. This was clearly a better way to research cloud formation and atmospheric conditions than having to piece together mosaics of images obtained from sun-synchronous, near-polar orbit satellites.

ATS-3, a larger version of ATS-1 that followed shortly after, was the first spacecraft to routinely transmit full disk color Earth-cloud images. Earlier series spacecraft had peak sensitivity in the green region of the visible spectrum. However, the ATS-3's Multicolor Spin Cloud Cover Camera provided new peak sensitivity to the red, blue, and green visible spectra using three photo-multiplier light detectors.

The NASA-funded ATS experimental satellite series continued development through six spacecraft. The main focus of this ATS-series was on environmental monitoring spacecraft although the ATS-6 was essentially an experimental communications satellite. NASA's office of international affairs worked actively to allow a wide range of international participation in these ATS experiments.

By the early 1970s, ATS imagery was being routinely used in operational forecast centers, with the first movie loops being used at the National Severe Storm Forecast Center (NSSFC) in the spring of 1972. Atmospheric motion depiction from geostationary satellite image loops was transferred into routine operations at the national forecast centers, and the resulting cloud motion vectors evolved into an important data source of meteorological information, especially over the oceans. This data supplemented the data that was provided by the Nimbus program.

The success of the meteorological experiments carried aboard the ATS series of satellites led to NASA's development of a new satellite specifically designed to make atmospheric observations in a geostationary orbit, 35,786 km (22,230 miles) above the equator. NOAA's Geostationary Operational Environmental Satellite (GOES) program thus sprang out of this cooperative experimental period.

In particular, NASA designed, built, and launched the first two geosynchronous meteorological satellites: Synchronous Meteorological Satellite-1 (SMS-1) in May 1974 and SMS-2 in February 1975. These two spacecraft were the prototypes for the NOAA GOES program. GOES-1 was launched on October 16, 1975, followed by GOES-2 and -3, which were similar in design and provided continuity of service. The primary instrument on the SMS-1 and -2 and GOES-1 to -3 spacecraft was what was called the Visible/IR Spin Scan Radiometer (VISSR), and was based on Professor Soumi's original conceptual design. The VISSR, a true radiometer, provided day and night observations of cloud and surface temperatures, cloud heights, and wind fields.

The fundamental shift from experimental projects to operational programs occurred in the early 1970s. The initial TIROS and the ATS programs, which ended around 1970 and in the mid 1970s respectively, were largely replaced by the ITOS (i.e., TIROS-N series) and NOAA's new Geostationary Operational Environmental Satellite (GOES) system starting in 1975. The GOES network has continuously evolved over the past 35 years into what is now a full-scale environmental monitoring network that also supports all of NOAA's goals and scientific functions with new capabilities being systematically added. The US GOES network and its other spacecraft work closely with the international community, especially with EUMETSAT and the World Meteorological Organization.

The Geostationary Operational Environmental Satellite (GOES) Network

The United States operates two meteorological satellites in geostationary orbit: one over the East Coast, GOES EAST around West Longitude – 75.0° and one over the West Coast, GOES WEST around West Longitude – 135° with overlapping coverage over the United States. These satellites are replaced when they reach the end of their operation lifetime by satellites with improved or additional new instruments.

A total of 15 GOES Satellites have been deployed by NOAA since 1975. These GOES Satellites as described below have become increasingly sophisticated over the past few decades.

These space assets play a key role in supporting the major operational and strategic goals of NOAA. These goals today include (Davis 2011):

- Supporting the needs of the United States and those of world society to obtain up-to-date and reliable weather, water, and other related information, including data on atmospheric and oceanic pollution, on a timely basis
- Protecting, restoring, and managing the use of coastal and ocean resources through an ecosystems approach to water use and management
- Understanding climate variability and enhancing society's ability to plan and respond to climate change
- Supporting United States' commerce by supplying accurate and timely information for safe, efficient, and environmentally sound transportation

The various spacecraft assets described in the next section (i.e., GOES and polar-orbiting environmental monitoring spacecraft) now play a critical role in meeting these strategic objectives. Increasingly capable environmental monitoring satellites operated by the United States and other entities – particularly EUMETSAT – allow NOAA to accomplish much more than provide weather data.

GOES-8 to -12 Satellites

The history of the early GOES satellites and their technical characteristics will not be covered in detail here. This is because these earlier spacecraft, although they

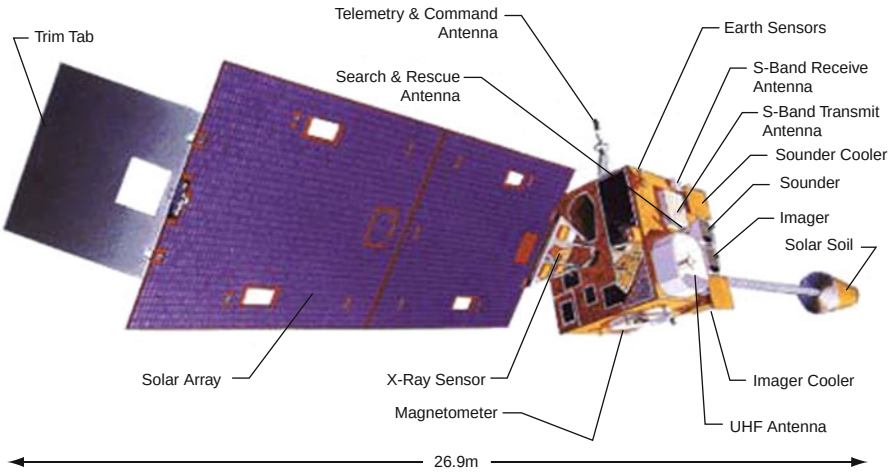


Fig. 37.2 GOES-12 (M) Launch date: July 23, 2001 (Graphic courtesy of NASA-GSF and NOAA GOES)

once performed a vital service, are now fully retired from service and their historical specifications are only currently relevant in that they helped to prove technologies that are incorporated in today's operational spacecraft.

Let's thus start with a review of the GOES-8 to -12 (known as GOES I to M prior to launch). These spacecraft are currently being phased out and are being replaced by the GOES- 13, -14 and -15 that will be described in the next section.

Beginning with launches in the mid-1990s, the GOES-8 to -12 series spacecraft provided weather and environmental monitoring for the United States for over a decade. These spacecraft were equipped to perform the following specific functions:

- Acquisition, processing, and dissemination of imaging and sounding data
- Acquisition and dissemination of Space Environment Monitor (SEM) data
- Reception and relay of data from ground-based Data Collection Platforms (DCPs) that are situated in selected urban and remote areas to the NOAA Command and Data Acquisition (CDA) station
- Continuous relay of Weather Facsimile (WEFAX) and other data to users, independent of all other functions
- Relay of distress signals from people, aircraft, or marine vessels to the search and rescue ground stations of the Search and Rescue Satellite Aided Tracking (SARSAT) system

These satellites, as exemplified by GOES-12 (M) in Fig. 37.2, have two major subsystem capabilities in the form of the GOES I/M Imager and a Sounder[1]:

The GOES I/M Imager is a multichannel instrument designed to sense radiant and solar-reflected energy from sampled areas of the Earth. The multielement spectral channels simultaneously sweep east–west and west–east along a north-to-south path by means of a two-axis mirror scan system. The instrument can produce full-Earth disk images, sector images that contain the edges of the Earth, and various sizes of area scans completely enclosed within the Earth scene using

Table 37.2 GOES I/M imager monitoring capabilities

Imager channels and capabilities					
Channel	1	2*	3*	4	5*
Wavelength (μm)	0.65	3.9	6.7	11	12
Product					
Clouds	x	x	x	x	x
Water vapor*			x	x	x
Surface temperature		o		x	o
Winds	x		x	x	
Albedo + IR Flux	x		o	x	o
Fires + Smoke	x	x		o	o

Key: *new operational data, *x* primary channel, *o* secondary channel

a new flexible scan system. Scan selection permits rapid continuous viewing of local areas for monitoring of mesoscale (regional) phenomena and accurate wind determination. Table 37.2 above indicates the specific technical performance characteristics of the GOES I/M Imager.[2]

The GOES I/M Sounder is the other major component of this type NOAA spacecraft. The "GOES I/M Sounder" is a 19-channel discrete-filter radiometer. This radiometer covers the spectral range from the visible wavelengths up to 15 μm. It is designed to provide data from which atmospheric temperature and moisture profiles, surface and cloud-top temperatures, and ozone distribution can be deduced by mathematical analysis. An engineering sketch of the GOES Sounder (or Radiometer) is provided in Fig. 37.3 below.

The GOES I/M series Sounder or Radiometer operates independently of and at the same time as GOES I/M Imager. The Sounder, in fact, uses a similar flexible scan system as is used in the Imager. The Sounder's multielement detector array assemblies simultaneously sample four separate fields or atmospheric columns. A rotating filter wheel, which brings spectral filters into the optical path of the detector array, provides the infrared channel definition. These capabilities are listed in Table 37.3 below.[3]

GOES 13, 14, and 15

The current generation of GOES satellites is now fully deployed. These spacecraft were known as GOES N/O/P prelaunch and as GOES-13, GOES-14, and GOES-15 postlaunch. These spacecraft (See Fig. 37.4 below) provide important new capabilities in terms of weather monitoring, environmental monitoring, and active search and rescue operations that are described below. In addition to an imager and sounder with expanded capabilities, these satellites also carry an upgraded Space Environmental Monitor and a Solar X-Ray Imager. They also kept a remote data collection capability from land and ocean-based DCPs. Finally, they kept the communications capability to support search and rescue operations. The network operates as a two satellite configuration from over the Pacific (GOES West) and

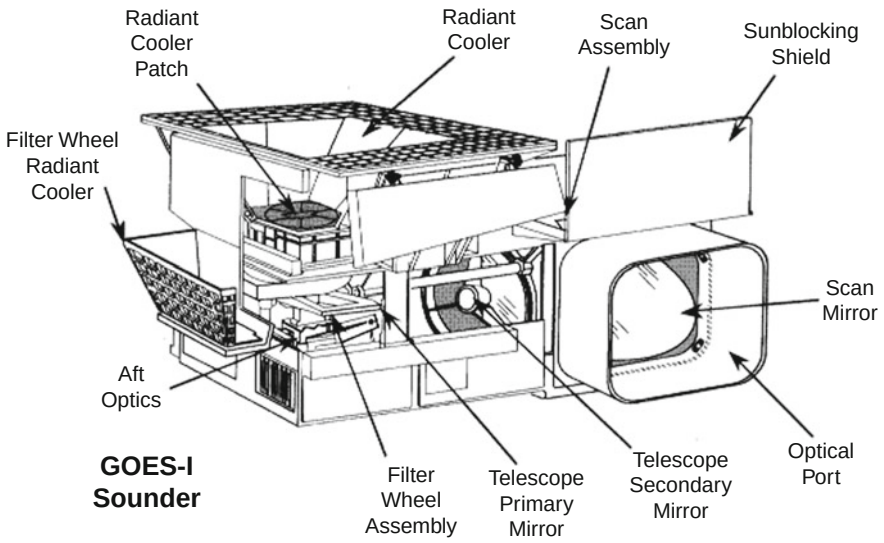


Fig. 37.3 Engineering sketch of the GOES I series sounder (Graphic courtesy of NASA-GSF and NOAA GOES)

Table 37.3 GOES I/M sounder environmental monitoring capabilities

THE GOES I/M sounder capabilities				
	Resolution (km)		Accuracy	
	Vertical	Horizontal	Absolute	Relative
Product				
Temperature				
Profile	3–5	50	2–3°K	1°K
Land	–	10	2°K	1°K
Sea	–	10	1°K	0.5°K
Moisture				
Profile	2–4	50	30%	20%
Total	–	10	20%	10%
Motion	3 layers	50	6 m/s	3 m/s
Cloud				
Height	2 layers	10	50 mb	25 mb
Amount	Total	10	15%	5%
Ozone*				
Total	–	50	30%	15%
Motion	1 layer	50	10 m/s	5 m/s
IR Flux*	total	50	10 W/m^2	3 W/m^2

Key: *potential future product



Fig. 37.4 The GOES 13 satellite (known as GOES N prior to launch) (Graphic courtesy of NASA-GSF and NOAA GOES)

Atlantic Ocean (GOES East) with the capability to continuously observe about 60% of the Earth's Surface. The third spacecraft is available as a spare.

A summary of improvements of the current GOES N/O/P spacecraft in comparison to the spacecraft they are replacing (i.e., the GOES I/M series) is as follows:

- The satellite power system (i.e., solar array and battery systems) has been upgraded in reliability and performance.
- The design lifetime has been upgraded from 7 to 10 years and sufficient fuel has been provided to support 13.5 years of life which should be fully achievable.
- The command data rate for each satellite has been increased from 250 bits/s to 2,000 bits/s to allow more complex commands at faster data rates. Telemetry data rates have also been improved for faster read outs of up to 4,000 bits/s if required.
- There is a star tracking system that will allow more precise pointing of the spacecraft. This will allow for more accurate registration and navigational capabilities.
- Solar data can now be collected by inclusion of a Solar X-Ray Imager.
- The Space Environmental Monitoring (SEM) subsystem has been upgraded in a number of ways to include an Extreme Ultra Violet (EUV) sensor, an Energetic Proton, Electron, and Alpha particle Detector (EPEAD), and a Magnetosphere Electron Detector (MAGED). This will greatly increase the ability to monitor space weather conditions.
- An additional transponder has been added to support an Emergency Manager's Weather Information Network (EMWIN).
- The data from the spacecraft will be distributed via a digital Low Rate Information Transmission (LRIT) system that will replace the analog WEFAX data distribution system that operated at lower speeds and greater potential for disruption.
- The data collection system will allow the collection of more data from more remote land and ocean-based platforms at higher speeds (i.e., up to 1,200 bits/s) through the use of an eight phase shift keyed system.

A full description of each of the four major sensing systems on the GOES 13–15 spacecraft, their technical characteristics, and their associated ground system capabilities are available on the NOAA and NASA web site.[4]

It is important to know that this fully functional system can provide NOAA meteorologists and experts on the ground with a wide and rich range of key data and actionable information. The read out systems are organized within the NOAA information network to include: Dust Storms, Fire Events, Flood Events, Iceberg Events, Ocean Events, Severe Weather Events, Snow Cover, Storm Systems, Tropical Cyclones, Unique Imagery, and Volcano Events.

The design and manufacture of the GOES spacecraft, from GOES 1–15, are in many ways similar to the design of the spacecraft used for telecommunications, satellite navigation, or other types of remote sensing activities. This can be clearly seen by undertaking a review of the power systems (i.e., solar arrays and batteries), the stabilization and tracking system, fuel used for stabilization, station-keeping and repositioning, thermal control and heat transfer, and tracking, telemetry, and command systems as noted in ▶ Chap. 44, "Common Elements Versus Unique Requirements in Various Types of Satellite Application Systems". That is, the satellite platforms are all quite similar in design and manufacture. This is really not surprising as most application satellites are manufactured by the same companies and the research scientists and engineers who work on the various subsystems develop technology for a "platform" that can be used on any application satellite. Just as a manufacturer of vehicles might use the same engine or chassis on various automobile models, the same is true for manufacturers of spacecraft. The many innovations that were applied to the design of the GOES N-P spacecraft such as star tracking, additional fuel for extended lifetime, faster data rates for command and telemetry, and improved power subsystems are found in other types of applications spacecraft being deployed today. The innovations that are unique to these spacecraft are those associated with the four sensors. The imager and sounder are improved versions of those found on GOES I-M series and the space environmental sensors are essentially new capabilities.

GOES-R Series

The current GOES System is expected to continue operating at least for another 5–7 years. The additional fuel and design upgrades to extend the life of GEOS 13, 14, and 15 should actually allow them to operate beyond this time. The GOES-R, expected to launch in 2015, will operate with a mission design life of at least 7–10 years and improve current GOES capabilities. These new capabilities of the GOES R series are described below.

Despite various discussions and initiatives in the past that considered the idea of trying to "privatize" weather and meteorological satellite operations within the United States, it is currently intended and broadly anticipated that the GOES-R series will continue to be collaboratively developed and acquired by NOAA and NASA. The acquisition of the end-to-end GOES system includes spacecraft, sensors, launch

services, and ground system elements consisting of mission management, product generation, product distribution, archive and access interface, and user interface.

New instrumentation that will be deployed within the GOES-R spacecraft includes that Advance Baseline Imager and the Geostationary Lightning Mapper. Both of these new instruments are expected to provide important new capability that will enhance public and transportation safety and lead to economic savings and benefits [5]:

Advanced Baseline Imager (ABI): The ABI is the primary instrument on GOES-R for imaging Earth's weather, climate, and environment. The ABI will be able to view the Earth within a very broad range of 16 different spectral bands, including two visible channels, four near-infrared channels, and ten thermal infrared channels. The ABI is almost revolutionary in its capabilities in that it provides three times more spectral information, four times better spatial resolution, and more than five times as many updates of its sensed data than the GOES-13 to -15 satellites in the current system.

The ABI is designed to observe essentially the entire western hemisphere in various time intervals and to do so at 0.5, 1, and 2 km spatial resolutions in 16 spectral bands as indicated above. The ABI has two main scan modes. The "flex" mode will provide full disk imagery every 15 min, while covering the continental US every 5 minutes. It also has the ability to provide data on demand as frequently as every 30 seconds. It is expected that two mesoscale regions will be continuously scanned, resulting in a 1 minute update for those sectors.

The ABI will be calibrated to an accuracy of 3% (1 σ) radiance for visible and near-infrared wavelengths. For infrared channels, the ABI will be accurate to 1^{o} K (1 σ) at 300^{o} K. The ABI on the GOES R satellite will thus actually improve every product compared to the current GOES Imager and introduce new products. Two new products to be produced by the ABI include the capability to indicate the probability of fog and its density and the accurate detection of precise vectors of atmospheric motion.

The GOES-R series satellites will also provide increased time-response capabilities with respect to fires, volcanoes, as well as hurricanes, tornadoes, and thunderstorms. The projected cost benefits of the GOES-R ABI instrument (and the GLM instrument described below) over the lifetime of the series is estimated by NOAA to be close to $5 billion (US). These projected benefits are expected to come from improved tropical cyclone forecasts, fewer weather-related flight delays, and airline incidences with volcanic smoke and ash plumes, improved production and distribution of electricity and natural gas, increased efficiency in irrigated water usage in agriculture, and higher protection rates and more efficient rerouting for airplanes and ships in the event of a tropical storm or hurricane.

Geostationary Lightning Mapper: The GLM is an optical transient detector and imager operating in the near-IR that maps total lightning (in-cloud and cloud-to-ground) activity with near uniform spatial resolution of approximately 10 km continuously day and night over the Americas and adjacent ocean regions. The GLM will provide early indication of storm intensification and severe weather events, improved tornado warning lead time of up to 20 min or more, and data for long-term climate variability studies. It is anticipated that GLM data will have

immediate applications to aviation weather services, climatological studies, and severe thunderstorm forecasts and warnings. The GLM will provide information to identify growing, active, and potentially destructive thunderstorms over land as well as ocean areas.

GLM measurements can provide vital information to help the operational weather, aviation, disaster preparedness, and fire monitoring communities in a number of different and quite significant ways:

- An increased ability to develop short range forecasts of heavy rainfall and flash flooding.
- New capability to provide near-real time detection of enhanced lightning activity that with associated improved models, can predict changes in the intensity change of tropical storms, hurricanes, and cyclones.
- Related improved warning capabilities for tornado and severe thunderstorm in terms of increased lead times as well as a corresponding reduction in false alarms and spurious information. (This is a particularly valuable new capability of importance for storm warning and transportation routing for oceanic regions, mountain areas, and areas where there might be radar outages).
- Improved routing of commercial, military, and private aircraft over oceanic regions, mountain areas, and sparsely populated and remote areas during severe storm conditions.
- More accurate and timely warning of lightning ground strike hazards.
- Development of improved and more accurate numerical weather prediction models increased identification of deep atmospheric convection patterns.
- Increased capability to develop what might be called "lightning climatology" and models of lightning intensity within storms.
- Improved ability to monitor and create mathematical models of a wide range of storm and lightning intensity patterns.

One will note that many of the projected benefits from the ABI and GLM are common and that in many instances the combined analysis of the ABI and GLM products will provide the optimum result.[6]

Polar-Orbiting Operational Environmental Satellite (POES) System

The Polar-orbiting Operational Environmental Satellite (POES) system supplements the GOES system. The POES offers the advantage of daily global coverage, by making nearly polar orbits roughly 14.1 times daily. Since the number of orbits per day is not an integer, the suborbital tracks do not repeat on a daily basis, although the local solar time of each satellite's passage is essentially unchanged for any latitude. Currently in orbit there are two satellites, one of which passes over a given point on Earth at the same local time in the morning and the other in the afternoon. These spacecraft, referred to as AM and PM satellites, provide global coverage four times daily. The POES system includes the Advanced Very High Resolution Radiometer (AVHRR) and the Tiros Operational Vertical Sounder (TOVS).

Because of the polar orbiting nature of the POES series satellites, these satellites are able to collect global data on a daily basis for a variety of land, ocean, and atmospheric applications. Data from the POES series supports a broad range of environmental monitoring applications including weather analysis and forecasting, climate research and prediction, global sea surface temperature measurements, atmospheric soundings of temperature and humidity, ocean dynamics research, volcanic eruption monitoring, forest fire detection, global vegetation analysis, search and rescue, and many other applications. These images and data supplement the information that GOES satellites provide.

In 1998, a new series of NOAA Polar Operational Environmental Satellites (POES) commenced with the launch of NOAA-K. The NOAA-K and its immediate successors, NOAA-L and NOAA-M, represent an improvement over the previous series of satellites that began with TIROS-N (1978), and continued with NOAA-6 through NOAA-14 (1994).

The NOAA K/L/M POES satellites begin a new era of improved environmental monitoring. The NOAA K/L/M satellites, NOAA-15 to NOAA-17 after launch, include improvements to instruments that are evolutionary. The initial concept was to add more passive microwave instruments and channels in place of the four-channel Microwave Sounding Unit (MSU) and the three channel Stratospheric Sounding Unit (SSU). Combined with command system security and frequency changes, NOAA K/L/M satellites look very much like previous satellites, but have significant changes to essentially every subsystem. A description of the instrumentation of the NOAA K/L/M satellites is presented below.

- The Advanced Microwave Sounding Units (AMSU-A1, AMSU-A2, AMSU-B) are state-of-the-art passive microwave sounders that significantly enhance NOAA's atmospheric sounding and non-sounding products suite. The AMSU instruments have better spatial resolution and upper atmospheric sounding capabilities than the previous MSU instrument flown on the TIROS-N series. The HRPT broadcasts at the old data rate of 665.5 kbps with the new AMSU data replacing what were previously spare words.
- The Advanced Very High Resolution Radiometer (AVHRR/3) provides spectral and gain changes to the visible channels that will allow improved low energy/light detection and adds a sixth channel, called 3A, at 1.6 μm for improved snow and ice discrimination. Channel 3A will be time shared with the previous 3.7 μm channel, now called channel 3B. The Automatic Picture Transmission (APT) user sees channel 3B as channel 6 using the wedge six grayscale modulation index.
- The High Resolution Infrared Radiation Sounder (HIRS/3) has spectral channel changes that were made primarily to improve soundings and to be congruent with the specifications developed for the GOES-I through -M Sounders. The HIRS/3 cooler set point was decreased to approximately 100° K which will improve the two infrared detectors' performance.
- The Space Environment Monitor (SEM-2) has improved calibration and particle detection capabilities. The Total Energy Detector (TED) measures to a lower

energy of 0.05 KeV and the TED integral F (ALPHA) has two ranges of 0.05–1 and 1–20 KeV. The Medium Energy Proton and Electron Detector (MEPED) have a fourth omnidirectional proton measure at 140 MeV.

- The Solar Backscatter Ultra Violet Spectral Radiometer (SBUV/2) has undergone relatively modest improvements. Its Programmable Read Only Memory (PROM) will be changed to a Random Access Memory (RAM) due to parts obsolescence and to provide more operational flexibility.
- The Data Collection System (DCS) data rate increased from 1,200 to 2,560 bps and the number of Data Recovery Units (DRUs) doubled from 4 to 8. DCS-2 bandwidth increased from 24 kHz to 80 kHz.
- The Search and Rescue Processor (SARP-2) Data Recovery Units increased from 2 to 3 to handle more global distress messages and to better detect interfering signals.

NOAA-19, designated NOAA-N (NOAA-N Prime) prior to launch, was launched on February 6, 2009, and is the last of NOAA's POES series of weather satellites. NOAA-19 carries a suite of instruments that provides data for weather and climate predictions. Like its predecessors, NOAA-19 provides global images of clouds and surface features and vertical profiles of atmospheric temperature and humidity for use in numerical weather and ocean forecast models, as well as data on ozone distribution in the upper part of the atmosphere, and near-Earth space environments – information important for the marine, aviation, power generation, agriculture, and other communities. The NOAA-19 primary instruments – the Advanced Very High Resolution Radiometer (AVHRR/3), High Resolution Infrared Radiation Sounder (HIRS/4), and the Advanced Microwave Sounding Unit (AMSU-A) – were all designed for a 3-year mission. The Solar Backscatter Ultraviolet Spectral Radiometer (SBUV/2) was designed for a 2-year mission, and the Microwave Humidity Sounder (MHS) was designed for a 5-year mission.

The POES series of satellites will be followed by a series of Earth observation satellites with greatly improved instrumentation, the National Polar-orbiting Operational Environmental Satellite System (NPOESS). The NPOESS Preparatory Project (NPP) satellite was launched on October 28, 2011. NPP's instruments are described later in this chapter.

Initial Joint Polar-Orbiting Operational Satellite (IJPS) System

Building upon the Polar-orbiting Operational Environmental Satellite (POES) program, an agreement is in place between NOAA and EUMETSAT on what is called the Initial Joint Polar-orbiting operational Satellite (IJPS) System. This program includes two series of independent but fully coordinated NOAA and EUMETSAT satellites. This program involves the exchange of instruments and global data, cooperation in algorithm development, and near real-time direct broadcasting. Under terms of the IJPS agreement, NOAA provides NOAA-18 and

NOAA-19 satellites for flight in the afternoon (PM) orbit and EUMETSAT provides MetOp-A and MetOp-2 (B) satellites for flight in the mid-morning orbit (AM). These satellites carry a common core of instruments that includes[7]:

- Third Generation Advanced Very High Resolution Radiometer (AVHRR/3): A six-channel imaging radiometer to detect energy in the visible and IR portions of the electromagnetic spectrum.
- High Resolution Infrared Radiation Sounders (HIRS/4): A multispectral atmospheric sounding instrument to measure scene radiance in the IR spectrum.
- Advanced Microwave Sounding Unit (AMSU-A): A cross-track scanning total power radiometer to measure scene radiance in the microwave spectrum.
- Data Collection System (DCS): To collect and store environmental study data from multiple platforms for transmission to the ground once per orbit to NOAA Command and Data Acquisition stations.
- Search and Rescue Satellite-Aided Tracking (SARSAT) Instruments: These are part of the international COSPAS-SARSAT system designed to detect and locate Emergency Locator Transmitters (LET), Emergency Position-Indicating Radio Beacons (EPIRB), and Personal Locator Beacons (PLB) operating at 121.5 MHz, 243 MHz, and 406 MHz to subsequently downlink to a Local User Terminal.
- Space Environmental Monitor (SEM): Provides measurements to determine the intensity of the Earth's radiation belts and the flux of charged particles at satellite altitude.
- Microwave Humidity Sounder (MHS): A five-channel microwave instrument to measure profiles of atmospheric humidity.

In addition, NOAA satellites fly a Solar Backscatter Ultraviolet (SBUV) Radiometer instrument (a nadir pointing, nonspatial, spectrally scanning, ultraviolet radiometer), while EUMETSAT's additional payloads include an infrared interferometer sounder, a scatterometer, an ozone instrument, and a Global Positioning System (GPS) occultation sounder.

Coordination on associated ground segments included in this agreement ensures the sharing of all mission data, blind-orbit data capture support, and telecommunications paths through each other's ground stations for backup command and control functions. The first MetOp satellite was launched on October 19, 2006, from Baikonur Cosmodrome, Kazakhstan. See ▶ Chap. 27, "Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing" for a more detailed description of the meteorological satellite program of EUMETSAT.

The Defense Meteorological Satellite Program (DMSP): Another Asset

The spacecraft of the Defense Meteorological Satellite Program (DMSP[8]) can "see" the best among all weather satellites with its ability to detect objects almost as "small" as a large oil tanker. Some of the most spectacular photos have been

recorded by the night visible sensor; city lights, volcanoes, fires, lightning, meteors, oil field burn offs, as well as the Aurora Borealis and Aurora Australis have been captured by this 830 km-high space vehicle's low moonlight sensor.

At the same time, energy monitoring as well as city growth can be accomplished since major and even minor cities, as well as highway lights, are conspicuous. This also informs astronomers of light pollution. In addition to monitoring city lights, these photos are a life-saving asset in the detection and monitoring of fires. Not only do the satellites see the fires visually day and night, but the thermal and infrared scanners on board these weather satellites detect potential fire sources below the surface of the Earth where smoldering occurs. Once the fire is detected, the same weather satellites provide vital information about wind that could fan or spread the fires.

NOAA also currently operates the Defense Meteorological Satellite Program near-polar-orbiting series of satellites. The satellites were initiated by the Defense Department in the mid1960s and were initially the responsibility of the US Air Force. Each DMSP satellite, orbiting at approximately 516 miles (830 km) above the Earth, crosses any point on the Earth up to twice a day. These satellites see such environmental features as clouds, bodies of water, snow, fire, and pollution in the visible and infrared spectra. Scanning radiometers record information that can help determine cloud type and height, land and surface water temperatures, water currents, ocean surface features, ice, and snow. Communicated to terminals on the ground, the data is processed, interpreted by meteorologists, and used in planning and conducting US military operations worldwide. On May 5, 1994, however, President Bill Clinton decided to merge America's military and civil polar-orbiting operational meteorological satellite systems into a single, national system that could satisfy both civil and national security requirements for space-based environmental data. Called the National Polar-orbiting Operational Environmental Satellite System (NPOESS), NOAA is responsible for the now integrated network.[9]

New and Future NOAA Satellites: The Joint Polar Satellite System (JPSS)

On February 1, 2010, the Executive Office of the President restructured the National Polar-orbiting Operational Environmental Satellite System (NPOESS) into two separate development programs: one aimed at the civilian community, the Joint Polar Satellite System (JPSS) and the Defense Weather Satellite System (DWSS) to satisfy Defense Department requirements. The civilian and scientific community program is led by NOAA who sets the requirements and NASA who is directing the acquisition. JPSS will provide operational continuity of satellite-based polar missions in the afternoon orbit that support its civil regional and global weather and climate requirements. In addition, JPSS will provide oceanographic, environmental, and space environmental information. The system's instrumentation is described below.

Planned JPSS Satellite Technical Characteristics

The currently planned capabilities for the JPSS satellites include six key subsystems as follows:

- *Visible/Infrared Imager/Radiometer Suite (VIIRS)*: VIIRS is an electro-optical imager having multiband imaging capabilities which collects calibrated visible/infrared radiances to produce data products for cloud and aerosol properties, land surface type, vegetation index, ocean color, land and sea surface temperature, and low light visible imagery. The 22-channel VIIRS will fly on the NPOESS Preparatory Project (NPP) and on all JPSS platforms to provide complete daily global coverage over the visible, short/medium-infrared, and long-wave infrared spectrum at horizontal spatial resolutions of 370 m and 740 m at nadir. VIIRS is the primary instrument for 21 different types of environmental data records.
- *Cross-track Infrared Sounder (CrIS)*: CrIS is a Fourier Transform Spectrometer that uses a Michelson interferometric sounder capable of sensing upwelling infrared radiances from 3 to 16 μm at very high spectral resolution (∼1,300 spectral channels) to determine the vertical atmospheric distribution of temperature, moisture, and pressure from the surface to the top of the atmosphere across a swath width of 2,200 km.
- *Advanced Technology Microwave Sounder (ATMS)*: ATMS is a cross-track high-spatial-resolution microwave sounder. ATMS data will support temperature and humidity sounding generation in cloud covered conditions. ATMS has 22 microwave channels to provide temperature and moisture sounding capabilities in the 23/31, 50, 89,150, and 183 GHz spectral range.
- *Ozone Mapping and Profiler Suite (OMPS)*: The OMPS monitors ozone from space. OMPS will collect total column and vertical profile ozone data and continue the daily global data produced by the current ozone monitoring systems, the Solar Backscatter Ultraviolet radiometer (SBUV)/2 and Total Ozone Mapping Spectrometer (TOMS), but with higher fidelity. The nadir sensor uses a wide field-of-view push-broom telescope to feed two separate spectrometers. The nadir total column spectrometer (mapper) measures the scene radiance between 300 and 380 nm with a resolution of 1 nm sampled at 0.42 nm and a 24-h ground revisit time. The limb sensor measures the along-track limb scattered solar radiance with 1 km vertical sampling in the spectral range of 290–1,000 nm. Three vertical slits sample the limb at 250 km cross-track intervals to provide for better than 7-day ground revisit times to improve the precision of the ozone profiles. The three slits are imaged onto a single charge-coupled device (CCD) (identical to both nadir CCDs). Due to limitations with flight hardware transferred from NPOESS to JPSS, the OMPS on JPSS J1will consist of a nadir sensor only.
- *Cloud and Earth Radiant Energy System (CERES)*: The CERES instrument seeks to develop and improve weather forecast and climate models prediction, to provide measurements of the space and time distribution of the Earth's Radiation Budget (ERB) components, and to develop a quantitative understanding of the links between the ERB and the properties of the atmosphere and

surface that define that budget. CERES consists of three broadband radiometers that scan the earth from limb to limb. Data from CERES will be used in conjunction with VIIRS to study changes in the Earth's energy balance and key changes in clouds and aerosols to determine the effect of changing clouds on the Earth's energy balance.

- *Total Solar Irradiance Sensor (TSIS)*: The Total Solar Irradiance Sensor (TSIS) will measure variability in the sun's solar output, including total solar irradiance. TSIS consists of two instruments: the Total Irradiance Monitor (TIM) that measures the total light coming from the sun at all wavelengths and the Spectral Irradiance Monitor (SIM) that will measure how the light from the sun is distributed by wavelength. These measurements are needed to understand how solar radiation interacts with the Earth's surface and atmosphere. TSIS is an important climate sensor that will help maintain continuity of the climate data record for space-based solar irradiance measurements that now spans over three decades.

JPSS will ensure continuity of crucial climate observations and weather data in the future. Data and imagery obtained from the JPSS will increase timeliness and accuracy of public warnings and forecasts of climate and weather events reducing the potential loss of human life and property damage. The data collected by JPSS will contribute to the unified and coherent long-term environmental observations and products that are critical to climate modelers and decision makers concerned with advancing climate change understanding, prediction, mitigation and adaptation strategies, policies, and science. JPSS, with its global view, will play a vital role in continuing these climate data records for the US and the international community.

The NPP Satellite

As mentioned above, the NPOESS Preparatory Project (NPP) satellite was launched on October 28, 2011. NPP will serve as an important link between the current generation of Earth-observing satellites and the next generation of climate and weather satellites of the JPSS. NPP observes the Earth's surface twice every 24-h day, once in daylight and once at night. In its orbit NPP flies 512 miles (824 km) above the surface in a polar orbit, circling the planet about 14 times a day. NPP sends its data once an orbit to the ground station in Svalbard, Norway and continuously to local direct broadcast users.

Of the six JPSS subsystems described above, NPP carries five instruments, VIIRS, CrIS, ATMS, OMPS, and CERES to monitor the environment on Earth and the planet's climate.[10] NPP measurements will be used to map land cover and monitor changes in vegetation productivity. NPP tracks atmospheric ozone and aerosols as well as takes sea and land surface temperatures. NPP monitors sea ice, land ice, and glaciers around the world. In addition to continuing these data records, NPP is also able to monitor natural disasters such as volcanic eruptions, wildfires, droughts, floods, dust storms, and hurricanes/typhoons. In all, NPP monitors the health of Earth from space – providing continuity to decades-long records and setting the stage for future Earth science missions.

An important design feature of NPP's VIIRS instrument is that it tracks land cover changes and vegetation productivity, extending the successful and widely used data records of NASA's Moderate Resolution Imaging Spectroradiometer (MODIS), a similar instrument launched aboard NASA's Terra and Aqua spacecraft in 1999 and 2002.

NPP's CERES instrument continues the Earth radiation budget data record started by the Earth Radiation Budget (ERB) instrument on the Nimbus-7 satellite in 1978 and continued through a series of NASA satellites since then, including CERES instruments on the satellites Terra and Aqua.

The CrIS and the ATMS instruments on board NPP work together, providing global high-resolution profiles of temperature and moisture. These advanced atmospheric sensors create cross sections of storms and other weather conditions, helping with both short-term "nowcasting" and long-term forecasting. CrIS measures continuous channels in the infrared region and has the ability to measure temperature profiles with improved accuracy over its predecessor instruments on operational satellites, and comparable accuracy to the Atmospheric Infrared Sounder (AIRS) on Aqua. NOAA will be using CrIS for numerical weather prediction and, because it is a brand new instrument, its use on NPP provides a real-world test of the equipment before NOAA's upcoming Joint Polar Satellite System (JPSS) missions. The ATMS instrument works in both clear and cloudy conditions, providing high-spatial-resolution microwave measurements of temperature and moisture. ATMS has better sampling and two more channels than previous instruments like the Advanced Microwave Sounding Units (AMSU), and it combines all of their abilities into one instrument. Working in concert, CrIS and ATMS together comprise the Cross-track Infrared Microwave Sounding Suite (CrIMSS).

The OMPS instrument measures the ozone layer in our upper atmosphere, tracking the status of global ozone distributions, including the in "ozone hole" region. It also monitors ozone levels in the troposphere, the lowest layer of our atmosphere. OMPS will extend a 40-year long record of ozone layer measurements while also providing improved vertical resolution compared to previous operational instruments. Closer to the ground, OMPS's measurements of harmful ozone will improve air quality monitoring and when combined with cloud predictions, help to create the Ultraviolet Index, a guide to safe levels of sunlight exposure for people. The complexity of the NPP satellite can be appreciated in the photograph shown as Fig. 37.5 when the NPP satellite was in its testing phase prior to its launch.

The GOES and POES Ground Systems

The GOES Ground System is, in fact, a "System-of-Systems" that comprises the end-to-end framework for collecting, processing, and disseminating critical environmental data and information from the satellites. It supports the launch, activation, and evaluation of new satellites and the in-depth assessments of satellite data quality. Data from the satellites are received at ground facilities, where the data are processed to monitor and control the satellite and to generate products that are used



Fig. 37.5 The NPP satellite prior to launch (Photograph courtesy of NASA)

by NOAA, its users, and the world meteorological community. The GOES ground system consists of components at the Satellite Operations Control Center (SOCC) at Suitland, Maryland; Command and Data Acquisition (CDA) facilities at Wallops, Virginia, and Fairbanks, Alaska; and Wallops Backup (WBU) facility at NASA Goddard Space Flight Center (GSFC) in Greenbelt, Maryland.

The POES mission operates with a NOAA- provided constellation of two operational satellites in circular, near-polar, sun-synchronous orbits that provide scheduled down-loads of environmental data collected from space to the POES Ground System for satellite monitoring and control and mission processing, analysis, and distribution. The POES Ground System is also a "System-of-Systems" that includes collecting, processing, and disseminating critical environmental data and information from the POES satellites. Operational elements are located at Fairbanks, Alaska; Wallops, Virginia; and Suitland, Maryland. It contains subsystems located in the following NESDIS Offices: Office of Satellite Operations (OSO), Office of Research and Applications (ORA), and the NOAA National Data Centers (NNDC). See ► Chap. 43, "Ground Systems for Satellite Application Systems for Navigation, Remote Sensing, and Meteorology" for a more detailed description of the ground system of the meteorological satellite program of the United States.

Conclusion

The US Meteorological Satellite System is now essentially integrated under the operation of the National Oceanic and Atmospheric Administration with NASA assisting with the design and launch of these satellites and the US Department of Defense, and particularly the National Reconnaissance Office and US Air Force, assisting with regard to the design and operation of the Defense Meteorological Satellite Program (DMSP). The US meteorological system consists of a combination of geosynchronous imaging and sounder satellites (GEOS system) that provide a near-real time image of the Earth disk on a 24 h a day (i.e., day and night) basis, while the POES system provides a sun-synchronous image of the entire Earth with some 14 passes over the Earth for each satellite. The combination of two US satellites and two European satellites provide very rapid updating of information. The coordination of various international programs is accomplished via the World Meteorological Organization. ▶ Chap. 39, "International Meteorological Satellite Systems" presents the status of various national and regional meteorological satellite systems and the global coordination of these systems.

Notes

1. GOES I/M Brochure, goes.gsfc.nasa.gov/text/goesimbroch.html
2. GOES I/M Imager, goes.gsfc.nasa.gov/text/imager.html
3. GOES I/M Sounder, goes.gsfc.nasa.gov/text/sounder.html
4. http://www.lib.noaa.gov/noaainfo/heritage/noaahistory.html
5. Geostationary Operational Environmental Satellites (GEOS-R) Program end-to-end technical study, GEOS-R Program requirements document and GEOS-R mission requirements document, see http://www.osd.noaa.gov/
6. Additional GOES-R information can be found at: http://www.goes-r.gov/History of NOAA
7. http://goes.gsfc.nasa.gov/text/history/History_NOAA_Satellites.pdf
8. http://www.ngdc.noaa.gov/dmsp/index.html
9. Meteorological satellites, http://www.centennialofflight.gov/essay/SPACEFLIGHT/metsats/SP35.htm
10. NPP building a bridge to a new era of earth observations, (http://www.nasa.gov/mission_pages/NPP/mission_overview/index.html)

References

G. Davis, NOAA, History of the NOAA satellite program, June 2011, http://goes.gsfc.nasa.gov/text/history/History_NOAA_Satellites.pdf

EUMETSAT Geostationary Meteorological Satellite Programs

38

Declan Murphy

Contents

Introduction 992
Overview of Past and Current EUMETSAT Programs 993
Current Status of EUMETSAT Programs 994
EUMETSAT Partnerships and International Cooperation 995
The MSG Satellites: An Overview 996
The MSG Space Segment 998
The Instruments 998
Spinning Enhanced Visible and Infrared Imager (SEVIRI) 998
Geostationary Earth Radiation Budget (GERB) 1001
Mission Communication Payload (MCP) 1001
Search and Rescue Transponder 1001
The MSG Ground Segment 1002
The Centrally Located Facilities 1002
The Ground Stations 1004
Product Extraction Facilities 1005
Data Collection and Data Dissemination 1008
The EUMETSAT Data Centre 1011
Meteorological Third Generation (MTG) 1013
The Instruments 1015
Cross-References 1019
Notes 1019
References 1019

Abstract

EUMETSAT, the European Organisation for the Exploitation of Meteorological Satellites, operates a range of satellite programs, among them the Meteosat series of geostationary satellites which has provided continuity of coverage over Europe

D. Murphy
Met Éireann (the Irish Meteorological Service), Glasnevin Hill, Dublin 9, Ireland
e-mail: declan.murphy47@gmail.com

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_55, © Springer Science+Business Media New York 2013

and Africa since 1977. Its current operational geostationary services are provided by the Meteosat Second Generation (MSG), consisting of the primary satellite, the Meteosat-9, the back-up and Rapid Scanning Service Meteosat-8, as well as the older generation Meteosat-7 satellite positioned over the Indian Ocean. It works closely in partnership with the European Space Agency and with NOAA in its programs. As a user-driven organization it places great emphasis on developing additional value from its products by sophisticated systems for processing of the satellite data, centrally at its headquarters in Darmstadt, and through a distributed network of Satellite Applications Facilities in its Member States. A successor program to the MSG, the Meteosat Third Generation, has been approved and will ensure coverage out to 2040.

Keywords

Calibration • Climate monitoring • EUMETSAT • European Space Agency • Further processing • Geostationary • Instruments • Meteorology • Meteosat • National meteorological services • Numerical weather prediction • Radiometer • Satellite applications facilities • Satellites • Weather forecasting

Introduction

Since the launch of the first meteorological satellites by the United States in the 1960s, meteorologists around the world have made extensive use of the images and data available from them. Weather systems that previously had just been drawings on weather maps took on real shapes from the satellite images. As weather forecasting techniques developed, more sophisticated use of the satellite data became possible. Particular applications were the use of the data in Nowcasting, i.e., forecasting for the very short range up to 12 h ahead which relies heavily on observational data, and Numerical Weather Prediction techniques, mathematical models of the atmosphere which ingest observational data of many types, including numerical data derived from satellite instrument measurements, and then produce forecasts up to 10 days ahead.

Of paramount importance is the accurate monitoring and forecasting of severe weather situations to help save human lives and property. The ravages of weather-related disasters take a heavy toll of life around the world and the images available from geostationary meteorological satellites, with a frequent repeat cycle, are of huge value in the vital task of predicting such events and helping to mitigate their impacts.

While the primary justification for investment in a meteorological satellite program continues to lie in the benefits that it brings to operational weather forecasting, increasingly, the wider benefits of the programs for Earth observation of all kinds (including of the oceans, atmosphere, and land) and for monitoring of climate and of climate change add considerably to the value of the programs.

In Europe, EUMETSAT operates programs of meteorological satellites that benefit not only Europe but also Africa and other regions and forms part of a global satellite coverage with other satellite operators. EUMETSAT is an intergovernmental

organization set up in 1986 to exploit the benefits of European operational meteorological satellite programs. Originally composed of 16 Member States, its membership had grown by the end of 2010 to 26 states. It also has five Cooperating States.

Current EUMETSAT Member States are: Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Luxembourg, the Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, and the United Kingdom.

Cooperating States are: Bulgaria, Estonia, Iceland, Lithuania, and Serbia.

The impetus for the creation of EUMETSAT came from the national meteorological services of Europe. Originally the intention was that the organization would fund meteorological satellite programs (both geostationary and polar-orbiting) that would be defined jointly with the European Space Agency (ESA), which would then develop and operate the satellites, EUMETSAT gradually came to play a greater part in the specification of the space segment of the programs and set up its own ground station infrastructure to operate its satellites and to develop useful products from the satellite data.

EUMETSAT headquarters is located in Darmstadt in Germany. The Director-General reports to the EUMETSAT Council which is the governing body of the organization and consists of delegations from all Member States.

Overview of Past and Current EUMETSAT Programs

The history of European meteorological satellites began before the establishment of EUMETSAT with the launch by ESA of the first Meteosat geostationary satellite in 1977 as part of its activity to develop space applications. Meteosat-1, as it came to be called, gave Europe the ability to gather weather data over its own territory for the first time with its own satellite. The 800-kg first-generation Meteosat was followed by two other satellites in what is termed the preoperational series of Meteosat satellites. When EUMETSAT was formed it took over the funding of the Meteosat series for Meteosat-4, Meteosat-5, and Meteosat-6. The next in the series, Meteosat-7 was launched in 1997 and became the first satellite program to be devised and funded by EUMETSAT in what was called the Meteosat Transition Programme. The transition refers to the bridging of a gap between the older "first-generation" Meteosat satellites and the planned Meteosat Second Generation (MSG).

The first-generation Meteosat was equipped with the three-channel Meteosat Visible and Infrared Imager and a repeat cycle of 30 min. This system operated successfully since 1977 providing almost continuous images and other services to the National Meteorological Services of EUMETSAT Member States and brought major improvements to weather forecasting in Europe. Meteosat has also served operational and research users throughout West and East Europe and Africa, with many other users in North and South America, the Middle East, and even in the Arctic and Antarctic areas. Technological advances and increasingly sophisticated weather forecasting requirements created demand for observations to be more frequent, more accurate, and in higher resolution. To meet this demand, the scope and



Fig. 38.1 Launch of Meteosat-9 satellite from Guiana Space Centre, 21 December 2005

objectives of the Meteosat Second Generation (MSG) Program were defined in 1993 to form the basis of cooperative programs established at EUMETSAT and ESA.

On 28 August 2002, EUMETSAT launched the first MSG satellite (renamed Meteosat-8 when it began routine operations) from the Guiana Space Centre in Kourou, French Guiana, on board an Ariane rocket. It was followed on 21 December 2005 by the second MSG satellite, Meteosat-9 (Fig. 38.1).

Current Status of EUMETSAT Programs

The operational geostationary meteorological satellite service over Europe is now exclusively provided by Meteosat Second Generation (MSG) satellites – Meteosat-8 and Meteosat-9. Together, the two satellites provide the operational service for Europe of a quality never before experienced from geosynchronous orbit. Follow-on satellites are scheduled for launch in 2012 and 2014 (www.eumetsat.int/home/main/satellites/MeteosatSecondGeneration; www.esa.int/esaMI/MSG).

Meteosat-9 provides the primary satellite coverage at a position above the equator at 0° longitude. Meteosat-8, situated at 9.5° East, is the backup for

the primary satellite. The 15-minute high-resolution image data generated by the operational Meteosat satellite, Meteosat-9, is complemented by the Rapid Scanning Service (RSS) of Meteosat-8. The RSS service was originally introduced in 2001 as a service that permitted more frequent imaging over a limited field of view. In the MSG era that imaging interval is 5 min, the same as European weather radars, allowing the monitoring of rapidly developing localized convective weather systems like thunder storms.

All MSG images and products are distributed via EUMETCast-Europe, EUMETSAT's broadcast system for environmental data, and on the Global Telecommunication System of the World Meteorological Organization (WMO) and are archived at EUMETSAT.

The first generation of Meteosat, however, is still providing an operational service through Meteosat-7, the survivor of that series following the retirement of Meteosat-6 in April, 2011. Meteosat-6 had set a record of more than 17 years for the duration of the operational life of a Meteosat satellite. Located at 57.5°E, Meteosat-7 provides the Indian Ocean Data Coverage (IODC) service. The operational service from this location began on 1 July 1998 and is currently planned to remain operational until the end of 2013. IODC data provide important information on monitoring cyclonic systems, dust storms, and other meteorological phenomena in the Indian Ocean region. As part of the IODC service, the Meteosat-7 satellite relays tsunami warnings covering the Indian Ocean region. As with all satellite data, the IODC data are also of value as input to numerical weather prediction systems.

Another major element of the EUMETSAT suite of meteorological satellites is the EUMETSAT Polar System (EPS) which currently has one polar-orbiting satellite, the Metop-A, in orbit. Together with the POES satellite operated by NOAA, the Metop-A provides a global coverage which greatly enhances the geostationary satellite activity. Two further satellites in the EPS program, Metop-B and Metop-C are due for launch in 2012 and 2016, respectively.

EUMETSAT is also part of a cooperation which operates an ocean altimetry satellite, Jason-2, which provides oceanographic and meteorological information. The other cooperation partners are NASA, NOAA, and the Centre National d'Études Spatiales (CNES, the French Space Agency). A follow-on satellite (Jason-3) will be launched in 2013.

EUMETSAT Partnerships and International Cooperation

The geostationary satellite programs of EUMETSAT (and also its polar-orbiting satellites) are carried out in close cooperation with the European Space Agency (ESA), representing an efficient use of European resources and reflecting the complementary roles of the two organizations. ESA is committed to research and development rather than to operational systems but the development of the first satellite in a program falls naturally within its mandate (www.esa.int/esaMI/MSG). EUMETSAT is a user-driven operational agency. The usual formula for

cooperation between the two organizations is for ESA to develop the first satellite, based on user requirements specified by EUMETSAT, with EUMETSAT also making some contribution to the costs. Development of new instruments (often pioneered on ESA research satellites) is also part of ESA's role. EUMETSAT covers the costs of recurrent satellites and provides the entire ground structure.

EUMETSAT also has close cooperation with the meteorological community, mainly through the World Meteorological Organization (WMO), the European Centre for Medium-range Weather Forecasts (ECMWF), and, of course, with the national meteorological services of its Member States and Cooperating States.

A close partnership exists with NOAA and manifests itself in the form of collaboration on the operational global polar-orbiting program and of a backup agreement between the two agencies that commits each side to help the other in the case of operational difficulty such as satellite failure.

EUMETSAT also has cooperation agreements with the other meteorological satellite providers and shares data with them.

It is also active in the international structures that facilitate cooperation in satellite meteorology. It is a member of the Coordination Group for Meteorological Satellites (CGMS) and, in fact, provides the secretariat for CGMS. It is also a member of the Committee for Earth Observation Satellites (CEOS) and contributes to other international activity such as the Group on Earth Observation (GEO) and the European Union's Global Monitoring for Environment and Security (GMES).

The MSG Satellites: An Overview

The MSG system is designed to support weather forecasting in all ranges including Nowcasting (out to 12 h ahead), short-range forecasting (out to 2 days), medium range (out to 2 weeks). It makes a valuable contribution to numerical weather prediction through the numerical data that are derived from the satellite instruments. It also contributes to climate applications over Europe and Africa.

The mission objectives are:

- Multi-spectral imaging of the cloud systems, the Earth surface, and radiance emitted by the atmosphere, with improved radiometric, spectral, spatial, and temporal resolution when compared to the first-generation Meteosat.
- Extraction of meteorological and geophysical fields from the satellite image data for the support of general meteorological, climatological, and environmental activities.
- Dissemination of the satellite image data and meteorological information, upon processing, to the user community in a timely manner.
- Collection of meteorological and environmental data from Data Collection Platforms (DCPs), and its distribution to appropriate users.
- Support to secondary payloads of a scientific nature (GERB) and to Search and Rescue (called GEOSAR). These payloads do not interfere with the primary objectives as laid out above (www.eumetsat.int/home/main/satellites/MeteosatSecondGeneration).

Like all geostationary satellites, the satellites in the MSG series are placed in orbit approximately 36,000 km above the equator to ensure that their orbits are synchronized with the Earth's movement and so have a fixed field of view of the Earth.

Meteosat Second Generation (MSG) consists of a series of four geostationary meteorological satellites, along with ground-based infrastructure, that will operate consecutively until 2020, with the possibility to extend until 2022. The main instrument carried by the MSG satellites is the imaging radiometer, the Spinning Enhanced Visible and InfraRed Imager (SEVIRI), which has the capacity to observe the Earth in 12 spectral channels. In addition, a scientific research instrument, the Geostationary Earth Radiation Budget (GERB), is included in the payload. A Search and Rescue transponder, GEOSAR, is also supported on the satellites (Schmetz et al. 2002).

In common with other programs devised by EUMETSAT, the MSG program is a partnership between EUMETSAT and the European Space Agency (ESA). The role of EUMETSAT was to establish requirements for the space segment based on users requirements, ensuring the consistency between those requirements and the MSG satellites, to contribute one third of the MSG-1 (Meteosat-8 in-orbit) funding, to procure the MSG-2/3/4 satellites via ESA, to procure all launch services and all services for post-launch early operations, to develop the ground segment, and to operate the system.

ESA developed the MSG-1 prototype satellite according to EUMETSAT requirements and acted, on behalf of EUMETSAT, as procurement agent for MSG-2/3/4 satellites.

The prime contractor for the manufacture of the MSG satellites is Thalès Alenia Space, France. Astrium SAS manufactured the SEVIRI instrument, while a European consortium, led by the Rutherford Appleton Laboratory (RAL) in the United Kingdom, was responsible for the Geostationary Earth Radiation Budget (GERB) instrument.

Each MSG satellite was designed to remain in orbit in an operable condition for at least 7 years. The current policy is to keep two operable satellites in orbit and to launch a new satellite close to the date at which the elder of the two comes to the end of its onboard fuel. Starting toward the end of the MSG in orbit lifetime there will be a follow-on series in geostationary orbit, Meteosat Third Generation.

The SEVIRI instrument observes the atmosphere over 12 spectral channels with repeat cycles of 15 min in nominal mode and 5 min in rapid-scanning mode. There are 11 spectral channels observing the earth with a sampling distance of 3 km, generating full earth images. The last of the 12 spectral channels observes the earth with a sampling distance of 1 km, generating a partial view of the earth in the east/west direction, with the possibility of splitting the observed area at a programmable latitude such that the top and bottom areas can be shifted in the east/west directions from each other.

MSG's enhanced channel and scan capacity, with the ability to transmit more than 20 times the information at twice the speed of its predecessor has opened up a range of improved applications for users.

The enhanced imagery delivered by MSG provides detailed maps to support operational weather forecasts by, for example, improved animations of developing weather conditions, such as potentially hazardous fog banks around shipping lanes or storms on an airplane's flight path. MSG is also proving to be an invaluable tool for climate monitoring. Satellite observation of clouds, precipitation, and temperature

provide valuable sources of data to climate researchers. MSG's ability to monitor atmospheric water vapor, dust, and surface features (such as the distribution of snow, ice and vegetation) is also proving to be essential in understanding the Earth's climate.

In order to operate the satellite and to collect data from it and to derive value-added products and to disseminate and archive the data, a sophisticated ground segment is a part of the program.

The MSG Space Segment

Like the previous generation of Meteosat satellites, MSG is spin-stabilized. When operating in geostationary orbit, the satellite spins counterclockwise at 100 rpm around its longitudinal axis, which is parallel with the Earth's rotational axis (Fig. 38.2).

The MSG body is cylindrical in shape, 3.2 m in diameter, and 2.4 m high, with the top antenna protruding to about 3.8 m. The satellite itself is built in a modular way around three main sub-assemblies:

– The Spinning Enhanced Visible and Infrared Imager (SEVIRI) instrument in the central compartment
– The Mission Communication Payload (MCP), including antennas and transponders, in the upper compartment
– The platform support subsystems, in the lower compartment

For its initial boost into geostationary orbit as well as for station keeping, the satellite uses a bipropellant system. This includes small thrusters, which are also used for attitude control. The MSG solar array, built from eight curved panels, encloses the satellite body.

The satellite platform provides the necessary housekeeping functions for accommodation and service of the payloads (enabling control from the ground), for orbital motion and stabilization, and for energy supply.

The communications payload provides multichannel transponders and antennas for:

- Downlink of onboard-generated payload data (within the raw data stream of approximately 3.3 Mbps) and onboard satellite monitoring data
- Uplink of telecommands to the satellite
- Relay of messages from DCPs to the MSG Ground Segment, of low rate information from the MSG Ground Segment to users, and of GEOSAR beacons

The Instruments

Spinning Enhanced Visible and Infrared Imager (SEVIRI)

As with all meteorological satellites, the principal instrument is the imaging radiometer which operates by collecting radiances from the Earth's surface and atmosphere.

On the MSG satellites, the radiometer is the SEVIRI which has 12 spectral channels (as opposed to three on the previous Meteosat systems), which provide

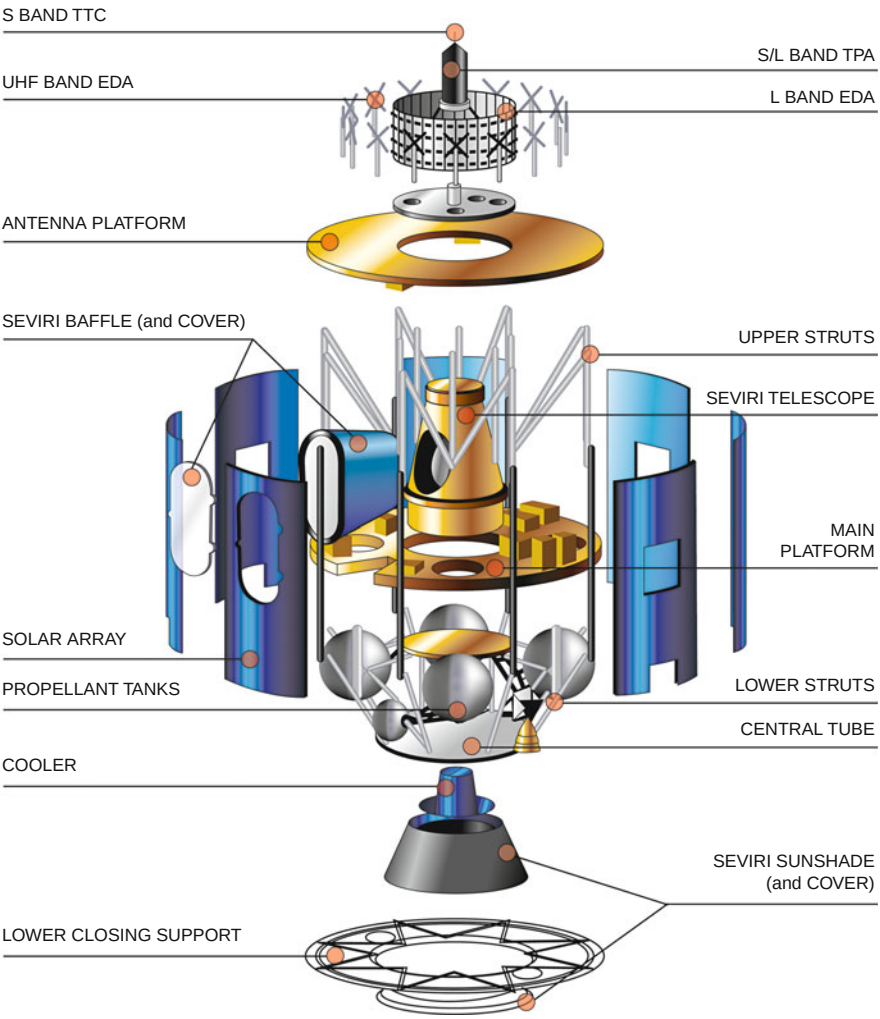


Fig. 38.2 Schematic of the MSG satellite

more precise data throughout the atmosphere giving improved quality to the starting conditions for numerical weather prediction models (Schmid 2000).

The 12 SEVIRI channels consist of 8 infrared (IR) detector packages (3 detectors each), 2 Visible and 1 Near-IR (3 detectors each) and 1 High Resolution in the Visible (HRV) channel (9 detectors). The full list is:
Visible band centered on 0.6 μm
Visible band centered on 0.8 μm
Near-infrared band centered on 1.6 μm
Infrared band centered on 3.9 μm
Water vapor band centered on 6.2 μm

Water vapor band centered on 7.3 μm
Infrared band centered on 8.7 μm
Ozone band centered on 9.7 μm
Infrared band centered on 10.8 μm
Infrared band centered on 12.0 μm
Carbon dioxide band centered on 13.4 μm
Broadband high-resolution visible band

The operating principle of the SEVIRI is that a scanning mirror is used to move the instrument line-of-sight (LOS) in the south-north direction. The target radiance is collected by the telescope and focused onto the detectors. Channel separation is performed at telescope focal-plane level, by means of folding mirrors. A flip-flop type mechanism is periodically actuated to place an IR calibration reference source in the instrument's field of view. The image data are directly transferred from the Main Detection Unit (MDU) to the satellite data-handling subsystem. The Functional Control Unit (FCU) controls the SEVIRI functions and provides the telemetry and telecommand interfaces with the satellite. The Earth imaging is achieved by means of a bidimensional Earth scan, relying on the spacecraft's spin and the scanning mirror. The rapid scan (line scan) is performed from east to west thanks to the spacecraft's rotation around its spin axis (spin rate 100 rpm). The spin axis is perpendicular to the orbital plane and is nominally oriented in the south-north direction. The slow scan is performed from south to north by means of a scanning mechanism, which rotates the scan mirror in steps of 125.8 μrads. A total rotation of the mirror of $\pm 5.5^\circ$ (corresponding to 1,527 scanning lines) is used to cover the entire field of view of the Earth plus margin, corresponding to 22° imaging range in the south-north direction, and 1,210 scan lines in the baseline repeat cycle.

The full Earth's disk image is obtained in about 12 min. The scanning mirror is then driven back to its initial position and the flip-flop mechanism is activated to insert a black body onboard the spacecraft into the optical path for the instrument calibration. The black body is removed from the calibration position after about 2 s and Earth observation is resumed, leading to an overall repeat cycle of 15 min.

From each channel geographical arrays of image pixels are received, each pixel containing 10 data bits, representing the received radiation from the Earth and its atmosphere.

The resulting data are buffered onboard and transmitted to the ground over a full revolution to moderate the downlink rate.

Using channels that absorb ozone, water vapor, and carbon dioxide, MSG satellites allow meteorologists to analyze the characteristics of atmospheric air masses making it possible to reconstruct a three-dimensional view of the atmosphere. The improved horizontal image resolution for the visible light spectral channel (1 km as opposed to 2.5 km on the previous system) also greatly aids weather forecasters in detecting and predicting the onset or cessation of severe weather.

Calibration of the satellite instruments is a vital part of the entire process. The spectral response of an instrument is a measure of the instrument's response to radiation at specific wavelengths and spectral response characterization is the most crucial aspect of satellite calibration. The responses are nonlinear and may change

over the lifetime of an instrument, making it necessary to correct for these changes after launch when producing images from the system. The accuracy of prelaunch spectral response characterization, and how well the on-orbit changes are understood, directly affects calibration accuracy and the quality of the data products. Even within the same satellite series, spectral response varies by instrument, sometimes dramatically. In fact, spectral response often varies by detector on the same instrument. Spectral responses are derived for all 12 channels of the SEVIRI instrument.

Geostationary Earth Radiation Budget (GERB)

In addition to the SEVIRI there is the Geostationary Earth Radiation Budget (GERB) instrument. The GERB is a visible-infrared radiometer for Earth radiation budget studies. It makes accurate measurements of the shortwave (SW) and longwave (LW) components of the radiation budget at the top of the atmosphere. It was the first Earth Radiation Budget experiment from geostationary orbit. The GERB provides valuable data on reflected solar radiation and thermal radiation emitted by the Earth and atmosphere (www.eumetsat.int/home/main/satellites/MeteosatSecondGeneration; Schmetz et al. 2002).

The GERB instrument is a scanning radiometer with two broadband channels, one covering the solar spectrum (0.32–4.0 μm), and the other covering a wider portion of the electromagnetic spectrum (0.32–40 μm). Together these channels are used to derive the thermal radiation emitted by the Earth in the spectral range 4.0–40 μm, as a difference between the two of them. Data are calibrated on board in order to support the retrieval of radiative fluxes of reflected solar radiation and emitted thermal radiation at the top of the atmosphere with an accuracy of 1%. The radiation budget represents the balance between incoming energy from the Sun and outgoing thermal (longwave) and reflected (shortwave) energy from the Earth.

Mission Communication Payload (MCP)

This package contains all antennas and transponders necessary to meet the demanding communication needs of the MSG mission. This includes telemetry, telecommanding and transmission relay links, in various frequency bands.

Search and Rescue Transponder

The satellite payload includes a transponder capacity for relay of distress signals. This is the Search and Rescue (GEOSAR) mission, which forms part of the International Satellite System for Search and Rescue (COSPAS-SARSAT). The Search and Rescue transponder receives distress signals from any mobile unit in difficulty within the MSG coverage zone in Europe, Africa, and the Atlantic Ocean.

The MSG Ground Segment

The functions of the MSG Ground Segment are to control and communicate with the satellites, to collect and validate the instrumental data, to enhance the value of the data by further processing, to disseminate the data to the users, and to archive them for future use by researchers.

The EUMETSAT MSG Ground Segment is made up of:

- *Centrally located facilities*, at EUMETSAT headquarters in Darmstadt, Germany, for satellite control, preprocessing of data, dissemination, and archiving
- Primary and backup ground stations
- *Product extraction facilities,* consisting of the Meteorological Products Extraction Facility (MPEF) based in Darmstadt, and the network of Satellite Application Facilities (SAFs), distributed among the EUMETSAT's Member States

The Centrally Located Facilities

The Mission Control Centre (*MCC*) controls the EUMETSAT satellites through the relevant ground stations and preprocesses all data acquired from these satellites. Among its tasks are control of the satellite orbit, monitoring the status of the satellite through checking various characteristics such as temperature, fuel consumption, etc., and monitoring the ground stations. In general, it ensures that all aspects of the satellite performance are monitored, controlled and, if necessary, corrected by the initiation of commands that are relayed to the satellite by the ground stations. It also performs all operations necessary for satellite tracking and ranging. A team of operations personnel man the Mission Control Centre on a continual basis (Fig. 38.3).

Raw sensor data are received from the ground stations and forwarded to the Image Processing Facility (IMPF). These level 1.0 data, as they are known, are then subject to further processing to generate level 1.5 data. In the case of Meteosat image processing, this entails line-by-line processing to ensure that imperfections are removed. In particular, the data from the various onboard sensors are realigned by resampling in order to make the image from each set of detectors coincide with the reference grid. At the same time, the sampling removes the slight perturbations caused by the movement of the spacecraft, and corrects for effects such as Earth's curvature and rotation, thereby rectifying the image so that it appears to come from the nominal location of the spacecraft (geometric correction). Further adjustments to the individual data values are made to correct for atmospheric effects and for sensor characteristics in accordance with calibration information (radiometric correction). Once each line of the level 1.5 products is complete, it is passed to the dissemination computers for immediate relay to users and to higher processing facilities, Meteorological Products Extractions Facility (MPEF) and the Satellite Application Facilities (SAFs).

Calibration is a vitally important aspect of operational meteorological satellite programs. In addition to prelaunch and initial orbit calibration of the instrument sensors, a routine calibration monitoring is performed for IR channels (using the



Fig. 38.3 Mission control centre at Darmstadt

internal black body of SEVIRI) and for visible channels (using vicarious calibration on dedicated stable earth targets). Based on the above calibration methods, radiometric corrective factors are used to generate the level 1.5 image from the level 1.0 image.

The Meteosat ground segment also supports the *Global Earth Radiation Budget* (*GERB*) mission, handling all communications with the instrument onboard the Meteosat, and the reception of raw GERB data. The GERB data are then sent to the central GERB ground segment at the Rutherford Appleton Laboratory (RAL) in the UK for processing and forwarding to other European institutions.

Another element of the MSG Ground Segment acquires Foreign Satellite Data and makes them available to the further processing systems or for dissemination to EUMETSAT users.

The *GEOSAR Ground Segment* has relayed to it messages from the COSPAS-SARSAT beacons on board the satellites. In this case the satellites act as simple transponders only.

The MSG Ground Segment is connected to *the Global Telecommunication System* (*GTS*) *of WMO*, so as to receive GTS messages for further dissemination or for use in the MPEF. It also ingests into the GTS network meteorological products produced by MPEF, as well as all international and selected regional DCP messages. The GTS connection is via the GTS Regional Telecommunications Hub at the Deutscher Wetterdienst in Offenbach, Germany.

The EUMETSAT data collection and dissemination systems are described below, as is the *EUMETSAT Data Centre* which archives images and meteorological products from all satellite programs, and provides access to these data to users via the *EUMETSAT Portal*.

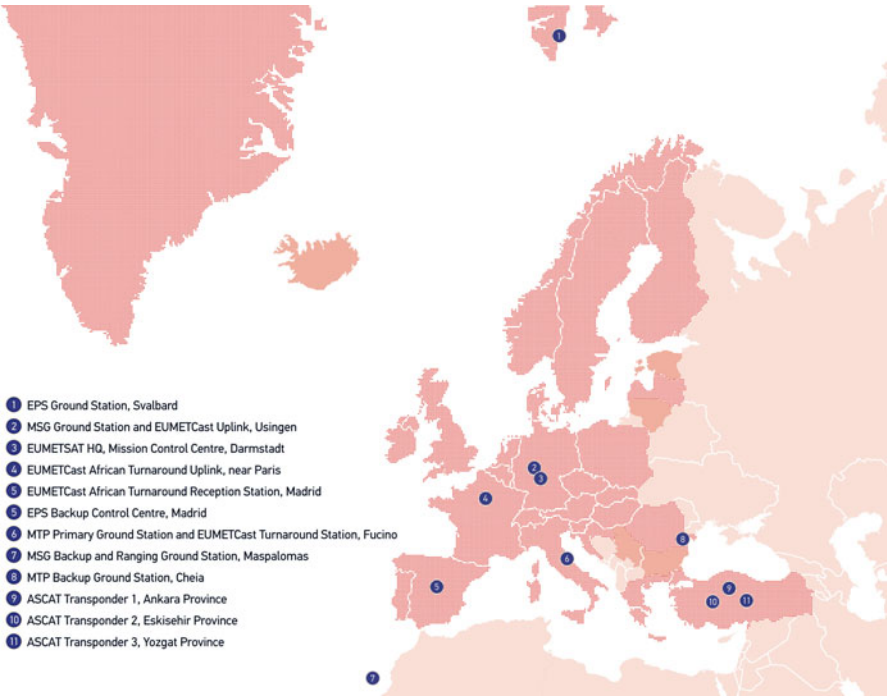


Fig. 38.4 EUMETSAT ground segment

The Ground Stations

A network of dedicated ground stations provides the communication channels between the satellites and the Mission Control Centre (MCC). The ground stations are an essential component of any satellite system and collect from the satellite all the information necessary for the assessment of satellite performance, as well as the scientific data from the instruments and they also relay command messages from the MCC to the satellite (Fig. 38.4).

The *Primary Ground Station* (*PGS*) is located in Usingen, Germany, around 30 km north of Frankfurt. It provides the primary interface between the satellites and the Mission Control Centre (MCC), including all ranging functions and communications lines. The PGS is unmanned and can be remotely monitored and controlled from the MCC in Darmstadt. The Backup Satellite Control Centre (BSCC), which can assume the functions of the MCC in the case of emergencies, is also located at Usingen.

The prime transmission channel between the MCC and the PGS is a 34 Mbit/s microwave link with a terrestrial-based back-up link. To accomplish these vital tasks to the required reliability standard, a considerable amount of redundancy is incorporated in the station design, which, to a great extent, can function completely automatically.

Three fully steerable 13-m diameter parabolic antennas are located at the PGS and used exclusively to support all communications with the MSG satellites. Each

antenna is capable of supporting all transmissions and data reception required for one spacecraft and is used for telemetry and telecommands, raw image reception, collection of Data Collection System reports and for low-rate direct dissemination.

A *Backup and Ranging Ground Station* (*BRGS*) is located in Maspalomas, Gran Canaria Island, Spain. This location is sufficiently separated from the PGS to allow accurate ranging measurements to be made to determine the precise location and orbit of the MSG satellites. The BRGS is also dedicated to provide telecommanding and telemetry support to the ground network for the operations of the satellites in case of a complete system failure at the PGS. A further backup station is located at Cheia, Romania, to overcome outages of the station in Maspalomas caused by the atmospheric scintillation.

Product Extraction Facilities

As an organization created and funded by the European meteorological community, it is not surprising that EUMETSAT places a very strong emphasis on developing products, based on the satellite images and data, which are of direct benefit to its users. Consequently, a comprehensive range of meteorological and geophysical products is derived by EUMETSAT's Application Ground Segment, both at its headquarters in Darmstadt and through the distributed network of Satellite Application Facilities (SAFs) (www.eumetsat.int/home/main/satellites/MeteosatSecondGeneration).

All of these products are extracted on a fully automated basis requiring a minimum of human intervention and are distributed via the EUMETSAT data dissemination system, EUMETCast, the Internet, and other means.

The product extraction mission is performed within the MSG Ground Segment by a combination of the Meteorological Products Extraction Facility (MPEF) in Darmstadt and the network of Satellite Applications Facilities (SAFs) distributed around the EUMETSAT Member States. In addition to the products that are based on MSG data, both the MPEF and the SAFs also make extensive use of the data available from EUMETSAT's Metop polar-orbiting satellite.

The product extraction processing uses the level 1.5 image and supporting data acquired from Image Processing Facility, along with observed and forecast meteorological data acquired from the GTS network if required, to generate a set of products. Data from both the primary satellite data (Meteosat-9) and the Rapid Scanning Satellite (Meteosat-8) are used.

In the case of Meteosat processing, the facility receives near-real-time level 1.5 data from the image processing system, which has performed a geometrical correction and removed imperfections in the images. A radiative transfer model is applied to compute expected radiances at the top of the atmosphere, as well as atmospheric correction tables. Scenes analysis provides information on the surface type in each cloud-free image pixel. This output is then presented to the various meteorological product applications as a classified pixel map and as segmented clustered scenes, and meteorological products are generated.

In addition to the EUMETSAT satellite data, input from sources external to EUMETSAT is required for the processing and verification of some products. Independent meteorological observations are used to verify a subset of the products, and verification results are stored along with the products themselves.

The *Meteorological Product Extraction Facility* (*MPEF*) provides the core functionality of the product extraction mission for an agreed set of meteorological products.

A sample of the products produced from MSG data by the MPEF is:

- Atmospheric motion vectors derived from analysis of cloud motion, yielding information on winds in the atmosphere
- Cloud analysis (including cloud cover, cloud top temperature and pressure)
- Clear sky radiance
- Precipitation index
- Climate data set
- High-resolution precipitation index
- Global stability index, providing a few hours of warning for potentially strong convective storms
- Fire monitoring

A product that came into prominence in April 2010 was the Volcanic Ash Detection product. The immense disruption to European air traffic caused by the eruption of the Eyjafjallajökull volcano in Iceland demonstrated the usefulness of the product and prompted efforts to enhance it.

The network of *Satellite Application Facilities* (*SAFs*) supplement the core set of centrally generated MSG products. The SAFs are located at the National Meteorological Services of EUMETSAT Member States. Utilizing specialist expertise from the Member States, SAFs form an integral part of the distributed EUMETSAT Application Ground Segment. Each SAF is a center of expertise along a specific theme in meteorological applications (Fig. 38.5).

Each SAF is led by the National Meteorological Service (NMS) of a EUMETSAT Member State in association with a consortium of EUMETSAT Member States and Cooperating States, government bodies, and research institutes. The lead NMS is responsible for the management of each complete SAF project. The research, data, and services provided by the SAFs complement the standard meteorological products delivered by the MPEF in Darmstadt.

EUMETSAT supervises and coordinates the overall activities of the SAF network and the integration of the SAFs into the various operations within the EUMETSAT Application Ground Segment. It manages and coordinates interfaces – between the SAFs themselves and between SAFs and other EUMETSAT systems – overseeing the integration of SAFs into the overall ground segment infrastructure.

The SAFs help deliver a variety of benefits including:

- Improvements to short-range forecasting of severe weather hazards
- Support to the aviation, agriculture, construction, gas, water, and electricity industries
- Better understanding of the causes and effects of pollution of the upper atmosphere and the depletion of ozone

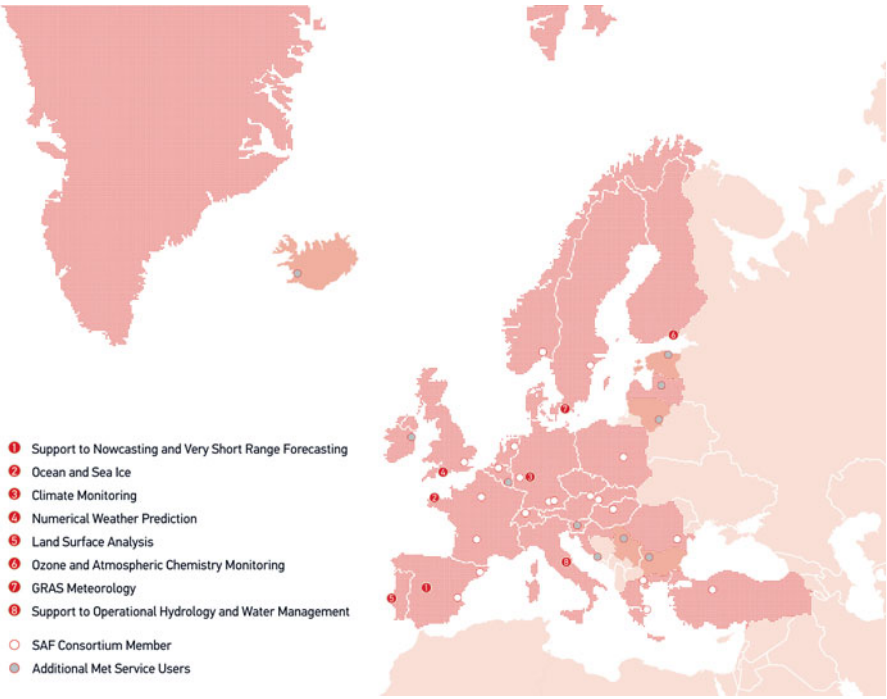


Fig. 38.5 Network of EUMETSAT satellite application facilities

- Early warning of hazards
- Better data for climate monitoring
- Improved information for land use, ecology, disaster monitoring, and agricultural forecasting
- Benefits for sea transport, fishing and offshore industries
- Improved data for input to Numerical Weather Prediction and the availability of user software packages for operational applications
- Improved software packages and near-real-time and offline products

There are currently eight SAFs providing products and services on an operational basis:

– *Support to Nowcasting and Very Short Range Forecasting SAF,* hosted by AEMET, Spain. The main goal of the NWC SAF is to produce software packages, for local installation at the user's site, that support Nowcasting and Very Short Range Forecasting.

– *Ocean and Sea Ice SAF*, hosted by Météo-France. The OSI SAF is an answer to requirements from the meteorological and oceanographic communities of EUMETSAT Member and Cooperating States for comprehensive information derived from meteorological satellites at the ocean–atmosphere interface. The OSI SAF offers a valuable complement to in situ data, based on continuously increasing temporal and geographical resolution products from coastal to global coverage.

- *Climate Monitoring SAF*, hosted by Deutscher Wetterdienst, Germany. The CM SAF generates and archives high-quality datasets for specific climate application areas, through the exploitation of satellite measurements with state-of-the-art algorithms to derive information about the climate variables of the Earth system. It aims to provide data that can be further used to assess the current climate, e.g., for infrastructure planning, to assess the climate variability and change including climate change detection and attribution, to support the development of climate models by validating long-range and short-term climate forecasts, to assess the impact of changing environment, and to provide evidence for policy actions.
- *Numerical Weather Prediction SAF*, hosted by the United Kingdom Met Office. The NWP SAF exists to develop techniques for more effective use of satellite data in numerical weather prediction models. To achieve this, the NWP SAF updates, assesses, and prioritizes user requirements and develops the satellite data processing modules needed to meet those requirements. These include preprocessing, retrieval and assimilation modules, modules for monitoring, tuning, and quality control, and modules for validation of satellite products and of observation operators. It also monitors the quality of many satellite data streams and makes the results available on the web.
- *Land Surface Analysis SAF*, hosted by Instituto Meteorologia, Portugal. The aim of the LSA SAF is to take full advantage of remotely sensed data relating to land, land–atmosphere interactions, and biosphere applications; a strong emphasis is put on developing and implementing algorithms that will allow an operational use of data from EUMETSAT satellites. The LSA SAF system generates, archives, and disseminates, on an operational basis, a set of parameters involved in the surface radiation budget, snow, vegetation cover, evapotranspiration, and fire-related products.
- *Ozone and Atmospheric Chemistry Monitoring SAF*, hosted by the Finnish Meteorological Institute. The O3M SAF produces, archives, validates, and disseminates ozone and atmospheric chemistry products to support the services of the EUMETSAT Member States in weather forecasting as well as monitoring of ozone depletion, air quality, and surface UV radiation.
- *GRAS Meteorology SAF*, hosted by the Danish Meteorological Institute. The GRAS SAF uses data from the GRAS (Global Navigation Satellite System Receiver for Atmospheric Sounding) instrument on the polar-orbiting Metop satellite to generate high-quality GPS Radio Occultation (RO) datasets for Numerical Weather Prediction (NWP) applications and specific climate application areas.
- *Support to Operational Hydrology and Water Management SAF*, hosted by USAM, Italy. The H-SAF generates high-quality data sets and products for operational hydrological applications from satellite data. Precipitation and soil moisture products are among its outputs.

Data Collection and Data Dissemination

The Meteosat Data Collection Service (DCS) enables environmental data to be collected via the Meteosat satellite from Data Collection Platforms (DCP), such as

ground or air-based observing stations which can be located anywhere in the field of view of that satellite (where their line of sight to the satellite is at least 5° above the horizon). The DCS is particularly useful for the collection of data from remote and inhospitable locations where a DCP may provide the only possibility for data relay.

A particular example of the usefulness of the service is its role in the Indian Ocean Tsunami Warning System, whereby the DCS on EUMETSAT's Indian Ocean Data Coverage satellite contributes to relaying data from some 52 DCPs to the Pacific Tsunami Warning Center in Hawaii, which issues tsunami alerts.

The DCP data is processed at the EUMETSAT Control Centre and routed for further dissemination either via EUMETCast, the WMO Global Telecommunication System(GTS), the Internet, or by Direct Broadcast via the MSG satellite (at 0° longitude).

The data are transmitted at 100 bps for a Standard Rate platform and 1,200 bps for a High Rate platform.

DCPs operate in two main bands – international and regional:

International DCPs are characterized by the possibility that they can move from one field of view of one satellite to that of another. They have to operate in the "self-timed" mode, presently defined as a 1.5 min repeat cycle per transmission, including a safety margin of 15 s at the beginning and the end. Eleven channels are available for this type of DCP. The bandwidth allocated for those International DCPs is 402.0355–402.0655 MHz.

Regional DCPs remain in the coverage area of one satellite. There are three types of DCP – Self-timed, Alert, and Hybrid (self-timed and alert). The "Alert mode" is to transmit short messages on a dedicated alert channel if one or more meteorological parameters exceed predefined thresholds. The bandwidth allocated for those Regional DCPs is 402.0685–402.4345 MHz, which can be used to support Standard Rate DCPs and High Rate DCPs. The bandwidth is divided into segments of 3 KHz for the MTP Standard Rate DCP, 2.25 KHz for the MSG High Rate DCP, and 1.5 KHz for the MSG Standard Rate DCP channels.

The EUMETSAT dissemination mission is responsible for the dissemination of data to the user community. Data to be disseminated include the processed image data, generated products, and meteorological data and products from other sources.

EUMETSAT's primary dissemination mechanism for the near real-time delivery of satellite data and products generated by the EUMETSAT Application Ground Segment is the EUMETCast system which also delivers a range of third-party products. Outside the EUMETCast footprint, they are disseminated by the World Meteorological Organization's Regional Meteorological Data Communication Network/Global Telecommunications System.

EUMETCast is a multiservice dissemination system which uses the services of a commercial satellite operator and telecommunications provider to distribute data files using Digital Video Broadcast (DVB) technology. One of the strengths of the system is the simplified user infrastructure and the resulting low cost for obtaining high-quality data. All available data can be received with a single reception station using off-the-shelf components.

The key features of EUMETCast are:

- Secure delivery allows multicasts to be targeted to a specific user or group of users thus supporting any required data policy
- Handling of any file format, allowing the dissemination of a broad range of products
- Use of DVB turnarounds allows the easy extension of geographical coverage
- One-stop-shop delivery mechanism allows users to receive many data streams via one reception station

There are currently over 3,000 EUMETCast user stations in operation.

EUMETCast is part of a global monitoring system. It is Europe's contribution to GEONETCast, the Group on Earth Observations' network of dissemination systems. Within the Global Earth Observation System of Systems (GEOSS), EUMETCast is a contributing system for global data dissemination and is also available for use by the European Global Monitoring for Environmental and Security (GMES) initiatives and other environmental data providers. EUMETCast is also a EUMETSAT contribution to the Integrated Global Data Dissemination Service (IGDDS), a component of the World Meteorological Organization (WMO) information service.

The Meteosat data disseminated via EUMETCast can be grouped into several classes, which are described here in order of priority for their dissemination:

- The level 1.5 image and supporting data, as generated by the imaging mission; they constitute the highest priority data and are expected to be disseminated within 5 min of the reception of the raw image data on the ground.
- Level 2.0 and 3.0 products (from the MPEF and SAFs), plus any supplementary data constitute the second priority data. The products to be disseminated comprise a predefined set of key products and a configurable set of additional products.
- "Foreign" satellite data. This category comprises data from other meteorological satellite systems for distribution within the MSG dissemination. These may include data from other geostationary satellites and from polar-orbiting systems, with particular regard to data from satellites overlapping in their field of view with MSG.
- GTS type data, including messages from DCPs and meteorological data service. The required retransmission responsivity of DCP messages will depend upon the criticality type of the message.

The data are broadcast via EUMETCast to users operating a receiving station in the field of view of commercial communication satellites according to the coverage agreed by EUMETSAT Member States. Both HRIT and LRIT data are made available.

The direct dissemination of LRIT data via the MSG satellite (at 0° longitude) is also available for those users who are equipped with dedicated LRIT stations in the field of view of the MSG satellite (at 0° longitude).

High Rate Information Transmission (HRIT) and Low Rate Information Transmission (LRIT) are the CGMS standards agreed upon by satellite operators for the dissemination of digital data originating from geostationary satellites to

users via direct broadcast. The distinction between the two standards, as their names suggest, is the data rate (bandwidth) necessary to convey the data content. LRIT data (mainly the meteorological products, a reduced set of images at low resolution) are typically disseminated at low bandwidth (typically 128 Kbps for direct dissemination via the MSG satellite at 0° longitude) while HRIT data (mainly the full earth and rapid scan images in full resolution) are typically disseminated at speeds up to 1 Mbps.

The EUMETSAT Data Centre

The EUMETSAT Data Centre stores all the organization's satellite data and derived products securely and helps users access the archived data. The Data Centre supports the data requirements of external users such as the National Meteorological Services, research organizations, universities, and commercial companies. It also supports internal users, like the Meteorology Division or Operations facilities which need the data for purposes such as research and reprocessing (www.eumetsat.int/home/main/satellites/MeteosatSecondGeneration).

The Data Center fulfills the following functions:

- Acquisition and archiving of Level 1.0 and 1.5 Images
- Acquisition and archiving of the MPEF Products
- Generation and maintenance of catalogs covering the archived data sets plus the SAF catalog information
- Provision of an online catalog query and product retrieval service (the EUMETSAT Portal) using a device known as the Product Navigator

User access to, and orders, of EUMETSAT data from the Data Centre are increasing continually. In 2009, over 1 PB was retrieved from the Data Centre in response to user ordering. The rate of retrieval of data clearly demonstrates that the Data Centre is more than a safe data store.

To cope with the growing access demands from users and new data streams from satellite missions planned in the future, the systems in the Data Centre are continually upgraded.

The Data Centre also hosts and operates the EUMETSAT Global Space-based Inter-Calibration System (GSCIS) collaboration server. GSCIS is an international collaborative effort to examine and harmonize data from operational weather satellites to improve climate monitoring and weather forecasting.

In addition to delivering real-time weather information over decades, Meteosat satellites have also been a source of observations relevant for climate and environmental monitoring with records dating back to 1981. These records now constitute a valuable dataset that is of great interest to climate scientists. These observations are used to analyze climate processes, climate variability, and climate change.

The first and second generations of Meteosat provide long series on many important Essential Climate Variables defined by the Global Climate Observing System (GCOS) which is under the aegis of WMO and other international bodies. This information is a great resource for climate studies, and the EUMETSAT Data

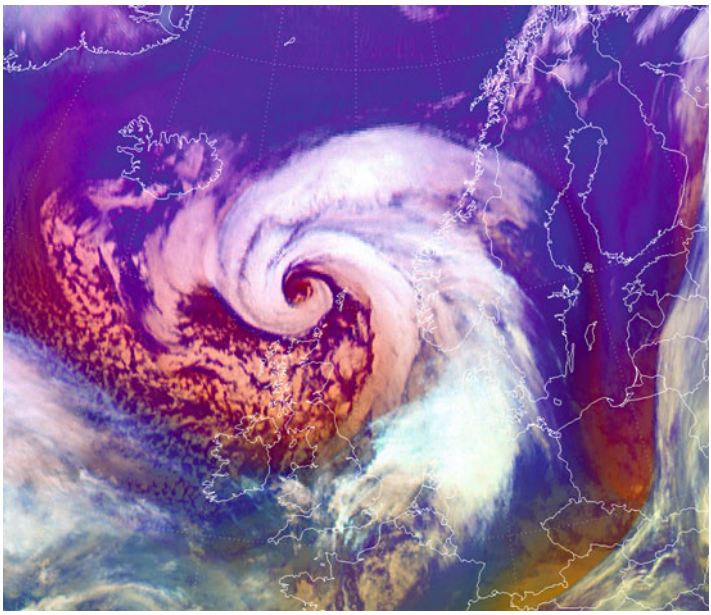


Fig. 38.6 Image of Storm "Tuva," which developed over the North Atlantic following rapid cyclogenesis, from Meteosat-9, 31 January 2008

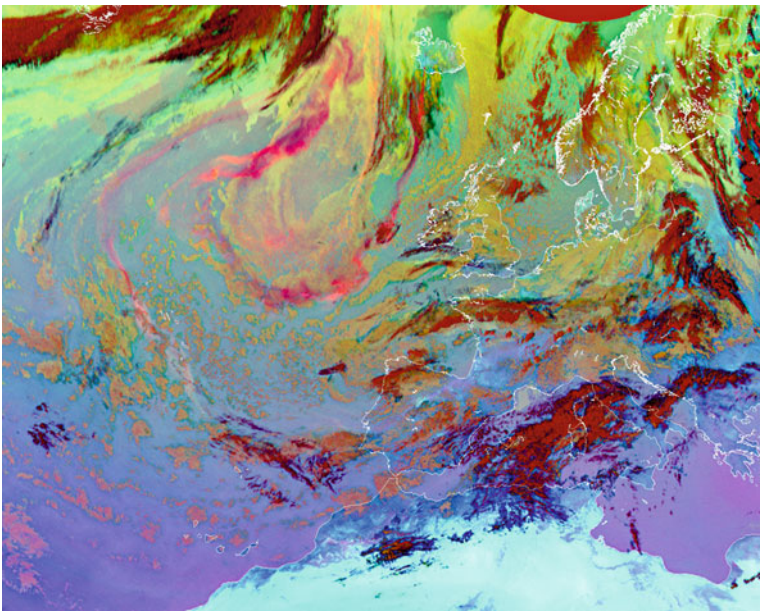


Fig. 38.7 Volcanic ash cloud (*seen in reddish hues*) emanating from the Eyjafjallajökull volcano in Iceland, detected from the rapid scanning service of Meteosat-8, 10 May 2010

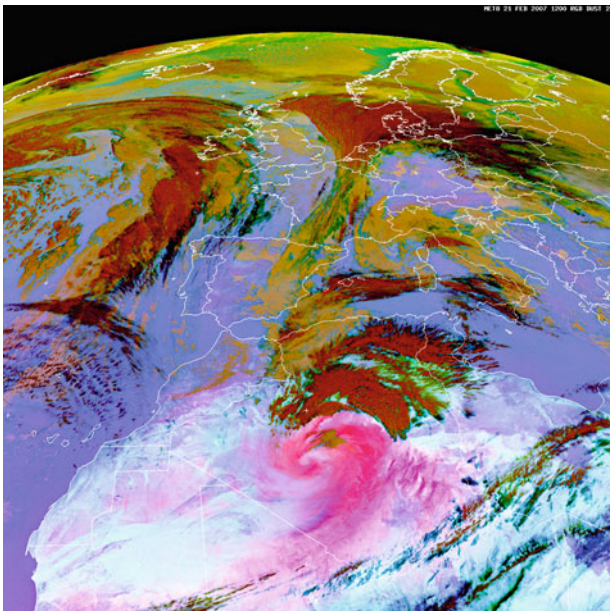


Fig. 38.8 Dust storm over North Africa, from Meteosat-8, 21 February 2007

Centre provides an efficient means of access to the data. The Earth's radiation budget is a major topic of study in relation to climate change and outputs derived from the Meteosat series, such as surface albedo, cloud properties and atmospheric humidity, are very relevant in this respect. MSG's Geostationary Earth Radiation Budget (GERB) mission provides additional valuable data on reflected solar radiation and thermal radiation emitted by the Earth and atmosphere and represents another valuable contribution by EUMETSAT.

The existing first-generation Meteosat Visible and Infrared Imager (MVIRI) and MSG's SEVIRI and the future MTG's Flexible Combined Imager will extend the Meteosat data record to 50 years in 2032, longer than most other satellite records (Figs. 38.6–38.8).

Meteorological Third Generation (MTG)

In June 2010, the Council of EUMETSAT approved the content of the Meteosat Third Generation (MTG) Program that will eventually supersede the MSG and give continuity of European meteorological geostationary satellite coverage until at least 2040.

As would be expected, the MTG will provide improved coverage in several ways. Breaking with tradition, the satellite in the series will have three-axis stabilization rather than spin stabilization. This of itself will ensure that the satellites instruments are constantly pointed to the Earth and enhanced images will result. Another major

departure from previous practice is that the program will utilize two satellites in orbit together to provide full operational coverage. This arises from the decision to include a wider range of instruments in the mission and to split the operational load between two satellites. The additions include an atmospheric sounding capability, a lightning detection mission, and an ultraviolet sounder (Aminou et al. 2009).

In contrast to the current MSG Meteosat satellites, using a 2-t class spacecraft and an imager with 12 spectral channels, the planned Meteosat Third Generation imaging satellite will be a 3-t satellite with 16 nominal spectral channels. The second operational platform will carry out atmospheric sounding to observe the different layers within the atmosphere. The sounder will be one of the key innovations in the new program, allowing Meteosat satellites for the first time not just to image weather systems but to analyze the atmosphere layer by layer and to perform far more detailed chemical composition studies.

As before the program is being established through cooperation between EUMETSAT and the European Space Agency (ESA). A European consortium led by Thales Alenia Space of France will build the MTG spacecraft.

The satellite series will comprise four imaging and two sounding satellites. The imaging satellites, MTG-I, will fly the Flexible Combined Imager (FCI) and an imaging lightning detection instrument, the Lightning Imager (LI). The sounding satellites, MTG-S, will include an interferometer – the Infra-red Sounder (IRS) with hyper-spectral resolution in the thermal spectral domain – and the high-resolution Ultraviolet Visible Near-infrared (UVN) spectrometer that will address the atmospheric chemistry requirements of the European Union's Global Monitoring for Environment and Security (GMES) program in respect of its Sentinel-4 mission.

The first spacecraft is likely to be ready for launch from 2017. At any time the operational requirement will be fulfilled by an in-orbit configuration of two parallel positioned satellites, the MTG-I (imager) and the MTG-S (sounder) platforms. The use of three-axis stabilization will enable more demanding user requirements on spatial resolution, repeat cycle, and signal-to-noise ratio to be met, and is a prerequisite to conduct soundings from geostationary orbit.

The first MTG-I and MTG-S prototypes are being developed by ESA as part of its MTG program. The EUMETSAT MTG program includes the procurement of the four recurrent satellites – three MTG-Is and one additional MTG-S – as well as six launches, the development of the ground segment and the operations of all satellites.

The EUMETSAT ground segment facilities for the MTG will be integrated into the existing multi-mission ground network infrastructure, which is common to first- and second-generation Meteosat and Metop and Jason missions.

Following on from its predecessors, MTG will also provide a Data Collection and Retransmission service to collect and relay environmental data from automated data collection platforms. It will also carry GEOSAR communications payload to relay distress signals to a central reception station in Europe that passes the signals on for quick organization of rescue activities. The geostationary relay allows a continuous monitoring of the Earth disk and immediate alerting.

The MTG data and products generated within the EUMETSAT Application Ground Segment will be made available in near real-time through future evolutions of current delivery mechanisms such as EUMETCast and the WMO Global Telecommunications System (GTS). For MTG, EUMETCast will be upgraded to cope with the greatly increased product throughput rates, without major new technological needs for users and their reception stations. Some terrestrial-based dissemination mechanisms are also likely to be used to deliver special data sets to key users, but the broadcast capability provided by EUMETCast will remain the primary delivery mechanism. The EUMETSAT Data Centre will continue to archive all EUMETSAT including the MTG data and make them available through mechanisms such as the Product Navigator.

The EUMETSAT Application Ground Segment will be upgraded to accommodate the future processing requirements of MTG imager and sounder data. Both the Meteorological Product Extraction Facility (MPEF) and the Satellite Application Facility (SAF) Network will be enhanced to accommodate the processing of MTG data.

The Instruments

Flexible Combined Imager

The Flexible Combined Imager (FCI) on the MTG-I satellite will continue the very successful operation of the Spinning Enhanced Visible and Infrared Imager (SEVIRI) on Meteosat Second Generation (MSG). The satellite's three axes stabilized platform will be capable of providing additional channels with better spatial, temporal, and radiometric resolution compared to the current MSG satellites.

Requirements for the FCI have been formulated by regional and global Numerical Weather Prediction (NWP) and Nowcasting communities. These requirements are reflected in the design which allows for Full Disk Scan (FDS) with a basic repeat cycle of 10 min and a European Rapid Scan Service which covers of one-quarter of the full disk with a repeat cycle of 2.5 min. The FCI takes measurement in 16 channels of which 8 are placed in the solar spectral domain between 0.4 and 2.1 μm delivering data with a 1 km spatial resolution. The additional 8 channels are in the thermal spectral domain between 3.8 and 13.3 μm delivering data with a 2 km spatial resolution. In the rapid scanning mode, there will be two additional channels in the solar domain with a spatial resolution of 0.5 km and two in the thermal domain with a spatial resolution of 1 km.

The additional channels within the solar domain will surpass current aerosol retrievals, including volcanic ash, thereby providing an important contribution to future air quality monitoring. The increased spatial resolution and range of channels will offer improved fire detection products and an increase in the quality of climate relevant products such as fire, radiative energy, and power, which are directly related to carbon dioxide production.

Lightning Imager

The Lightning Imager (LI) will offer improvements for Nowcasting by delivering information on total lightning (Inter Cloud – IC and Cloud to Ground – CG).

The instrument will bring full hemispheric near real-time total lightning detection capabilities.

The benefit of the LI mission is that it will continuously and simultaneously observe total lightning over the hemisphere providing the information to the users with an extremely high timeliness.

It is expected that the Lightning service will combine complementary information on total lightning measured by the Lightning Imager and the surface-based networks measuring global distribution of cloud-to-ground lightning, as for instance measured by the surface-based Arrival Time Difference network (ATDnet), which together should strongly improve the quality of information essential for air flight safety. Each lightning stroke initiated by an electrical discharge in or below clouds causes radiation in the visible spectrum that will be detected by the Lightning Imager on board MTG-I. The LI will measure optical pulses initiated by lightning strokes emitting energy above a set threshold, at a wavelength of 777.4 nm with a spatial resolution of 10 km. The information delivered to the users will be time, position, and intensity of detected optical pulses.

Lightning observations will also benefit climate monitoring. One approach to assess the impact of climate change on thunderstorm activity is to globally monitor and long-term analyze the lightning characteristics, which would require a long-term stable and spatially homogeneous lightning observing system. Lightning is a major source of harmful nitrogen oxides (NOx) in the atmosphere which play a key role in the ozone conversion process and acid rain generation. A detailed knowledge of the global distribution of the total lightning (CG + IC) is a prerequisite for studying and monitoring the physical and chemical processes in the atmosphere regarding NOx.

Lightning observations from the geostationary orbit, delivered with spatially homogenous and well-characterized quality, are specifically suited to support these climate and atmospheric chemistry applications. The LI observations on MTG will complement the NOAA Geostationary Lightning Mapper (GLM) placed on the GOES-R and the GOES-S satellites and the information obtained from the ground detection networks.

Infrared Sounder

The new geostationary sounder service from the Infrared Sounder (IRS) on MTG-S is based upon requests from the NWP community to deliver spectral information and/or retrieved products as horizontal and vertical gradients of humidity and temperature. Making available 3-D information on humidity, temperature, and wind will support Nowcasting applications, particular in situations of water vapor convergence and convective instability, giving improved warnings on location and intensity of convective storms. The deduced information on atmospheric dynamics (e.g., Atmospheric Motion Vectors (AMVs) with a vertical resolution of about 2 km which surpasses the current products) will be invaluable to numerical models used in operational forecasting in the future. In particular, a breakthrough regarding better precipitation forecasts is expected by using this

new information within these advanced models coupled with advanced data assimilation capabilities.

Another example of the expected benefits of the IRS instruments is information on ozone, carbon monoxide, and volcanic ash compositions within the atmosphere. The user community has already identified the crucial role of infrared instruments for future volcanic-ash monitoring. The improved MTG capabilities of the imagery mission will provide further details on the extent of the ash plume, whereas the capabilities of the MTG sounding mission will be essential for the derivation of quantitative products with additional information on the composition and density of the ash cloud.

Retrieving highly resolved vertical structures of humidity (~2 km resolution with 10% accuracy) and temperature (~1 km with 0.5–1.5° accuracy) by remote sensing techniques requires measurements within the water vapor and CO_2 absorptions bands with extremely high spectral resolution and accuracy. The IRS is based on an imaging Fourier interferometer with a hyper-spectral resolution of 0.625 cm^{-1} wave number and a spatial resolution of 4 km. The IRS will deliver over the Full Disk in a total of 1,720 channels in the infrared spectrum, with a basic repeat cycle of 60 min.

Ultraviolet Sounder

The Ultraviolet, Visible and Near-Infrared Sounding (UVN) instrument supports the Global Monitoring for Environment and Security (GMES) Sentinel 4 mission for geostationary chemistry applications. The primary objective of the Sentinel 4 mission is to support air quality monitoring and forecasting over Europe with a high revisit time (~1 h or better). The primary data products will be ozone, nitrogen dioxide, sulfur dioxide, formaldehyde, and aerosol optical depth.

The UVN will fly onboard the MTG-S satellites. Funding for the UVN is provided by the European Commission in cooperation with European Space Agency (ESA).

The UVN is a spectrometer taking measurements in the ultraviolet (UV: 305–400 nm), the visible (VIS: 400–500 nm), and the near infrared (NIR: 755–775 nm) spectra with a spatial resolution of better than 10 km. Its observations are restricted to Earth area coverage from 30° to 65° N in latitude and 30° W to 45° E in longitude. The observation repeat cycle period will be shorter than or equal to 1 h.

ESA is responsible for the definition of the Sentinel 4 mission and provision of the UVN Instrument, whereas EUMETSAT takes responsibility for the operational processing, delivery, and management of the instrument data.

Conclusion

Since EUMETSAT was established, its policy in relation to geostationary satellite programs has been to ensure long-term continuity to the users of its services based on proven technology and on effective exploitation of the satellite data. Building on the success of the original Meteosat series, initiated by the European Space Agency at the urging of the European national meteorological

services, EUMETSAT's Meteosat Second Generation program provides a high-quality imaging facility over Europe, Africa, and the eastern Atlantic, as well as the Indian Ocean Data Coverage. It also supports the scientific and climate mission known as the Global Earth Radiation Budget (GERB).

Considerable effort is expended on extracting additional value from the data. The centralized product extraction facility at Darmstadt produces highly valuable sets of products based on analysis of the images. The innovative Satellite Application Facilities add immense value through their focus on specialized aspects of the data application.

With continuity beyond the MSG an imperative for EUMETSAT, the Meteosat Third Generation (MTG) has been initiated and will provide continuity of European geostationary meteorological satellite programs up to 2040.

Throughout its history, EUMETSAT has fostered international cooperation in satellite meteorology and has forged excellent relationships with other satellite providers in an effort to promote global coverage and to optimize the effectiveness of the investments in the satellite programs. Within Europe it works very closely with the European Space Agency and is taking an active role in the ambitious GMES (Global Monitoring for the Environment and Security) plan of the European Union. It has also addressed the needs of the developing world through various initiatives, such as the Preparation for the Use of MSG in Africa (PUMA) and AMESD (African Monitoring of the Environment for Sustainable Development).

The general direction that future developments in relation to EUMETSAT geostationary meteorological programs will take is to a large extent already evident in the agreed content of the MTG program. New satellite technology accompanied by enhancement of existing design solutions (e.g., three-axis stabilization rather than the spin stabilization) will enable a wider range of instruments of greater sophistication and higher spectral resolution at the level of affordability comparable to the past programs.

Some other factors which dictate the trends affecting EUMETSAT programs are the need to utilize technology and instruments that are well proven rather than experimental because of the operational dependence on them. Another factor, evident, for example, in the cooperation between EUMETSAT and the EU GMES Program on MTG (and the next generation of the EUMETSAT polar-orbiting systems), is the broadening of meteorological satellite programs to include a wider range of Earth observation. Within meteorology, the mathematical models use an increasingly diverse range of data sources including those derived from ocean and land observation. Such cross-over usage of data is also likely to become evident in other disciplines. This kind of development is likely to become a stronger feature as more and more operational systems become reliant on space-based observation. In Europe, this trend may be particularly strong as the full GMES program is rolled out.

The benefits of international cooperation are likely to become even more evident. The close cooperation between EUMETSAT and NOAA is an outstanding example of this at present, as demonstrated by the collaboration on

geostationary and polar-orbiting programs. The growing role of other operators such as China, Japan, India, and South Korea are likely to lead to an increased level of global coverage and to closer collaborative links.[1]

Cross-References

► International Meteorological Satellite Systems

Notes

1. An agreement is in place between EUMETSAT and NOAA on what is called the Initial Joint Polar-orbiting operational Satellite (IJPS) System. This program includes two series of independent but fully coordinated EUMETSAT and NOAA satellites and the exchange of instruments and global data, cooperation in algorithm development, and near real-time direct broadcasting. Under the terms of the IJPS agreement, NOAA provides NOAA-18 and NOAA-19 satellites for flight in the afternoon (PM) orbit and EUMETSAT provides MetOp-A and MetOp-B satellites for flight in the mid-morning orbit (AM). The IJPS Agreement is complemented by a Joint Transition Activity Agreement, covering the Metop-C satellite on the European side and the NOAA (NPOESS) Preparatory Project (NPP) satellite, launched on 28 October 2012, and JPSS-1 satellite on the U.S. side. These satellites carry a common core of instruments (see ► Chap. 26, "Introduction and History of Space Remote Sensing"). With NPP, NOAA will continue to share complementary data and information on weather and climate from low earth orbit with EUMETSAT to maximize each agency's investments in space and EUMETSAT will disseminate NPP data in Europe. NOAA and EUMETSAT are also discussing the continuation of their cooperation, under a Joint Polar System Agreement, which will cover the upcoming NOAA JPSS-2 and follow-on satellites and the EUMETSAT EPS SG satellites. For further information on EUMETSAT please visit www.eumetsat.int

References

D. Aminou et al., Meteosat third generation (MTG) status of space segment definition, in *2009 EUMETSAT Meteorological Satellite Conference*, Oslo, 2009, http://www.eumetsat.int/Home/Main/AboutEUMETSAT/Publications/ConferenceandWorkshopProceedings/

ESA MSG Website, www.esa.int/esaMI/MSG

EUMETSAT Website, www.eumetsat.int/home/main/satellites/MeteosatSecondGeneration

J. Schmetz et al., An introduction to Meteosat second generation (MSG). Bull. Am. Meteor. Soc. **83**, 977–992 (2002)

J. Schmid, The SEVIRI instrument, in *2000 EUMETSAT Meteorological Satellite Conference*, Bologna, 2000, http://www.eumetsat.int/Home/Main/AboutEUMETSAT/Publications/ConferenceandWorkshopProceedings/

International Meteorological Satellite Systems

39

Sergio Camacho-Lara, Scott Madry, and Joseph N. Pelton

Contents

Introduction 1022
The World Weather Watch Programme 1023
China: The Fengyun Meteorological Satellite System 1026
India: The INSAT Satellite System 1028
The Japanese Geostationary Meteorological Satellite System (Himawari) and the QZSS Network 1030
The Russian Geostationary Operational Meteorological Satellite (GOMS) and Polar-Orbiting Meteorological (Meteor) Systems 1033
South Korea's Communication, Ocean, and Meteorology Satellite (COMS) 1038
Cross-References 1040
Notes 1040

Abstract

The oldest and most extensive meteorological satellite systems are those of the United States and of Europe, as operated by the Eumetsat system. These are addressed in detail in the preceding two chapters. This chapter describes the meteorological satellite systems of China, India, Japan, Russia, and South Korea. These meteorological satellite systems are extensive and provide a number of sophisticated meteorological satellite sensing capabilities both from

S. Camacho-Lara (✉)
Centro Regional de Enseñanza de Ciencia y Tecnología del Espacio para América Latina y el Caribe (CRECTEALC), Luis Enrique Erro 1, Tonantzintla, Puebla, Mexico
e-mail: sergio.camacho@inaoep.mx

S. Madry
International Space University, Chapel Hill, North Carolina, USA
e-mail: Scottmadry@mindspring.com

J.N. Pelton
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: peltonjoe@gmail.com

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_56, © Springer Science+Business Media New York 2013

geostationary and polar orbiting satellite systems. Today all of these various satellite systems – those of China, Europe, India, Japan, Russia, South Korea, and the United States are in various manners linked together and share data. This international coordination of meteorological data is accomplished through the World Weather Watch (WWW) program of the World Meteorological Organization and the Coordination Group for Meteorological Satellites (CGMS).

These international cooperative efforts – supplemented by bilateral or regional agreements – allow a degree of standardization with regard to the formatting and display of meteorological data and a systematic process for sharing of vital weather data. This sharing of meteorological data is important on an ongoing basis – but this can be particularly important – when there is a failure of a meteorological satellite, a launch failure, or a delay in the deployment of a replacement satellite. In some cases, countries such as the United States have even "loaned" meteorological satellites to other countries when failures or launch delays have created gaps in critical coverage areas.

The various international satellites around the world that are deployed in different orbital locations and with varying periodicity provide a very useful redundancy of coverage that is particularly important in tracking major storms and obtaining the most up-to-date information of atmospheric, oceanic, and of arctic conditions.

This chapter provides a description of the meteorological satellite systems of China, India, Japan, South Korea, and Russia and their current status. Researchers can also consult the various universal reference locations (i.e., URLs) for these various meteorological satellite systems which can be useful in obtaining the more recent information about the deployment and operation of these systems.

Keywords

China Meteorological Administration (CMA) • Geostationary Operational Meteorological Satellite (GOMS) • Elektro Satellites of Russia • Fengyun Meteorological Satellite System of China • INSAT System of India • Himawari System of Japan • Japanese Geostationary Meteorological Satellite (GMS) Systems • Communications, Ocean, and Meteorological Satellite (COMS) of South Korea • Meteor Satellites of Russia • MTSAT of Japan • World Meteorological Organization (WMO) • World Weather Watch (WWW)

Introduction

Weather is of vital interest to people, as it affects agriculture, industry, transportation, and many of our daily life activities. Violent weather sometimes threatens our safety and even our lives when extreme meteorological events come upon us with little warning, such as hurricanes, typhoons, ice storms, tornados, and tropical and winter storms, and affect the areas where we live. There are also weather events that are longer time in the making. These are flooding and drought events that can be due to periodic phenomena like El Niño or La Niña or may be due to weather patterns caused by climate change.

For thousands of years, people have been observing weather patterns to determine when to plant, travel, store food, and even how to use the acquired knowledge for strategic military advantage. These observations, rudimentary at the beginning, gave rise to the discipline called meteorology, which at first was based on records of in situ obtained data such as temperature, precipitation, wind speed, direction, and solar radiation.

To predict the weather, modern meteorology depends upon the acquisition of in situ and space-obtained data and on near instantaneous exchange of weather information across the entire globe. To better understand the global climate system and to anticipate its future evolution, not only do we need global data sets, we must also coordinate climate analysis, modeling, and predictions on an equally global basis in order to establish the correct climate state and to create powerful modeling tools for climate prediction.

The World Weather Watch Programme

The World Weather Watch (WWW) was established in 1963 and this activity represents the core of the WMO's combined observing system. The WWW includes the telecommunication facilities, data processing, and forecasting centers that are operated by its Members. The WWW, thus, serves to make available meteorological and related environmental information to all countries. The WWW is a unique achievement in international cooperation. In few other fields of human endeavor has there ever been such a truly worldwide operational system to which virtually every country in the world contributes for the common benefit of humankind and does so every day of the year for decades on end. These arrangements, as well as the operation of the WWW facilities, are coordinated and monitored by WMO with a view to ensuring that every country has available all of the information it needs to provide weather services on a day-to-day basis as well as for long-term planning and research.[1]

An increasingly important part of the WWW Programme provides support for developing international cooperation related to global climate and other environmental issues, and to sustainable development.

As a result of the implementation of the WWW, a significant number of projects and activities have evolved from the need to efficiently coordinate efforts and to make the data acquired and information derived from satellite data available to those who need them, or can use them to further knowledge of weather-related phenomena.

The Global Observing System (GOS) provides observations of the state of the atmosphere and ocean surface for the preparation of weather analyses, forecasts, advisories, and warnings, for climate monitoring and environmental activities. It is operated by National Meteorological Services and national or international satellite agencies. Developing the space-based part of the GOS is one of the main components of the WMO Space Programme.[2]

The space-based Global Observing System includes four operational near-polar-orbiting satellites and six operational geostationary environmental observation

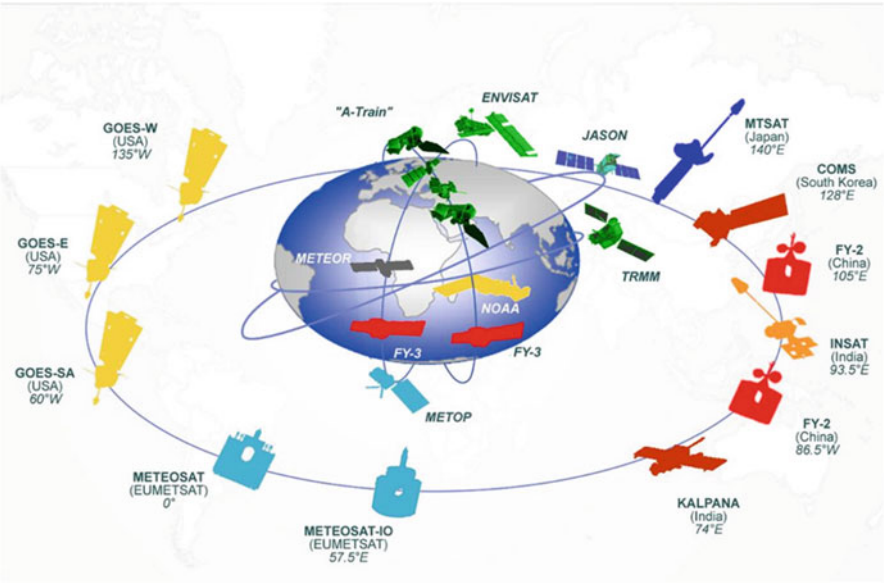


Fig. 39.1 The space-based component of the global observing system (Courtesy WMO)

satellites as well as several research and development satellites. These systems are shown in Fig. 39.1.

Polar orbiting and geostationary satellites are normally equipped with visible and infrared imagers and sounders, from which many meteorological parameters can be derived. Several of the polar-orbiting satellites are equipped with sounder instruments that can provide vertical profiles of temperature and humidity in cloud-free areas. Geostationary satellites can be used to measure wind velocity in the tropics by tracking clouds and water vapor. Satellite sensors, communications, and data assimilation techniques are evolving steadily so that better use is being made of the vast amount of satellite data. Improvements in numerical modeling, in particular, have made it possible to develop increasingly sophisticated methods of deriving the temperature and humidity information directly from the satellite radiances. Research and Development (R&D) satellites comprise the newest constellation in the space-based component of the GOS. R&D missions provide valuable data for operational use as well as for many WMO supported activities. Instruments on R&D missions either provide data not normally observed from operational meteorological satellites or improvements to current operational systems.

GOS also includes solar radiation observations, lightning detection, and tide-gauge measurements. In addition, wind-profiling and Doppler radars are

proving to be extremely valuable in providing data of high resolution in both space and time, especially in the lower layers of the atmosphere. Wind profilers are especially useful in making observations at times between balloon-borne soundings, and have great potential as a part of integrated networks. Doppler radars are used extensively as part of national, and increasingly of regional networks, mainly for short range forecasting of severe weather phenomena. Particularly useful is the Doppler radar capability of making wind measurements and estimates of rainfall amounts.

While programmatic coordination is done through the WWW, coordination regarding compatibility and complementarity among polar-orbiting and geostationary meteorological satellites is done through the Coordination Group for Meteorological Satellites.

The Coordination Group for Meteorological Satellites (CGMS) came into being on September 19, 1972, when representatives of the European Space Research Organization (since 1975 called the European Space Agency, ESA), Japan, the United States of America, Observers from the World Meteorological Organization (WMO), and the Joint Planning Staff for the Global Atmosphere Research Programme met in Washington to discuss questions of compatibility among geostationary meteorological satellites.

CGMS provides an international forum for the exchange of technical information on geostationary and polar orbiting meteorological satellite systems. It consists of 15 member organizations and two observers. The members of the CGMS are China Meteorological Administration (CMA), Centre National d'EtudesSpatiales (CNES), China National Space Administration (CNSA), European Space Agency (ESA), EUMETSAT, India Meteorological Department (IMD), Intergovernmental Oceanographic Commission/UNESCO (IOC/UNESCO), Japan Aerospace Exploration Agency (JAXA), Japan Meteorological Agency (JMA), Korea Meteorological Administration (KMA), National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), Russian Federal Space Agency (ROSCOSMOS), Russian Federal Service for Hydrometeorology and Environmental Monitoring (ROSHYDROMET), and World Meteorological Organization (WMO). EUMETSAT has run the Secretariat since 1987.

The CGMS Secretariat represents CGMS Members in a number of other international coordination bodies such as the Committee on Earth Observation Satellites (CEOS) and its related Earth Observation International Coordination Working Group (EO-ICWG), the Group on Earth Observation (GEO), and the Space Frequency Coordination Group (SFCG).

The meteorological polar-orbiting and geostationary satellites are provided by States through their national and international organizations. The meteorological systems of the United States and Europe (through EUMETSAT) are presented in previous chapters. The sections that follow present the meteorological polar-orbiting and geostationary satellite systems of China, India, Japan, South Korea, and Russia.

China: The Fengyun Meteorological Satellite System

The China Meteorological Administration (CMA) has responsibility for weather forecasting and monitoring of meteorological conditions including those related to climate change. The National Satellite Meteorological Center (NSMC), which is affiliated to the CMA, is responsible for the operation of China's meteorological satellite network.

China has a number of international agreements to obtain meteorological information from other countries' satellite systems that are coordinated through the World Meteorological Organization. Nevertheless, for strategic reasons, China has been for some years implementing a fully functional national satellite system of its own that is known as the Fengyun Meteorological Satellite System. Fengyun means "wind" and "cloud." The Fengyun I series has been fully deployed since September of 1988 with the launch of FY-1A followed by FY-1B in September of 1990. This satellite network is a polar-orbiting meteorological system. This was then followed by the Fengyun II series which was fully deployed as of early 2012. The Fengyun II is a geostationary-orbiting satellite system. The last in the Fengyun II series of geostationary satellites the FYII-7, renamed as FY-2F, was successfully launched on a Long March 3 launch vehicle on January 13, 2012 and subsequently placed at 112°E above the Equator.[3] Figure 39.2 shows the launch of FYII-7.

The CMA is now in the process of implementing the Fengyun III meteorological satellite series. This is an upgraded polar-orbiting meteorological satellite network that constitutes the replacement for the initial Fengyun I series.[4]

According to the China Meteorological Administration, the Fengyun III satellite series will have a number of expanded capabilities over the Fengyun I series. The defined objectives for the Fengyun network, once it is deployed in orbit, will be as follows:

- To provide global measurements of temperature gradients in three dimensions
- To collect moisture soundings of the atmosphere, and thereby to measure cloud and precipitation parameters in support of Numerical Weather Prediction (NWP)
- To provide global imagery of large-scale meteorological and/or hydrological events and biosphere environment anomalies (by integrating imaging from FY-III and FY-II satellites)
- To provide geophysical parameters in support of global meteorological change and climate monitoring
- To provide global and local meteorological information for specialized meteorological users working in such areas as aviation, marine transportation, and fishing
- To collect and relay environmental data from the ground segment to national and international users and scientific analysts

The FY III operational network will initially consist of two polar-orbiting satellites. These satellites will be deployed in phased orbits so that one of the satellites provides coverage in the daytime (AM) and the other in the evening (PM). The payload will be different for AM/PM satellites with sensors optimized for operation in the sunlight and for the one designed for nighttime operations.



Fig. 39.2 The launch of the geostationary Fengyun II-7 satellite in January 2012 (Graphics courtesy of Chinese National Space Agency)

The appropriate time slots for the AM and PM satellite operations have been coordinated through the World Meteorological Organization (WMO) (FY-3 (Fengyun 3) 2nd Generation Polar-Orbiting Meteorological Satellite Series https://directory.eoportal.org/get_announce.php?an_id=12759),

The FY-III series began with a developmental phase. In this developmental phase, the FY-IIIA satellite (launched on May 27, 2008) and the FY-IIIB satellite (launched November 4, 2010) gathered experience but with sounders that had limited capabilities. Table 39.1 shows the dates of launch and projected launch and the type of mission of the FY-III series of satellites.

The second generation of China's polar-orbiting meteorological satellite (FY-3), with a three-axis stabilization mode, carries 11 observation sensors and provides the functions of global, all-weather, multispectral, three-dimensional and quantitative Earth observations. A description of the sensors, as well as access to the data, is provided by the Fengyun Satellite Data Center.[5]

Table 39.1 Overview of Fengyun-3 spacecraft series of CMA/NSMC

Spacecraft	Launch (Projected launch)	LTDN (Local Time on Descending Node)	Mission service type
FY-IIIA	*May 27, 2008*	10:00 h	R&D (Experimental)
FY-IIIB	*Nov. 04, 2010*	14:00 h	R&D (Experimental)
FY-IIIC	2012	10:00 h	Operational
FY-IIID	2014	14:00 h	Operational
FY-IIIE	2016	10:00 h	Operational
FY-IIIF	2018	14:00 h	Operational

India: The INSAT Satellite System

The Indian Space Research Organization (ISRO) has deployed INSAT meteorological satellites for three decades. The first INSAT 1-A satellite was launched on April 10, 1982. This spacecraft was a hybrid geostationary satellite that was capable of providing telecommunications services but it also contained a meteorological package. Following a misfunction in INSAT 1-A, an identical INSAT 1-B was launched on August 30, 1983. On April 3, 1999 the INSAT 2E, an upgraded hybrid communications and meteorological satellite was also launched. Up to this point all of the satellites deployed by India for meteorological sensing had been hybrid telecommunications satellites that also included a meteorological package. Further all of these satellites had been supplied by overseas suppliers.

On September 12, 2002 the Indian Space Research Organization launched a dedicated meteorological satellite that had been designed and manufactured within ISRO. This satellite was initially named Metsat but was renamed Kalpana-1 in 2003 in the honor of the Indian-born American Astronaut Dr. Kalpana Chawla who perished in the Columbia Shuttle accident. This satellite – like all of the spacecraft in the INSAT satellite series – was launched into geostationary orbit. This Indian designed and manufactured spacecraft was also unique in that it was launched on the Indian Polar Satellite Launch Vehicle in its first launch and successful mission. This dedicated Kalpana-1 satellite contains a very high resolution radiometer (VHRR) that operates in the visible bands as well as the thermal and water-vapor infrared bands. It also contains a data relay transponder to transmit data to a number of ground locations (Fig. 39.3).

On April 28, 2003 the INSAT 3A, placed at 93.5°E longitude, another hybrid satellite capable of providing telecommunications, broadcasting, and meteorological services was launched to complete India's current meteorological satellite configuration. The INSAT 3A, as the latest of these satellites, includes very high resolution radiometers (VHRRs) operating in multibands and charge coupled device (CCD) multispectral cameras as well as a package to obtain Search and Rescue signals by downed pilots, ships in distress, or other emergency beacon



Fig. 39.3 Kalpana-1 meteorological satellite designed and launched by ISRO (Graphics courtesy of ISRO)

signals. This satellite's meteorological package included the following instruments:

- A Very High Resolution Radiometer (VHRR) with imaging capacity in the visible (0.55–0.75 μm), thermal infrared (10.5–12.5 μm), and water vapor infrared (5.7–7.1 μm) channels. This radiometer can provide 2 × 2 km and 8 × 8 km ground resolutions respectively.
- A CCD camera that provides 1 × 1 km ground resolution, in the visible (0.63–0.69 μm), near-infrared (0.77–0.86 μm), and shortwave infrared (1.55–1.70 μm) bands.
- A Data Relay Transponder (DRT) having global receive coverage with a 400 MHz uplink and 4,500 MHz downlink for relay of meteorological, hydrological, and oceanographic data from unattended land and ocean-based automatic data collection-and-transmission platforms.

The combined network of the INSAT 2E, the Kalpana 1, and the INSAT 3A, thus, provides comprehensive geosynchronous meteorological data for India and surrounding areas.

The entirety of the Indian meteorological satellite capabilities – unlike most other meteorological satellite networks of other countries – relies exclusively on geostationary satellite sensing and thus, do not include polar-orbiting meteorological satellites to obtain data from much lower orbits. For most national applications the 1 × 1 km resolution is considered adequate for interpreting major

weather formations. The ability to collect data from ocean buoys and land-band sensors also allows more precise interpretation of meteorological data via the data relay transponders on the Kalpana-1 and INSAT 3A satellites.

The meteorological data derived from this meteorological satellite network is processed and disseminated by the INSAT Meteorological Data Processing System (IMDPS) operated by the India Meteorological Department (IMD).[6] The above described satellites are able to provide up-to-date information on upper atmosphere winds, sea surface temperature, and precipitation index data. The products derived from the combined network include cloud motion vectors, sea surface temperature, outgoing long-wave radiation, and quantitative precipitation indices. These products are used for weather forecasting that employs both synoptic and numerical weather prediction.

INSAT-VHRR imageries are used extensively by Indian news agencies to provide localized weather forecasts. At present, the most detailed and synoptic weather system observations over the Indian Ocean from geostationary orbit are provided by the INSAT system. INSAT's Very High Resolution Radiometer (VHRR) data in visible and other spectral bands is currently available in near real time at 90 Meteorological Data Dissemination Centers (MDDC) in various parts of the country. With the commissioning of direct satellite service for processed VHRR data, MDDC type of data can be provided at any location in the country. A low cost and very low data rate (300 bits/s) reception unit has been developed for national users wishing to receive data directly from this and other Indian meteorological satellite packages. A cooperative agreement has been signed with EUMETSAT for using meteorological data from Meteosat-5 at 63° East in exchange for weather images collected by INSAT.[7]

IMD has installed 100 meteorological Data Collection Platforms (DCPs) and other agencies have installed about 200 DCPs all over the country and even on the Indian base station in Antarctica. DCP services are provided using the Data Relay Transponders of Kalpana-1 and INSAT-3A.

For quick dissemination of warnings against impending disaster from approaching cyclones, specially designed receivers have been installed at the vulnerable coastal areas in Andhra Pradesh, Tamil Nadu, Orissa, West Bengal, and Gujarat for direct transmission of warnings to the officials and public in general using the broadcast capability of INSAT. IMD's Area Cyclone Warning Centers generate special warning bulletins and transmit them every hour in local languages to the affected areas. Three hundred and fifty such receiver stations have been installed by IMD.

The Japanese Geostationary Meteorological Satellite System (Himawari) and the QZSS Network

The Japan Meteorological Agency (JMA) carries out a number of missions under the Japanese Meteorological Service Act as well as the broader Act for Establishment of the Ministry of Land, Infrastructure, Transport and Tourism (MLIT). In

addition to collection of meteorological and weather-related data, JMA is charged with an active role in several specific areas. These areas include the following:

- Prevention and mitigation of natural disasters
- Safety of transportation
- Development and prosperity of industry
- Improvement of public welfare

To meet these goals, JMA focuses its efforts on monitoring the Earth's environment and forecasting natural phenomena related to the atmosphere, the oceans, and indeed the entire Earth. It is also charged with conducting research and technical development in related fields and to this end the JMA has an active partnership with the Japanese Aerospace eXploration Agency (JAXA). JMA also engages in international cooperation activities regarding both meteorology and seismology to meet Japan's international obligations and to promote partnerships with various national meteorological and hydrological services as well as with various related international agencies – particularly the World Meteorological Organization (WMO) and the United Nations Environmental Programme (UNEP).

Particular emphasis is placed on the prevention and mitigation of natural disasters, as Japan is prone to a variety of natural hazards such as typhoons, heavy rains, tsunamis, and earthquakes. JMA, as the sole national authority responsible for issuing weather/tsunami warnings and advisories, is required to provide reliable and timely information to governmental agencies and residents for the purposes of natural disaster prevention and mitigation.

In this way, JMA plays a vital role in natural disaster mitigation and prevention activities in the country through cooperation and coordination with relevant authorities, including the central government and individual local governments. Thus, in addition to the collection of meteorological data via meteorological agencies, JMA also seeks to use satellite remote sensing data to investigate earthquake, volcano, and other disaster phenomena.[8]

JMA collects meteorological data from an extensive number of earth and ocean-based sensors as well as via upper atmosphere sensing devices. Meteorological satellites are a very important part of its overall observation and data collection process.

In 1977, Japan launched its first geostationary meteorological satellite (GMS) into geostationary orbit with an orbital location to cover the western Pacific and East Asia as part of a space-based component of the Global Observation System (GOS) under the auspices of the WMO World Weather Watch (WWW) programme. Since then Japan's meteorological satellite sensing capabilities have continued to expand and to date there have been five satellites in orbit. The Japanese meteorological satellites are known by a number of different names and thus, it is important to note that these satellites are variously known as the Japanese Geostationary Meteorological Satellites (GMS), the Himawari (meaning Sunflower in Japanese) system and Multifunctional Meteorological Satellites (MTSATs). Currently the MTSAT 1R (Himawari 6) and MTSAT 2 (Himawari 7) are the primary meteorological satellite system for Japan. During the period around 2003 and 2004 the service coverage for Japan was quite

disordered for over an 18 month period. This was due to the delays in the manufacture of the MTSAT 1R satellite that was being constructed by Space Systems Loral while this corporation was going through bankruptcy. There were additional delays due to problems with the performance of the Japanese IIA launch vehicle. During this period the United States loaned Japan the GOES 9 satellite to temporarily serve in the stead of MTSAT 1R until this satellite was successfully launched in 2005.

The current Japanese meteorological satellite network provides a wealth of information, including data on cloud height and distribution, upper-air wind, and sea surface temperature distribution. The observational data received from the spacecraft allows JMA and other national meteorological and hydrological services to continuously monitor significant meteorological phenomena such as typhoons, storm fronts, and low-pressure systems. The data collected by Japan's meteorological satellites are also directly assimilated into the numerical weather prediction system, which in turn contributes to the timely issuance of disaster prevention information and weather forecasts from JMA and related weather agencies.

The Multifunctional Meteorological Satellite (MTSAT-1R) was launched in 2005 after 2 years of delay. This allowed the GOES 9 to be returned to US operation. The MTSAT 1R was capable of performing observations every 30 min with imaging channels consisting of a visible band and four infrared bands. MTSAT-2, launched in 2006, took over many of the imaging functions of MTSAT-1R in 2010 and is now the prime Japanese meteorological satellite while MTSAT-1R remains on standby. MTSAT-2 has the same imaging channels as MTSAT-1R and thus, provides similar observation data as its predecessor. Additional data is obtained by JMA from various polar orbiting satellites, such as the NOAA and POES series, operated by the United States as well as the MetOps satellite operated by EUMETSAT of Europe.

Further, JMA obtains key information from two other "experimental" satellites. The Tropical Rainfall Measurement Mission (TRMM) satellite provides data on global rainfall and precipitation patterns to JMA. This satellite which provides very high resolution images because it is deployed in a circular orbit that is only 350 km in altitude utilizes microwave radar to obtain its measurements. The TRMM satellite is a joint mission between NASA and JAXA (Fig. 39.4).

Another NASA research satellite named AQUA also provides detailed oceanographic and tropical rainfall data as well as atmospheric images to JMA utilizing the onboard moderate resolution imaging spectroradiometer(MODIS) (Fig. 39.5).[9]

Data from these satellites are indispensable in observing typhoons, monitoring the global and marine environment, and producing initial fields for numerical weather prediction.

Currently the Himawira 8 and 9 Meteorological Satellites are in manufacture by the Mitsubishi Electric Company with the Boeing Corporation serving as subcontractor. These advanced satellites are planned for launch dates on the Japanese IIA launch vehicle in 2014 and 2016 respectively.[10]

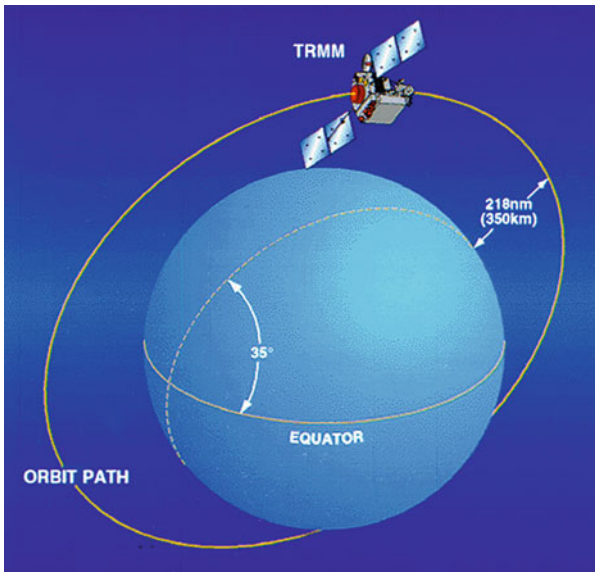


Fig. 39.4 The TRMM satellite in low earth orbit (Graphic courtesy of NASA)

The Russian Geostationary Operational Meteorological Satellite (GOMS) and Polar-Orbiting Meteorological (Meteor) Systems

The Russian meteorological satellite network was once one of the most robust in the world but budgetary constraints led to the decrease in the number of satellites in orbit. This has led to the Russian Federation relying on meteorological data from Europe, the United States, and other satellite networks. The Russian government has now strongly committed to restoring the network of weather satellites that existed during the time of Soviet Union. This is a large challenge in that the Russia needs to monitor weather and climate conditions across the country's 11 time zones which is by far the greatest challenge in terms of meteorological forecasting that any nation in world must face.

In the last few years, Russia has concentrated on developing and launching a full array of new meteorological satellites in both geostationary and polar orbit. This program is now underway but will not be fully completed for another decade.

Russia announced in September 2010 that it plans to fully restore its weather satellite network by 2030 under a State-sponsored program. As a first priority the Russian Federation State has undertaken to deploy a near-polar orbit constellation of Meteor-M satellites, beginning with the Meteor-M No 1 working on an 830-km circular sun-synchronous orbit, in order to bolster meteorological service for the vast expanse of the country and also to deploy and operate the latest Geostationary Operational Meteorological Satellite (GOMS) system. The geostationary satellites are to be comprised of the Elektro-L (launched in 2011) and Elektro-M (still to be launched).

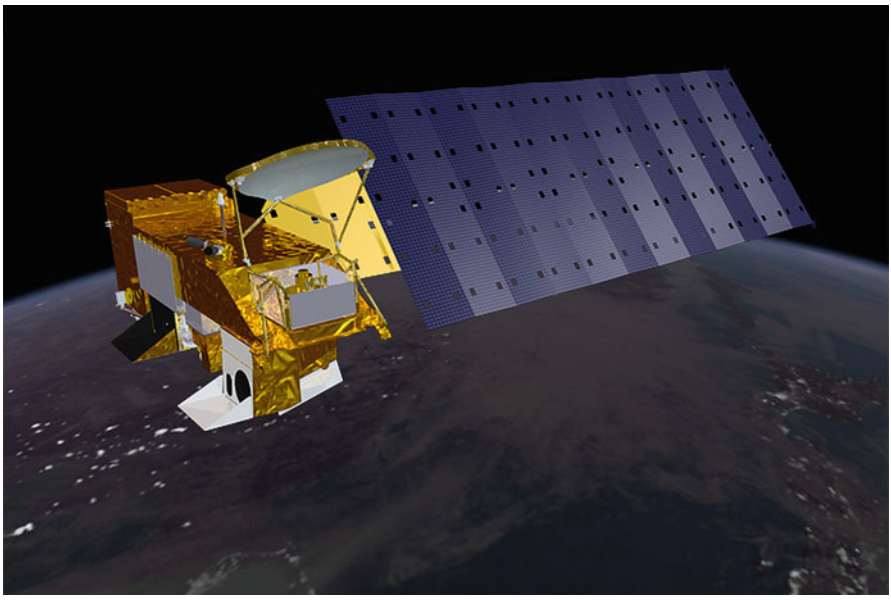


Fig. 39.5 The AQUA satellite over the Pacific Ocean at night (Graphic courtesy of NASA)

The Meteor-3 Polar-Orbiting Meteorological Satellite System

The polar-orbiting meteorological satellites, Meteor-3 presently operating in Russia, provide possibilities for acquiring data for hydrometeorological and helio/geophysical support as well as global environmental monitoring. The system's spacecraft are located on near-polar circular orbits (height of approximately 1,200 km, with an inclination of 82.5°). The characteristics of the instruments onboard the Meteor-3 series of satellites are shown in Table 39.2.

The scientific instrument package onboard the Meteor-3 spacecraft enables regular instant acquisition of images of cloudiness and of the Earth's surface in visible and infrared bands, data on air temperature and humidity, and sea surface temperature and cloud temperature. Acquired corpuscular and X-ray irradiance and total emitted radiation energy data are used for geophysical studies.

Beyond regular scientific hardware, the Meteor-3 spacecraft are often equipped with experimental and research instruments. The Meteor-3 satellite No.5, which was launched on August 15, 1991, carries the still operating scanning spectrometer for global ozone distribution mapping TOMS instrument, developed by NASA. The TOMS instrument has six spectral bands in the UV spectral band (312.5; 317.5; 331.3; 339.8; 360.0; 380.0 nm), a ground resolution 63 × 63 km^2 at nadir; 83 × 83 km^2 average; and a swath width of approximately 2,900 km. The Meteor-3 spacecraft No.8 carries the NASA-produced instrument SAGE-II to monitor distribution of aerosols, ozone, water

Table 39.2 Weather satellite Meteor-3 permanent onboard equipment

Instrument	Spectral band, um	Ground resolution, km	Swath width, km	Operating schedule
Scanning TV-sensor with onboard data recording system for global coverage mode	0.5–0.8	0.7 × 1.4	3,100	Recording, direct transmission
Scanning TV-sensor for automatic data transmission mode	0.5–0.8	1 × 2	2,600	Direct transmission
IR-radiometer for global coverage and direct data transmission modes	10.5–12.5	3 × 3	3,100	Recording, direct transmission
Scanning 10-channel IR-radiometer	9.65–18.7	35 × 35	400	Recording, direct transmission
Radiation Measuring System	0.17–600 MeV	–	–	Recording, direct transmission
Radiochannel	466.5 MHz – data transmission to control centers			
	137.850 MHz – data transmission to local acquisition stations			

vapor, and carbon dioxide. The SAGE-II instrument operates in seven spectral bands (centered at 385, 448, 452, 525, 600, 936, and 1,020 nm) and allows for restoration of vertical profile of atmospheric components with a vertical resolution of 1 km in altitude and a span of 10–40 km based on attenuation of direct solar radiation passing through the Earth's atmosphere. Russia's Hydrometeorological Center, with the help of the Russian Federal Space Agency, Roscosmos, is now deploying a total of six Meteor-M weather satellites operating in a low-earth orbit constellation. This constellation is to be complete by 2013 with four of the satellites now deployed. These satellites are being launched utilizing the new Soyuz-2 high-performance Soyuz booster. These satellites have service lifetime of 5–7 years. Subsequent generations of these polar orbit satellites will have a longer life of perhaps 12–15 years.[11]

The Meteor-M satellites are being manufactured by the Moscow-based VNIIEM, NPP (Science and Production Enterprise "All-Russian Scientific and Research Institute of Elektro-mechanics") under contract to the Russian Federal Space Agency, Roscosmos. VNIIEM/NPP was also responsible for manufacturing the earlier Meteor 1, Meteor-2, Meteor-3 series of satellites and the Geostationary Operational Meteorological Satellite (GOMS) weather satellite also known as Elektro-1.

The Russian Federal Space Agency successfully launched the latest Meteor-M weather satellite on the Soyuz-2.1b booster and anticipates using this launcher for the rest of the constellation. This 2,700 kg, 5–7-year life span spacecraft forms part of a constellation of weather observation spacecraft. Advanced Meteor satellites with longer lifetimes are planned to replace the Meteor-M satellites. In parallel with the new Meteor-M constellation are Russia's expanded capabilities in geostationary orbit.

The Russian Geostationary Operational Meteorological Satellite

The latest GOMS satellites, the Elektro-L and Elektro-M, have been designed by Roscosmos scientists and engineers in conjunction with climatologists and meteorologists to provide a wide variety of data, including weather analysis and forecasting on a global and regional scale. These satellites will be able to monitor changes in the climate as well as day-to-day weather patterns plus data from the Sun and information on cosmic radiation.[12]

These satellites are designed to provide synoptic images of the entire visible hemisphere of Earth at a resolution of 1 × 1 km per pixel (in the visible light band) and 4 × 4 km (in the Infrared band) and do so every 30 min. The weight of these spacecraft in operational mode is about 1,500 kg and their service lifetime is projected to be about 10 years.[13]

Elektro-L, also known as GOMS-2, was developed by ROSHYDROMET/PLANETA/Roscosmos. The Elektro-L is a successor spacecraft to the Geostationary Operational Meteorological Satellite (GOMS) that is also referred to as Elektro-1. Elektro-1 was launched on October 31, 1994 but was never brought to full operational service due to technical problems.

The overall mission objectives of Elektro-L and the Elektro-M (that are yet to be launched) are as follows:

- To provide an operational basis multispectral imagery (hydrometeorological data) of the atmosphere (including the cloud-covered sky)
- To provide complete updated images of the Earth's surface within the hemispheric coverage region (visible disk) of the spacecraft
- To provide information on high-energy cosmic radiation
- To collect heliospheric, ionospheric, and magnetospheric data
- To provide the required communication services for the transmission/exchange of all data with the ground segment
- To provide the services of data collection for the Data Collection Platforms (DCPs) in the ground segment as well as to provide the services of the COSPAS/SARSAT program which are to pick up the emergency search and rescue signals of pilots, distressed people at sea, or other isolated travelers in distress.[14]

The Elektro-L and Elektro-M spacecraft have been built by NPO Lavochkin Research and Production Association of Moscow in association with Roshydromet/Plantera/Roscosmos. The spacecraft employs the so-called navigator platform. This is a general purpose bus, which is three-axis stabilized, and provides a pointing accuracy of better than 0.05°. The angular drift is on the order of 5×10^{-4} o/**s**. A deployable solar array provides a power of 1.7 kW at end of life, while the spacecraft's mean power consumption is estimated to be about 700 W. The total mass of the spacecraft is about 1,620 kg with a payload mass of 435 kg. The Elektro-L and Elektro-M spacecraft design lifetime is projected to be 10 years. Figure 39.6 shows the Elektro-L spacecraft.

The Elektro-L spacecraft was successfully launched on January 20, 2011 on a Zenit-2 launch vehicle with a Fregat-SB booster from the Baikonur Cosmodrome,

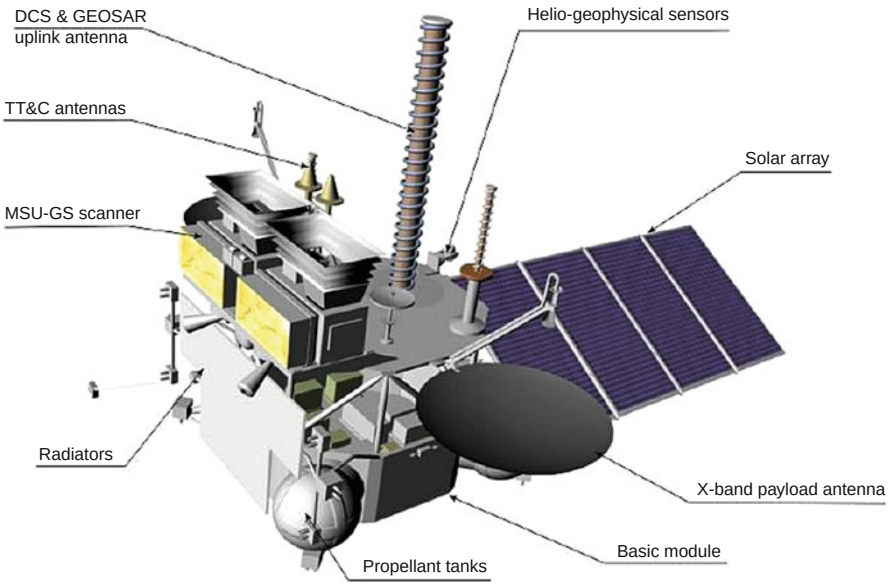


Fig. 39.6 The Elektro-L spacecraft and some of its components (Image credit: Roshydromet/Planeta)

Kazakhstan and began operations in geostationary orbit over the Indian Ocean in March 2011.

The Elektro-L spacecraft has several key component subsystems. These are the Onboard Radio Engineering Complex (OREC), the Multispectral Scanning Unit – Geostationary Scanner(MSU-GS), the Helio/geophysical Instrument Complex (GGAK-E), the Onboard Data Sampling System (ODSS), and the Geostationary Search and Rescue System (GS&RS). The first three of these components, that are most critical to the meteorological and climatological mission, are briefly described below (https://directory.eoportal.org/get_announce.php?an_id=1544).

OREC (Onboard Radio Engineering Complex). The objectives of the RF communication system are to provide all data transmission, relay, and retransmission services with the ground segment. These RF relay functions include the following:

- The sensor data downlink to the Ground Acquisition and Distribution Center is in X-band (7.5 GHz) at a data rate of 2.56–15.36 Mbit/s.
- Data reception from ground segment Data Collection Platforms (DCPs) at 400 MHz (UHF), or DCP data relayed via LEO satellites at a frequency of 470 MHz. This data is transmitted from Elektro-L to the Ground Acquisition and Distribution Center in S-band at 1.7 GHz.
- Onboard reception of processed hydrometeorological data products in X-band (8.2 GHz) and relay of this data (in S-band at 1.7 GHz) to all customers.
- Exchange of hydrometeorological data and remote sensing data between regional centers in X-band (at 8.2 and 7.5 GHz) with data rates of up to 15.36 Mbit/s.

- Data reception of COSPAS-SARSAT messages at 406 MHz and retransmission of these messages at 1.54 GHz.

The Multispectral Scanner Unit (MSU-GS) is a 10-channel radiometer. The objectives of this radiometer unit are to obtain solar-reflected imagery and brightness temperature measurements from the top of Earth's atmosphere and from the Earth's surface (ocean and land). In addition, the tropospheric moisture content is also determined.

The MSU-GS instrument is a multispectral scanner to take imagery in three visible and seven infrared bands. These measurement bands are closely parallel to the instruments now operating onboard the MeteoSat-8/MSG-1 spacecraft of EUMETSAT.

Helio/geophysical Spectrometry Instrument Complex GGAK-E

The system is designed to monitor the penetrating radiation's spectra and density in the near-Earth environment and the magnetic field state. The system records the following helio/geophysical information (HGI):

- The density of electron fluxes with energies in four bands from 0.04 to 1.7 MeV
- The density of proton fluxes with energies in four bands from 0.5 to 90.0 MeV
- The density of alpha particles fluxes with energies from 2 to 12.0 MeV
- Intensity of galactic cosmic radiation with energies greater than 600 MeV
- Solar X-ray radiation intensity with energies from 2 to 10 Å°
- Intensity of solar ultraviolet radiation in four wave bands up to 1,300 Å°

These very sophisticated subsystems allow the Russian meteorologist to have access in near real time to a wealth of data. The Russian meteorological network will continue to improve as the Elektro-M GOMS satellite and the full Meteor-M and Advanced Meteor-M satellite constellation are deployed (https://directory.eoportal.org/get_announce.php?an_id=1544).

South Korea's Communication, Ocean, and Meteorology Satellite (COMS)

South Korea's first Communication, Ocean, and Meteorological Satellite (COMS-1), dubbed Cheollian, was launched successfully by Arianespace using an Ariane 5 ECA rocket on June 27, 2010 from the Guiana Space Center in Kourou, French Guiana. The COMS-1, or Cheollian, is a multipurpose geostationary satellite capable of performing communication, ocean, and meteorological functions. The COMS satellite has been placed in geostationary orbit at 128° East. Its mission is scheduled to last 7 years; however, the satellite has a design life of 10 years.[15] The COMS satellite will be operated by the Korea Aerospace Research Institute (KARI).

The COMS satellite has three payloads: one for meteorology, one for ocean observation, and one for communications. COMS will provide meteorology data to end users around the globe and oceanography data for the Korean peninsula. This multipurpose satellite will also carry out experimental satellite communications services in Ka-band. As prime contractor for COMS, EADS Astrium was

responsible for the design and building of the satellite including both the meteorology and ocean imagers. The communications payload was provided by KARI in Korea, as a Customer Furnished Equipment.[16]

- COMS provides continuous image monitoring with the extraction of high-resolution meteorological data from its multispectral imager. It will give early warning of hazardous weather conditions including storms, floods, sandstorms, etc., and provide data on the long-term changes in sea surface temperatures and cloud patterns. Earth observation data from COMS will be relayed to a processing station. Once processed, the data will be resent via the COMS satellite to weather forecasters, Earth observation centers around the world.
- COMS will also carry an Ocean Imager to monitor marine environments around the Korean peninsula and provide data (on chlorophyll, etc.) to assist the fishing industry in the region. It will also monitor both long- and short-term changes to the marine ecosystem.
- The communications payload onboard COMS will allow "in-orbit verification" of advanced communication technologies and will support experiments covering wide-band multimedia communication services.

COMS carries the following payloads:

- Meteo Imager: The Imager is a multispectral channel two-axis scanning radiometer, and is capable of providing imagery and radiometric information of the Earth's surface and cloud cover over five channels – one visible channel (of 1 km ground resolution) and four Infrared channels (of 4 km resolution).
- Ocean Color Imager: The advanced Ocean Imager has a sophisticated focal plane providing for ocean data aquisition from geostationary orbit. The Ocean Imager will provide data over eight imaging bands in the visible spectrum. Ground resolution over Korea is 500 m.
- Meteorology Data Dissemination Function: Using an S-band-receiving antenna and an L-band-transmitting antenna, this function will allow dissemination in HRIT and LRIT format of weather data.
- Communications Payload: The Ka-Band payload will provide three regional beams simultaneously. The Ka-Band Payload will provide the beam switching function for high-speed multimedia services including the Internet via satellite in the public communications network for all coverage.

COMS is KARI's first geostationary satellite and will provide Korea with its own meteorology and ocean data, thus giving increased independence. COMS is part of a 15-year Korean space plan begun in the 1990s, and followed systematically ever since.

After a period of early operation, satellite communication and meteorological/ocean data services will be offered for public use. According to the national long-term plan for space development, a second geostationary multifunction satellite will be launched sometime after 2014.

Conclusion

The in-orbit global meteorological satellite resources represented by the United States, Europe, China, India, Japan, Russia, and South Korea are today quite

considerable and in the coming years, these satellite systems will continue to grow in scope and capability. The addition of new sensor capabilities to monitor heliographic and cosmic radiation and the deployment of more satellites in various relevant orbits will help to chart various elements of climate change, the melting of the arctic ice caps, the dimensions of the holes in the ozone layer, and increasing temperatures on land and in the ocean around the world. These expanded space facilities will play a critical role in providing better and longer-term weather forecasts, but also in developing new strategies to cope with the many adverse affects of climate change itself.

The collaborative efforts that come from the United Nations World Meteorological Organization (WMO), the World Weather Watch (WWW), the United National Environmental Programme (UNEP), and similar international and regional organizations can be of enormous value. Already most meteorological satellite data from around the world is freely shared in the common cause of dealing with the effects of tropical storms, monsoons, typhoons, hurricanes, and tornadoes. As new tools such as hyper-spectral sensing, lightning strike intensity, and cosmic radiation monitoring, the opportunities for even more international collaboration will continue to evolve. This progress will come in many ways such as improved instrumentation, more sophisticated data analysis and formatting, and better ways to monitor not only weather patterns, but longer-term trends in climate change.

Today all of the countries involved in satellite meteorology as discussed in this and previous chapters can design, build, and launch their own satellites – a significant change from early in the space age.

In the future, additional countries will deploy sophisticated meteorological satellites which will thereby enrich tomorrow's space capabilities. In 20 years, longer-term weather forecasts and much more sophisticated sensing of climate change will surely emerge from all of today's efforts to strengthen international cooperation in this vital area.

Cross-References

- ▶ Ground Systems for Satellite Application Systems for Navigation, Remote Sensing and Meteorology
- ▶ Introduction to Space Systems for Meteorology
- ▶ United States Meteorological Satellite Program

Notes

1. http://www.wmo.int/pages/prog/www/index_en.html
2. http://www.wmo.int/pages/prog/www/OSY/GOS.html
3. Successful launch of the Fengyun II-7 satellite http://english.cntv.cn/20120113/111472.shtml and http://www.nsmc.cma.gov.cn/NewSite/NSMC_EN/Home/Index.html

4. FY-3 (Fengyun 3) 2nd Generation Polar-Orbiting Meteorological Satellite Series https://directory.eoportal.org/get_announce.php?an_id=12759
5. http://satellite.cma.gov.cn/ArssEn/Ord/Satellite.aspx
6. http://www.imdpune.gov.in/research/ndc/ndc_index.html
7. Listing of ISRO Satellites http://www.isro.org/satellites/allsatellites.aspx
8. The Mission of the Japanese Meteorological Agency http://www.jma.go.jp/jma/en/Background/mission.html
9. Aqua Satellite by Nasa http://www.google.com/imgres?imgurl=http://earthobservatory.nasa.gov
10. Himawari 8 and 9 http://space.skyrocket.de/doc_sdat/himawari-8.htm
11. Russia to have five weather satellites by 2013, Moscow, RIA Novosti, October 20, 2008, pp. 1–2, http://wwwspacedaily.com and http://www.globalsecurity.org/space/world/russia/meteor-m1.htm
12. http://sputnik.infospace.ru/goms/engl/goms_1.htm
13. http://en.rian.ru/science/20100915/160591614.htmlhttp://en.rian.ru/science/20100915/160591614.html and also see http://www.globalsecurity.org/space/world/russia/meteor-m1.htm
14. https://directory.eoportal.org/get_announce.php?an_id=1544
15. http://en.wikipedia.org/wiki/Chollian#cite_note-GSP-2
16. http://space.skyrocket.de/doc_sdat/coms.htm

Section 5

Spacecraft Bus and Ground Systems

Overview of the Spacecraft Bus

40

Tarik Kaya and Joseph N. Pelton

Contents

Introduction 1046
Spacecraft Structures 1046
Orbital Control and Pointing Accuracy 1048
Power Systems 1051
Batteries 1054
Nuclear and Isotope Power Systems 1056
Thermal Control and Heat Dissipation 1059
Onboard Heaters and Cooling Systems 1062
Cross-References 1064
Notes 1064

Abstract

The evolution of application satellites has hinged on the development of more and more sophisticated spacecraft buses or platforms. The development of three-axis body stabilized platforms have allowed the deployment of more capable and much higher gain communications antennas, high resolution remote sensing and meteorological sensors, and more precise navigational payloads. The most important development in spacecraft buses has been the development of precisely oriented body stabilized platforms that allow the deployment of very high powered solar arrays and very accurate pointing of high-gain antennas and sensor systems. Other challenges have included developing lower mass and structurally strong spacecraft

T. Kaya (✉)
Mechanical and Aerospace Engineering Department, Carleton University, Ottawa, ON, K1S 5B6, Canada
e-mail: tkaya@mae.carleton.ca

J.N. Pelton
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: peltonjoe@gmail.com

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_87, © Springer Science+Business Media New York 2013

bodies, improved and longer life thrusters, better performance power systems with greater density of charge, and improved thermal control systems. This chapter explores the development of the spacecraft bus and their technologies. The following chapters discuss tracking, telemetry, and command; reliability testing; and the adaptability of essential multipurpose platforms to different applications.

Keywords

Battery systems • Carbon/epoxy composites • Despun platforms • Fuel cells • Fuel slosh • Heat dissipation • Heat pipes • Inertial wheels • Isotope power systems • Momentum wheels • Nuclear propulsion • Orbital control • Power systems • Quantum dot technology • Redundancy • Reliability and lifetime testing • Remote sensing sensors • Solar arrays • Solar cells • Spacecraft platforms • Spacecraft structures • Thermal control • Three-axis body stabilization • Thrusters

Introduction

Spacecraft Structures

The main goal of building spacecraft structures has been low mass and high strength. The evolution of carbon/epoxy composites and other similar hybrid materials have allowed engineers to construct satellite bodies that are up to three times lighter than the earliest models built with steel and titanium yet the spacecraft structure can still be as strong as – or stronger than – the earliest satellite designs. The key to developing the strong, but lightweight composite materials is to utilize them not only to build the structure that holds the solar arrays, the electronics, the antennas, the batteries, and momentum wheels, but to also use these materials elsewhere if possible. Thus these materials may also be used in antenna masts, radiator panels, or in other parts of the satellite. Any mass saved is a true bonus in the field of space applications.

The initial satellite bodies were largely shaped as drums. These satellites maintained their stability for pointing to the Earth by spinning around then using this angular momentum to achieve their orientation – much like a spinning top. In the case of communications satellites despun antenna systems revolved inside of the structural drum. This drum with the batteries, solar cells, and stabilization and positioning thrusters served as the satellite bus. These "internal antenna systems" spun in the opposite direction to the satellite bus. By spinning at exactly the same rotational speed but in the reverse direction, the effect was a completely synchronized system that could be constantly pointed to the Earth from GEO. Power was supplied to the antenna or sensing system from the bus via what was called a Bearing and Power Transfer Assembly (BAPTA).

This despun spacecraft design worked well and reliably in the early years. But this early design was shown in time to be inferior to the three-axis body stabilization design for a variety of reasons. First, in the case of three-axis body stabilized

design, the solar array could be deployed to achieve 100% sun illumination rather than 40% illumination achieved with the rotational design. With a spinning satellite the solar cells mounted on the exterior drum were illuminated only a part of the time. The rest of the time they were behind the satellite until they reemerged after the outside drum rotated back into a view of the sun. Moreover, the angle of the drum could also shadow the solar cells and prevent optimum illumination.

Secondly, the spinning spacecraft structures were also found to be inferior in design to the box-like structures which could be stabilized on all three axes by means of high-speed reaction or momentum wheels. On the spinning spacecraft the fuel tanks that supply the onboard thrusters spin with the spacecraft at speeds of up to 60 rpm. The fuel slosh from this action can be sufficient in some instances to cause sufficient precession to occur to send the satellite into a flat spin in the "X" axis rather than the intended "Y" axis. The design of reaction or momentum wheels, with their very high spin rate (i.e., 4,000–5,000 rpm) help to keep the entire spacecraft box and fuel tanks quite stable. Thus fuel slosh is not a problem.

The third reason in favor of three-axis body stabilization – that of greater pointing accuracy – was, in fact, the main driver that led to the design of the three-axis body stabilized platforms in the first place. The details why ever more precise orientation and positioning capabilities were required for later generations of application satellites will be discussed in the next section.

As a result, the box-like structures have been extensively used in various space missions. The rectangular monocoque body (i.e., one in which the outer structure carries most of the load and stress requirements) results in better strength and stiffness characteristics. The structural design of the satellites is mostly standardized. In the case of some special science missions, very different external configurations are of course possible. However, in general, the satellites are manufactured around a primary structure. The primary structure carries the main external loads and is usually made from a central tube in combination with flat panels. This central tube is usually the major load path between the launcher and spacecraft components. The secondary structure includes mounting platforms, solar panels and other deployable parts. The secondary structure also serves as a closure panel for the satellite. Finally, for housing electronics and supporting electric cables, etc., some smaller structures are also used (Fig. 40.1).

Research continues to develop even lighter and stronger materials for spacecraft structures. Also certain problems with epoxy or resin composites in orbit are being researched. Atomic oxygen that reacts with these composites causes microscopic deterioration in certain resins over time. The solution to this problem has been to apply a protective coat to shield the composite materials in the spacecraft structure from the atomic oxygen. Just a thin veneer prevents the atomic oxygen from chemically interacting with the composite materials of the spacecraft structure. Also research continues to develop even more reliable momentum wheels with "magnetic bearings" and thus no rotational friction. These high-reliability momentum wheels are designed to maintain quite accurate pointing capability and a high level of stability for the entire spacecraft – i.e., the bus itself, the deployed solar arrays and especially for the payload antennas, sensors, and meteorological and navigational payload systems.

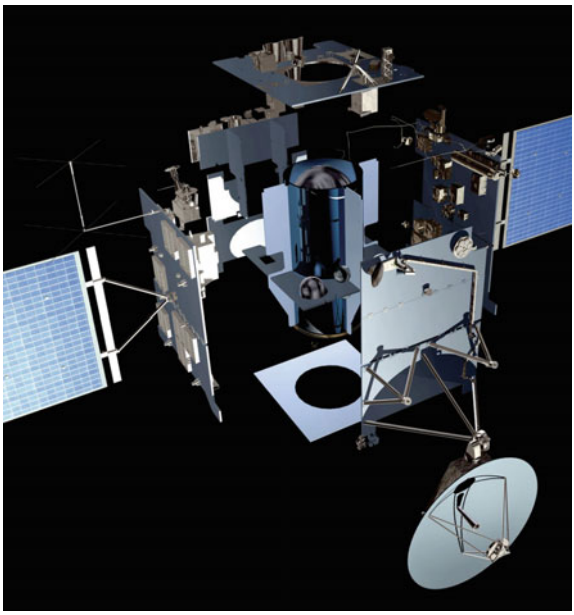


Fig. 40.1 An exploded view of Rosetta spacecraft with the primary and secondary structures (Graphic Courtesy of ESA)

Deployable structures are also a current area of research for several space applications since they can potentially decrease the spacecraft mass and volume, which is very important to lowering launch costs. It is also possible to launch larger antennas by using deployable structures for the GEO communications spacecraft, which can help further reducing size and power requirements of mobile ground receivers.

Orbital Control and Pointing Accuracy

The original development of three-axis body stabilized design was an attempt to build a more efficient spacecraft, which is capable of very long distance data relays, associated with planetary missions. The greater distance involved in planetary research vehicles requires much greater pointing accuracy than could be achieved by a spin stabilized spacecraft. The creation of a body stabilized spacecraft created the bonus of allowing the "wings" of a solar power array to be oriented to achieve maximum solar exposure. The use of gravity-gradient or spin stabilized spacecraft are not generally used for most sophisticated and high-value application satellites, simply because they do not offer the pointing accuracy required by most of today's satellites.

The pointing accuracy of a satellite depends not only on the speed and performance of a momentum or reaction wheel but also the ability of the spacecraft to be oriented exactly in the desired direction over time. The very high torque of the reaction or momentum wheel allows very small adjustments. By adding or subtracting a small amount of energy to a reaction wheel on a single axis stability can be achieved. Once this is done in all three axes the satellite can be perfectly



Fig. 40.2 A momentum wheel that is used to achieve spacecraft stability (Photo courtesy of NASA) (www.sti.nasa.gov/tto/spinoff1997/t3.html)

stabilized and pointed in the desired direction. A momentum wheel rotates at a higher speed than a reaction wheel (up to 10,000 rpm) to provide the required stability. Momentum wheels are quite small in diameter (i.e., in the range of 15–20 cm in diameter) but their very high rotational speed of 4,000–5,000 rpm give them a great deal of inertial force or torque. The screwdriver in Fig. 40.2 helps to picture the scale of these momentum wheels.

The initial application satellites were oriented in the desired direction by the use of Earth, sun, and star sensors. These sensors worked reasonably well. In the case, however, of a satellite losing orientation these sensors could lead to difficulties in reestablishing the proper orientation with sufficient speed. The main problem, however, was that such sensors only allowed a pointing accuracy of about 0.5–1.0°. This was all of the pointing accuracy that despun platforms could achieve, but three-axis body stabilized platforms could be pointed with greater precision. It therefore followed that a more precise system to orient satellites would be needed. This led to the use of RF beacons to align satellites with precise locations on the ground. These RF beacons can allow a body stabilized spacecraft to be oriented from GEO to a quarter degree of pointing accuracy. Currently there are alternative methods being researched that would use laser systems that could achieve even more precise pointing accuracy.

The other aspect of orbital positioning and orientation depends on onboard thrusters to keep the satellite in the desired orbital position and properly pointed. This is the greatest challenge for satellites in GEO since this type of spacecraft is positioned in equatorial orbit at truly great distances – approximately one tenth of the distance from the Earth to the moon. For many years orienting thrusters on

spacecraft have used hydrazine fuel or a combination of hydrazine and other hypergolic fuels such as nitrogen tetroxide in bipropellant systems. Hypergolic fuels are almost always quite toxic, but have the special feature of exploding with great force on contact with an oxidizer. Vernier thrusters with tiny jets could be used with precision to achieve great accuracy with such propellant systems. There has been research at NASA Glenn Research Center to develop nontoxic and lower cost fuels that still achieve spontaneous combustion. Based on this research NASA has now patented an innovation that overcomes the problem of using fuels such as methanol as a propellant for satellite thrusters. Their research has developed a satellite thruster that offers the ability to catalytically decompose a reduced-toxicity propellant into hot gases but still having the ability to spontaneously react with an oxidizer to begin the combustion process. This new system can be used for both bipropellant and monopropellant satellite propulsion and in the process this approach can also help to reduce the cost and complexity of satellite missions.[1]

Satellites in GEO are constantly being tugged away from their specifically assigned equatorial position. The gravitational forces of the sun and the moon as well as the Earth's irregular shape and composition serve to move GEO satellites East or West along the equator away from their assigned position. Much more to the point these various gravitational forces are much stronger in the North and South vertical directions in or out from GEO. These forces are at least ten times more powerful in the North and South direction than in the East and West directions.

The answer to this problem is to develop propulsive systems with much longer lives. The development of electric propulsion and ion propulsion systems have been the answer to longer lived systems. Electric propulsion provides thrust capabilities for much longer, but with much lower levels of force. In an ion engine ions from a heated gas such as Xenon are accelerated by the electrostatic force to tremendous speeds to provide the thrust (Fig. 40.3).

In the longer term ion engines provide more than two to three times the energy per unit of mass than chemical propulsion, but the thrust force at any particular time is much weaker. This often ends with the need to combine chemical propulsion systems to achieve initial orbital location. It is left to ion engines to keep the satellite oriented and positioned properly over the longer term. Application satellites have grown in sophistication, lifetime, power, and performance and this has also meant that they have grown more massive and large. Although electronics, processors, and application-specific integrated circuits (ASICs) have shrunk in size, antenna and power systems have increased greatly in mass and volume. This means that positioning, stabilization, and orientation systems have also grown in size and performance capabilities have had to improve. Ion engines have certainly increased longer term capabilities, but there has been increased research into the possibility of developing nuclear-powered capabilities. Such nuclear-powered systems could provide for not only orbital positioning but also provide for the power supply as well. There remain serious concerns about nuclear power safety, but nuclear-powered propulsion as well as power systems to support remote sensing (especially radar systems) and very large communications satellites remain a future possibility.

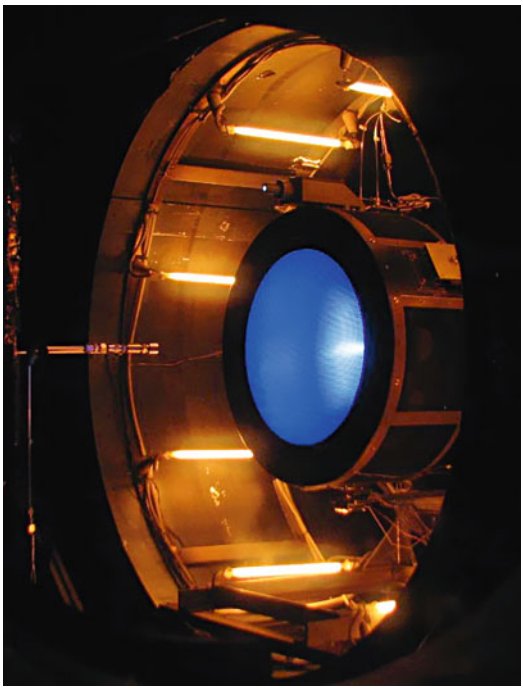


Fig. 40.3 A long-life Xenon ion thruster (Photo courtesy of NASA) ("Structure and Properties of Matter: Ion Thruster" Ion Thruster from NASA http://dawn.jpl.nasa.gov/DawnClassrooms/2_ion_prop/index.asp)

Power Systems

All application satellites require a considerable amount of power to operate and thus providing a reliable, long-lived source of power is critical to mission success. Some types of satellites require greater amounts of power than others. In the case of remote sensing satellites radar satellites that represent "active sensing" that initiates a signal from the satellite that "bounces back" requires much more power than other remote sensing satellites. In the case of satellite communications direct broadcast satellites and mobile satellite systems require much more power than fixed satellite systems. All application satellites benefit from a longer lifetime. Thus power systems that can be recharged and last 15 years – if not longer – are certainly welcome.

The most common type of failure in the field of space applications are those related to loss of power – either in space or in ground systems. It is thus important to design systems that have high performance but also have the greatest possible reliability. Most application satellites are powered by solar arrays with backup battery systems that supply power during periods of eclipse or short-term outages that might occur. There has been a constant effort to upgrade the performance of solar arrays that has seen the use of ever more efficient solar cells. The first solar cells were silicon based. These were upgraded from amorphous structure silicon

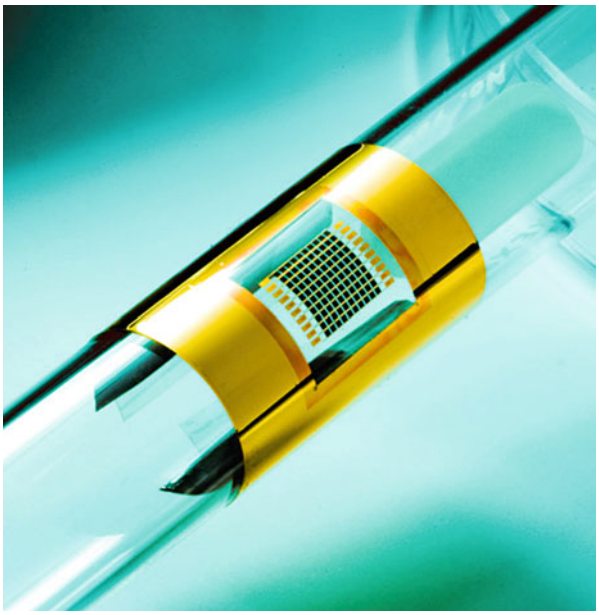


Fig. 40.4 Rapid production of flexible gallium arsenide solar cells will reduce cost (Photo courtesy of Prof. John Rogers, University of Illinois at Urbana) (A. Beleicher, Rubber-stamping makes creating solar cells, transistors, and infrared detectors easy. IEEE Spectrum (2010) http://spectrum.ieee.org/semiconductors/materials/thinfilm-trick-makes-gallium-arsenide-devices-cheap. Accessed May 2010)

solar cells to structured crystalline structure solar cells that produced higher efficiency conversion of photon received from the sun. In time even higher performance gallium arsenide solar cells were used. Although these solar cells are more expensive to manufacture the higher performance justifies the investment when the high cost of launch to orbit is considered. Some designs have included solar concentrators that allow the solar array to see the equivalent of two or even three suns. Today there is continuing research to improve solar array performance even further. One of the prime objectives is the reduction of the production cost for high-performance gallium arsenide solar cells. Recent breakthroughs in the production of flexible gallium arsenide solar cells promise the relatively near-term reduction in costs for all times of applications (Fig. 40.4).

Work at research labs has identified methods that allow the mass production of gallium arsenide cells that can be stripped off and quickly applied to substrates. The production of gallium arsenide solar cells with higher efficiencies and much lower costs are likely within the next 2–3 years.

There are ongoing efforts to increase these efficiencies to even higher levels. Quantum dots that can be "grown" on solar cells to increase their efficiency up to an estimated 50% are also under development. These quantum dot developments that allow more photons to be absorbed would translate into higher energy conversion efficiencies for high-performance solar cells. This type of design coupled with an addition of more junctions – especially junctions that capture energy in the ultraviolet range – are also in development (Fig. 40.5).[2]

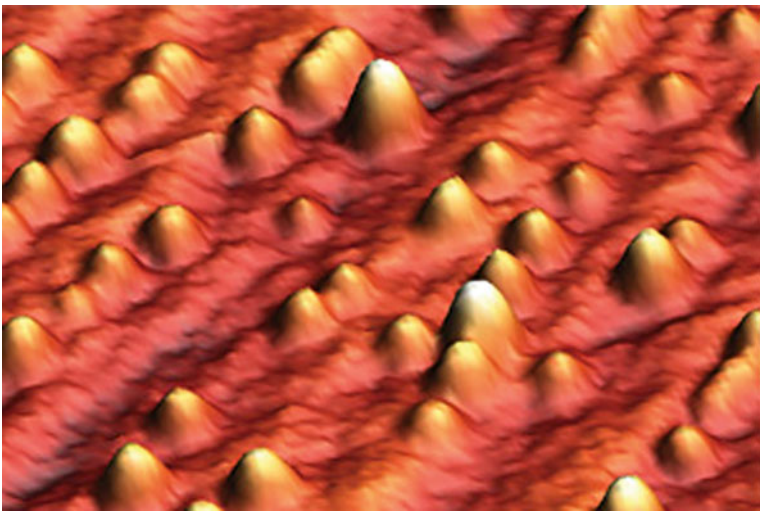


Fig. 40.5 Quantum dots can be "grown" on solar cells to increase their efficiency (Graphic courtesy of Rochester Institute of Technology Nano Power Research Labs)

By choosing the most suitable materials for cell junctions, it has been shown that more efficient multi-junction solar cells can be manufactured. The materials are chosen to absorb as much solar energy as possible, thus taking advantage of a larger part of the incoming solar power. In these cells, the junctions are connected in series instead of having a single junction of the conventional solar cell. Thus, the multi-junction cells generate lower current than the single junction cells but much higher voltage and power. As a result, the triple-junction cells have high conversion efficiencies on the order of 30%. The triple-junction cells have been used in several space missions. A recent example is NASA's Mars Exploration Rovers: Spirit and Opportunity. Research is under way to manufacture cells with four or higher junctions to increase further cell efficiencies. Multi-junction cell efficiencies up to 42% in laboratory conditions have been reported. The combination of the best materials, quantum dot technology, and multi-junction solar cells may allow the development of solar cells with over 50% conversion efficiencies (Fig. 40.6).

The other objective with regard to solar cell power system is to develop more advanced solar array deployment systems. The latest designs can roll out solar cells on a thin plastic substrate reel that is flexible rather than a rigid one. Another alternative new design would be to have inflatable solar arrays that could be quite low in mass and create effective deployment systems for flexible arrays. Flexible solar arrays, which are deployed with low-mass thin-film solar cells, allow the overall mass of the solar cell system to be less and more reliable. Current state of the art in these flexible solar arrays is on the order of 150–200 W/kg.[3]

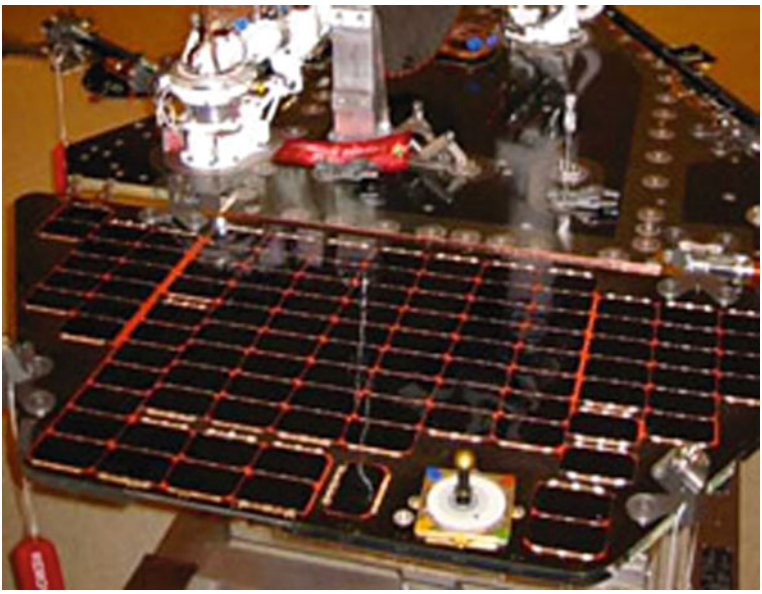


Fig. 40.6 The rear panel of the solar array on the MER Rover, showing the high-efficiency triple-junction solar cells (Photo courtesy of NASA)

Batteries

The onboard batteries are a critical component of the spacecraft's power system. Although a satellite can operate exclusively off of a solar array when it is being fully illuminated, the problem is that satellites of various types and in various types of orbits go into eclipse and then batteries become essential. LEO and MEO satellites operating in constellations – especially LEO satellites – are constantly going into and out of eclipses as they move behind the Earth. Thus satellites in LEO and MEO are constantly dependent on their batteries on an orbit to orbit basis. The Iridium constellation system in LEO was designed so that when these satellites are in the polar region they can be reconditioned and recharged. For satellites in the LEO and MEO orbits conditioning of batteries and consideration of when and how to discharge are a much more crucial matter and certainly plays into spacecraft lifetime and reliability considerations.

For GEO satellites the considerations are much different. For long periods of time during the year the satellite is in continuous solar illumination. There is an annual cycle, however, which twice a year brings a GEO satellite into a daily eclipse. During the maximum eclipse period the satellite is without solar-based power for over an hour. Battery-based operation must sustain the satellite operations during these periods. At the beginning of the cycle there is only a very brief period of eclipse that builds up to over an hour as of the Winter Solstice, for instance, and then the eclipse gradually subsides.

At the outset of space applications spacecraft were equipped with nickel cadmium batteries that provided reasonably good power density and lifetime. Over time higher energy density nickel hydrogen batteries were developed and implemented on more and more spacecraft. A shift to different battery types is not a decision taken lightly since the battery system represents a single point of failure. Extensive lifetime testing is undertaken, but since spacecraft lifetimes of up to 15 years are desired, this becomes a major barrier to innovation. In the case of batteries, testing under elevated temperatures can be utilized as a means of compressing the reliability and lifetime testing process. Most recently the transition to lithium ion batteries for spacecraft power storage has taken place. This type of batteries that are in wide usage for cell phones, laptop computers, etc., contain the highest energy storage density and have proved to be reliable over long lifetimes and quite adaptable to recharging.

Batteries are key elements of a satellite design for spacecraft engineers. They represent a critical resource to the overall functioning of the payload and thus a potential single point of failure. They, however, also constitute an element of mass and volume that limits the size of the payload or increase the cost of the launch services. This has led to an ongoing effort to develop batteries that contain a higher energy density, are lighter in weight, and still quite reliable over time with a mean time to failure in excess of 15 years.

There have even been thoughts that if one could develop sufficiently lightweight and high-performance batteries (or perhaps fuel cells as will be addressed immediately below), one could eliminate solar cell arrays and develop satellites that were entirely powered by batteries and/or fuel cells or combine such systems with thermo-ionic converters. As the power requirements for applications have ascended to perhaps tens of kilowatts, the feasibility of such a design approach has begun to seem more and more unlikely.

The research for the future is aimed at developing new technology for energy storage that could be more efficient than current battery technology as we know it today. Prime in this regard is the concept of developing unitary regenerative fuel cells. Fuel cells typically take hydrogen and oxygen and combine them via an electrochemical system that is triggered by a catalyst to make this process happen. In a way this electrochemical process can be likened to battery, but in this case the product is both electricity and water. The water, within what is called a "regenerative fuel cell" can then be electrolyzed to produce hydrogen and oxygen once again and start the process over again. Recently, researchers have, however, experimented with other possible regenerative chemical interactions. One of these fuel cell design processes has found that hydrogen peroxide (H_2O_2) and sodium borohydride ($NaBH_4$) can produce long-lived and effective reactions. Although this system is currently operating at an efficiency of 1,000 W-h/kg, it is projected that this system (for possible spacecraft use) might be able to operate at efficiencies that are almost three times greater. Other approaches to hydrogen peroxide fuel cells have found that methanol can operate efficiently with a lower cost alkaline-based catalyst[4].

Various types of regenerative fuel cell systems have been under development for many years and indeed successful fuel cells have been developed for use in ground vehicles, aircraft, and even spacecraft. The prime power storage systems on the Space Shuttle were fuel cells. The most recent research suggests that breakthroughs in the next few years can see fuel cell technologies used not only in space and in vehicles but perhaps to serve as energy systems for buildings and other industrial uses.

The problem to date has been that the catalysts have typically involved quite expensive materials. A quantitative objective of developing a fuel cell that could generate up to a kilowatt of power for under $1000 has long eluded developers working in this field. Recent efforts to develop lower cost catalysts and to perfect systems with longer life and lower mass have increased hopes that viable regenerative fuel cell technology for application satellites can be implemented in near future. Perhaps the hydrogen peroxide fuel cell system might prove both feasible and cost effective for such purposes in the next few years.

Another area of research with regard to cost-effective energy storage is the development of very high velocity flywheels that can store a significant amount of energy. Flywheel systems are increasingly being used on the ground as backup energy storage systems for emergency power restoration in emergency communications systems and even conventional power storage systems. Such flywheel systems do not require an electrochemical process to take place. Nor do these systems require almost continuous operational management with periodic discharge and other maintenance efforts, which is the case with batteries. The most attractive aspect of this approach with regard to application satellites is the possibility that the flywheels could serve a dual purpose of not only storing energy, but also utilizing these flywheels as reaction wheels for stabilization and orientation purposes as well. NASA and other space agencies are thus pursuing research in this area with the objective of in-orbit tests of reaction or momentum wheels that could also serve as energy storage systems.[5]

Nuclear and Isotope Power Systems

The story of application satellites has been in many ways a history of "technology inversion." This means that application satellites have evolved from small, low-powered devices that work with large, sophisticated, and high-powered ground systems to the "inverse situation." This means that today's application satellites, for the most part, are large, sophisticated, and high-powered systems in space that work to increasingly small, mobile, and low-powered systems on the ground. In some cases the ground systems can even be handheld mobile systems. This trend has pushed ever upward the demand for onboard power systems – especially for active radar systems for remote sensing and mobile and broadcast communications satellites. The first communications satellites such as Early Bird (1965) generated 100 W of power. Today's most capable application satellites

may typically require on the order of 12–15 kW. The various technologies discussed above have been able to deploy power systems capable of generating such levels of powers, but the future suggests that even higher power levels may well be required.

One possible pathway forward would entail the use of nuclear reactors or radioisotope systems to power the application satellites of the future. As of this time only a few small nuclear reactors designed to power manned missions or planetary explorations have been launched into space. Although Russia has launched as many as 35 compact nuclear reactors to power space missions, the United States has launched one such system. This was the SNAP 10A (Systems for Nuclear Auxiliary Power) launched on April 3, 1965. This was, in fact, the first nuclear reactor launched into space and the only US mission of this type. The launch of a nuclear reactor into space has been considered politically controversial around the world because of the radioactive contamination that could occur in the event of a launch failure or in the event the spacecraft was not deorbited safely (Fig. 40.7).

The SNAP 10A spacecraft had several components. These consisted of a compact nuclear reactor, the nuclear reactor controls and associated reflector system, and a heat transfer and power generator system. The nuclear reactors launched by the Soviet Union and Russia are considered to be similar in design. For reasons of safety, cost, and performance requirements it seems unlikely that a complete nuclear reactor system would be used to power future application satellites.[6]

There have been a number of radioisotope-powered systems launched into space by both the United States and the Soviet Union/Russia. These Radioisotope Thermoelectric Generators (RTGs) have been developed under the SNAP Program within the United States. A variety of the RTG systems powered by a number different SNAP radioisotope fuelled power plants have, in fact, been launched by the United States since the 1960s. In fact over 40 such radioisotope-powered systems have been launched – essentially all in support of planetary exploration missions where the duration of the mission and remoteness from the sun has limited the feasibility of solar cell arrays serving as the reliable, longer-term power supply.

The New Horizon spacecraft that will fly by Pluto and its satellite Charon in 2015 and then exit the Solar System into the Kuiper Belt is powered by an RTG. This RTG operates by heating a number of thermoelectric converters that generate electricity with about a 7% efficiency.[7] This low efficiency implies that an RTG will generate more heat than electrical power. The excess heat can be used to maintain certain components warm but in most of the cases the thermal subsystem needs to deal with this waste heat (Fig. 40.8).

These RTG power sources produce heat as the source of electricity when solar power is not a viable power supply and the mission is of a very long duration. The isotopes used in RTGs can be many different types, but the general purpose heat supplies (GPHS) for the Cassini Mission and the New Horizons probe are Plutonium 238 oxide pellets.



Fig. 40.7 SNAP 10A the first US nuclear reactor launched into Earth orbit (Photo courtesy of NASA)

When selecting an isotope for an RTG, it is necessary to consider several mission parameters including mission life, sensitivity of equipment to radiation, and desired temperature range. Plutonium 238 has been the fuel of choice for nearly all RTG-powered spacecraft. Plutonium 238 when decaying emits alpha particles which are easily shielded. In addition, it has a half-life of 87.8 years. Although this is much more than required, it gives the best available power density and half-life time combination. However, there have been difficulties in obtaining the required amount of Plutonium 238 in recent missions, leading to the significant delays. An alternative isotope is Polonium 210, which has in fact much higher power density but the half-life time is much shorter. For missions with long lifetimes, this is a serious limitation. Table 40.1 provides a list of some common RTG isotopes.

The design of an RTG power supply that provides sufficient heat to a surrounding grid of thermoelectric converters is shown in Figure 40.9.[8] To date no applications satellite has had sufficient power need or size to justify the use of an

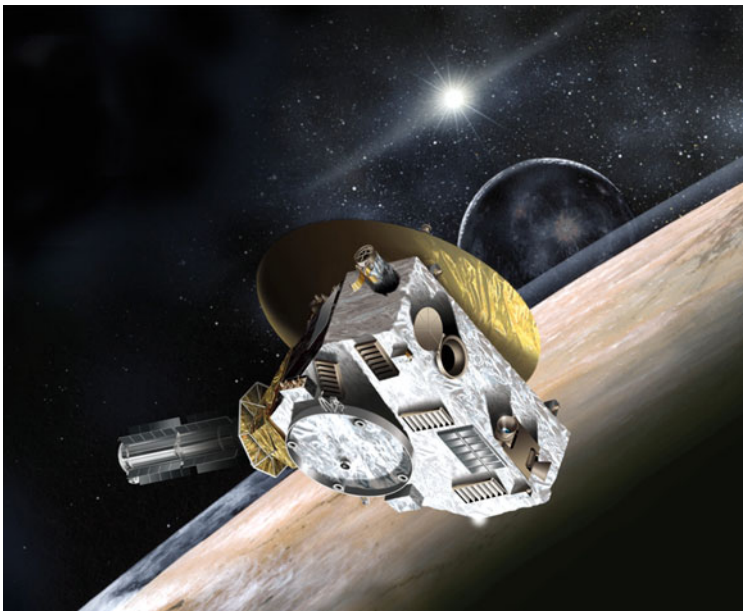


Fig. 40.8 The New Horizon exploratory spacecraft with an RTG power source (Photo courtesy of NASA)

RTG. It is believed by some engineers that if and when electric power system needs reach a level of perhaps 40 kW, radioisotope-powered heat sources working with thermoelectric converters with 10% efficiencies could justify the use of such devices.

This approach would require additional thermal control systems to dissipate the excess heat and would of course entail special safety procedures at satellite launch and with deorbiting procedures. The current applications that involve sending spacecraft into deep space, of course does not raise the deorbiting issue.

Thermal Control and Heat Dissipation

The key element to remember about all application satellites is that they operate in a very hostile environment and there are no repair crews to fix a malfunction. One of the major environmental problems that satellite designers must address is the thermal environment. Regulating the thermal environment is key because many electronics must be kept within a modest temperature range to operate reliably. Further there are special conditions that must be considered. One of these is the sharp thermal gradient that occurs when a satellite moves from an eclipse condition to full solar illumination where the outside of the satellite may heat rapidly within a few minutes. There can also be special conditions where

Table 40.1 Information regarding various radioisotopes

Radioisotope	Class of emitter	Half-life	Watts per gram (W/g)
Plutonium 238	α	87.8 years	0.39
Polonium 210	α	138.4 days	140
Curium 242	α	165.6 days	120
Curium 244	α	18.1 years	2.27

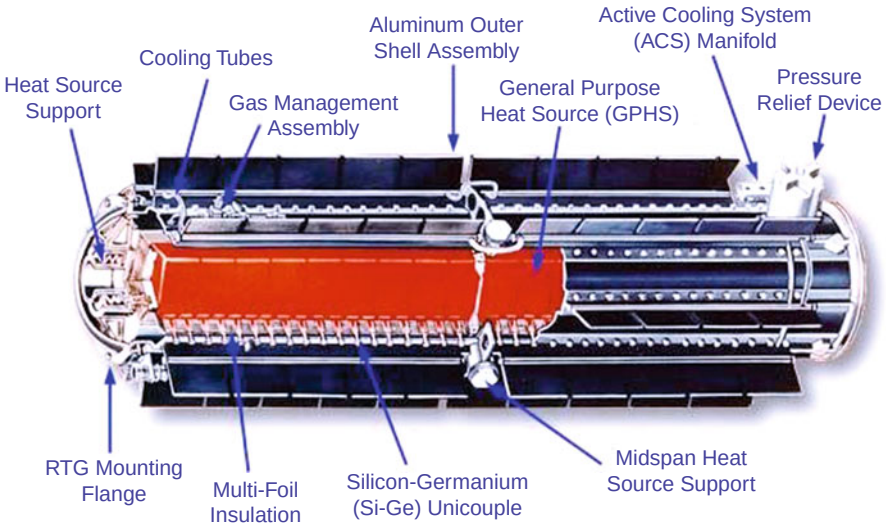


Fig. 40.9 A schematic diagram of a general purpose heat supply RTG (Graphic courtesy of US Department of Energy, Office of Nuclear Energy, Science and Technology)

a remote sensing instrument must be kept in cryogenic conditions to obtain precise information.

Outer space is cold but an object in space is warmed by the sun's irradiation. Thus the spacecraft must be designed so that it absorbs some degree of heat but it also reflects some of the heat so that the spacecraft remains neither too hot nor too cold. Further the power system of the satellite generates electricity and heat. The interior heat of a satellite can build up over time and unless there is a mechanism to transfer the heat inside the spacecraft to the outer shell. There are several ways of achieving this goal. One of the most efficient methods uses a "heat pipe."

A conventional heat pipe mechanism is depicted in Fig. 40.10 that indicates how these devices operate. Heat is applied to the outer casing of the evaporator side, leading to the evaporation of the liquid. The resulting vapor is pushed toward the condenser. The meniscus formed at the surface or inside the capillary structure naturally adjusts itself to establish a capillary head that matches the total pressure drop. The subcooled liquid from the condenser returns to the evaporator as a result

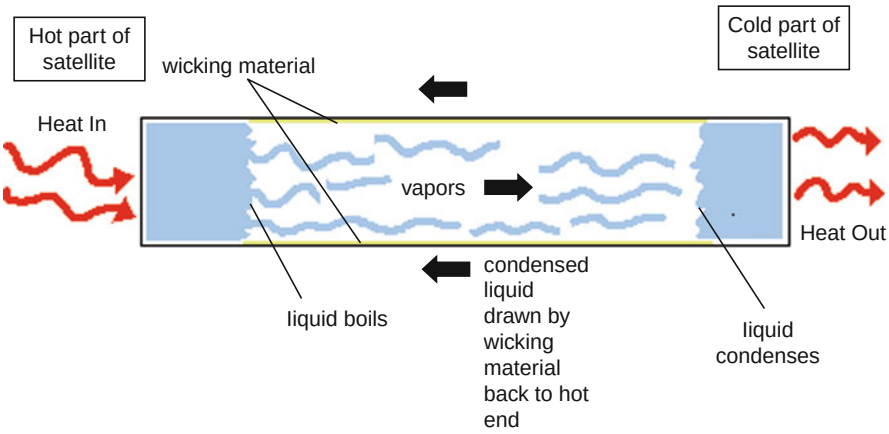


Fig. 40.10 The basic schematic concept for a spacecraft heat pipe

of the capillary pressure, completing the cycle. Thus, the heat pipe operation is passive (no moving parts). One of the main advantages of these types of systems is high-reliability operation without consuming a great deal of the mass budget for the satellite with a minimum demand on the volume that these devices consume. Although the passive and active elements for thermal control and regulation of a satellite design typically involves only about 2–3% of the total mass budget, there are still research efforts to reduce the mass and volume of these systems so as to increase the satellite's payload.[9]

On the other hand, if the thermal controls do not work properly then the entire mission can be jeopardized. The first operational communications satellite with a despun antenna, the Intelsat III had a spinning antenna bearing freeze that was thought to be, in part, the fault of inadequate thermal control. In this case the entire satellite was declared a complete failure and the satellite's bearing and thermal control system had to be rapidly redesigned against production and launch schedule constraints.

Heat pipes have been standard tools in spacecraft thermal control and there is sufficient flight heritage. One important drawback of the heat pipes is that they do not work at adverse elevation under gravity. This is an important limitation for ground testing and may also limit operations, e.g., exploration missions on planetary surfaces or adverse acceleration forces during the spacecraft maneuvers. A more recent version of the heat pipes is a Loop Heat Pipe (LHP), which was invented in Russia in the late 1970s. The operation principle of an LHP is similar to the heat pipe described above. In addition, the reservoir or compensation chamber as shown in Fig. 40.11 provides additional volume for the fluid volume changes, allowing variable conductance operation. The porous structure is only limited to the evaporator section, thus very high capillary pumping can be obtained by using metal-sintered wicks without introducing unnecessarily high-flow resistance. LHPs offer several advantages which make these systems very attractive for the thermally

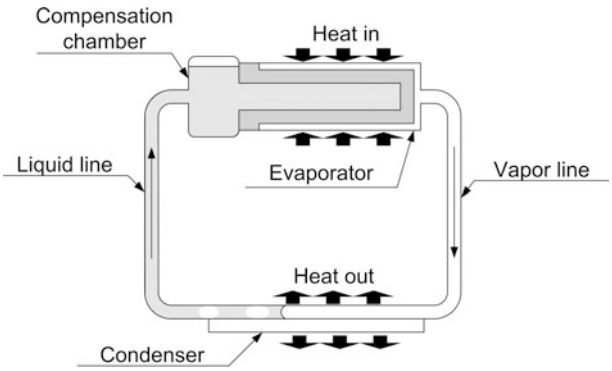


Fig. 40.11 Loop heat pipe (LHP) schematic (not-to-scale)

demanding missions. In addition to the passive operation, they can work against gravity, up to several g loads. They exhibit diode behavior (i.e., if the condenser is warmer than the evaporator, the fluid circulation automatically stops). Therefore, they can be used as a thermal switch. For instruments requiring fine temperature control, a tight temperature range as narrow as ± 0.1 K is also feasible.

LHP transport lines (liquid and vapor lines) can employ flexible tubing. This allows easy integration and more importantly this design allows deployable radiators. The Boeing-702 communications bus was the first US commercial satellite to employ a LHP-based deployable radiator system. Research in this area continues because of the importance of this technology. The development efforts are focusing on the multiple evaporator and multiple condenser systems, operation at higher heat flux, and miniaturization. Figure 40.12 shows a small LHP unit.

Onboard Heaters and Cooling Systems

One of the greatest challenges of maintaining a long life in a spacecraft that is equipped with a large amount of electronics is keeping the temperature within a range that is actually quite narrow. In most cases it is desirable to keep electronic gear within a range of about 5–25°C. There is a large temperature range when the satellite must make the transition during launch from quite hospitable temperatures to the coldness of outer space in a matter of only a few minutes. Once a satellite is in orbit it may fly around the Earth in LEO that every 90 min or so brings it from full sun illumination to a very cold environment as the satellite travels behind the Earth. The use of passive techniques such as Multilayer Insulation (MLI) on the exterior of the satellite and the use of heat pipes can help a good deal, but most application satellites must have more elaborate heating and cooling devices to help keep the satellite's interior in a safe and viable temperature range. These can be closely akin to electric heaters and refrigerant systems used on the ground except they are subjected to extensive lifetime testing processes to insure long life and the ability



Fig. 40.12 A small HP unit for spacecraft thermal control (229 mm × 127 mm) (Photo courtesy of NASA)

to withstand the forces of liftoff. There are also often power converters from DC to AC current and these can also serve the dual purpose of assisting with heating devices.

Conclusion

The optimum design of an application satellite is not only a difficult and demanding process, but one that is constantly evolving as research into a wide range of technologies and materials continues over time. This chapter has briefly addressed spacecraft structure and materials, positioning and orientation systems, sun, star, and Earth sensors, and other systems for maintaining spacecraft pointing, power systems, and thermal control systems.

New materials, not only for the spacecraft bus structure, but also antennas, and solar cell deployment arrays have allowed application satellites to become lighter (with regard to these component parts) and thus allowed more mass (and sometimes volume) for mission payload. Alternatively these advances in structural materials have allowed for reduced payload service costs. Spacecraft are also better able to be accurately positioned and pointed to allow higher precision monitoring or for higher gain telecommunications antennas to function more effectively. Currently the most precisely pointed application satellites, operating from GEO, can be pointed with a directional accuracy of up to 0.25°.

The objective for positioning and orientation systems is not only precision but long life as well. High performance ion engines that are capable of sustained operation over long periods of time allow precise deployment and operation. Over the longer term ion engines may not only provide station keeping and pointing control, but also assist with deployment of satellites from LEO to GEO over a longer term period. Overcoming and solving the orbital debris problem is becoming an increasingly serious issue – especially in LEO. This slow

deployment from LEO to GEO, which increases the risk of collision, poses a deterrent to such a deployment method.

One of the most critical and demanding aspects of an application satellite is reliable and sustained electrical power supply. Power is key to all types of application satellites and failure of spacecraft or short-term outages are most frequently related to power supply difficulties. Over time power supply technologies have increased greatly in capability. Solar cells are lighter and convert energy with higher efficiency. Flexible array structures are lighter in weight. Batteries now have greater power density and last longer, but there are new power storage technologies under development. These new technologies include regenerative fuel cells, flywheel systems, and possibly radioactive thermoelectric generators and thermo-ionic converters.

Finally the performance and reliability of application satellites are dependent on the thermal environment of the spacecraft being maintained within a narrow range of temperature gradients. The thermal vacuum testing of integrated spacecraft and components and thermal control systems have proven valuable over the years. Efforts continue to develop thermal control systems that can handle higher power densities as well as to develop improved passive and active units to help control temperature extremes. Research also continues to develop thermal control systems which can provide temperature control within a very narrow range (0.1 K) for demanding payloads.

The TTC&M systems on spacecraft are critical to their operation and this subject will be addressed in the following chapter. There will be additional chapters that address the reliability testing and engineering of satellites to achieve longer life, the problems of orbital debris, and other issues that are critical to performance of application satellites.

Cross-References

- ▶ Common Elements Versus Unique Requirements in Various Types of Satellite Application Systems
- ▶ Lifetime Testing, Redundancy, Reliability and Mean Time to Failure
- ▶ Telemetry, Tracking, and Command (TT&C)

Notes

1. "Reduced Toxicity Fuel Satellite Propulsion System Including Fuel Cell Reformer with Alcohols Like Methanol" NASA Glenn Research Center, http://technology.grc.nasa.gov/tech-detail-coded.php?cid=GR-50278mini=y
2. RIT: Solving the World's Energy Crisis by Improving the Efficiency of Photovoltaics" https://www.rit.edu/showcase/index.php?id=36
3. P. A. Jones, S. F. White, T. J. Harvey, B. S. Smith, A High Specific Power Solar Array for Low to Mid-Power Spacecraft http://www.aec-able.com/corpinfo/Resources/ultraflex.pdf

4. “A Methanol and Hydrogen Peroxide Fuel Cell Using Non Noble Gas Catalyst in Alkaline Solution http://etd.lsu.edu/docs/available/etd-11052006-193341/unrestricted/Sung_thesis/PDF
5. “Fuel Cells and Hybrid Energy Systems”, NASA Space Architecture Strategic Research Plan, http://www.macrovu.com/image/PVT/NASA/RPC/uc%3DFuelCellV4.pdf
6. G.L. Bennet, Space Nuclear Power: Opening the Final Frontier Fourth International Energy Conversion Engineering Conference (IECEC) San Diego, California http://www.fas.org/nuke/space/bennett0706.pdf
7. “NASA New Horizon: Mission to Pluto and Kuiper Belt” www.nasa.gov/mission_pages/newhorizons/main/index.html
8. “RTG History and New Horizons” www.osti.gov/accomplishments/rtg.html
9. Spacecraft Thermal Control, http://webserver.dmt.upm.es/isidoro/tc3/STC%20missions%20and%20needs.htm

Telemetry, Tracking, and Command (TT&C) 41

Arthur Norman Guest

Contents

Introduction 1068
Telemetry: Providing Health and Status Updates for the Satellite 1069
Tracking: Locating and Following the Satellite 1072
Control: Commanding the Spacecraft Bus and Payload of the Satellite 1073
TT&C System Design Aspects 1076
Cross-References 1077
Notes 1078
References 1078

Abstract

The telemetry, tracking, and control (TT&C) subsystem of a satellite provides a connection between the satellite itself and the facilities on the ground. The purpose of the TT&C function is to ensure the satellite performs correctly. As part of the spacecraft bus, the TT&C subsystem is required for all satellites regardless of the application. This chapter describes the three major tasks that the TT&C subsystem performs to ensure the successful operation of an applications satellite: (1) the monitoring of the health and status of the satellite through the collection, processing, and transmission of data from the various spacecraft subsystems, (2) the determination of the satellite's exact location through the reception, processing, and transmitting of ranging signals, and (3) the proper control of satellite through the reception, processing, and implementation of commands transmitted from the ground. Some advanced spacecraft designs have evolved toward "autonomous operations" so that many of the control functions have been automated and thus do not require ground intervention except under emergency conditions.

A.N. Guest
International Space University, San Francisco, USA
e-mail: guest.arthur@gmail.com

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_69, © Springer Science+Business Media New York 2013

Keywords
Command • Communications • Computer • Control • Data handling • Guidance • Navigation • Range • Satellite • Telemetry • Tracking

Introduction

Regardless of the application it is being used for, satellites represent a complex system of hardware and software. However, the satellite is only one component in the larger system required to provide the service for which it was built and launched. There are three specific segments that must work together for the larger overall system to provide communication, navigation, or any other service of interest (Army Space Reference Text 1993):

- The *space segment* consisting of all satellites required for the application and the launch vehicles used to deliver those satellites to orbit.
- The *command segment* consisting of all the personnel, facilities, and equipment that are used to monitor and control all the assets in space.
- The *user segment* consisting of all the individuals and groups who use and benefit from the data and services provided by the payloads of the satellite and the equipment that allows this use (Fig. 41.1).

Onboard each satellite, the connection between the spacecraft and the command segment is achieved by the Telemetry, Tracking, and Control (TT&C) subsystem. As can be deduced from its name, this subsystem has three specific tasks that must be performed to ensure the ability of the satellite to successfully achieve any application:

1. *Telemetry*. The collection of information on the health and status of the entire satellite and its subsystems and the transmission of this data to the command segment on the ground. This requires not only a telemetry system on the spacecraft but also for a global network of ground stations around the world to collect the data, unless, of course, the application satellite network includes inter-satellite links that are capable of relaying the data to a central collection point.
2. *Tracking*. The act of locating and following the satellites to allow the command segment to know where the satellite is and where it is going. Again this requires a ranging system on the spacecraft and a data collection network on the ground that allows this ranging and tracking function to work.
3. *Control*. The reception and processing of commands to allow the continuing operation of the satellite in order to provide the service of interest. Again a ground system is required.

These tasks must be performed for both of the major components of the satellite: the payload and the spacecraft bus. Earlier chapters detailed the principles and technologies related to specific payloads and applications and the previous chapter introduced the key subsystems present in any spacecraft bus required to support the payload (structure, attitude control, power, thermal controls, etc.). The TT&C

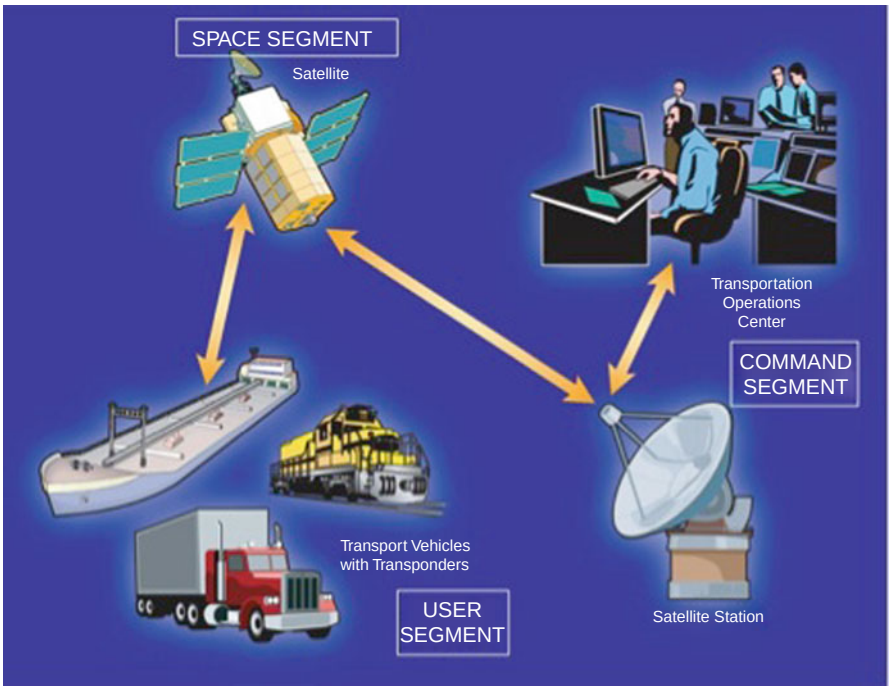


Fig. 41.1 An example of the three segments of a space system (Adapted from energy.gov)

subsystem must be able to gather information from and transmit commands to all of the subsystems in the spacecraft bus as well as to the payload itself. The following sections in this chapter focus on the principles and factors involved with each of the three tasks that are performed by this subsystem.

The design of TT&C systems become more and more complex as one moves from a single geosynchronous satellite operating over a single country to a global geosynchronous network involving a number of such satellites. The most complex and demanding TT&C network involves the operation of a global network of satellites in medium earth (MEO) or low earth orbit (LEO). This is because there are many more satellites in orbit and the orbital trajectories of the satellites are more complex. The relay of data from such MEO and LEO systems is more demanding in almost every respect. The particular strategies that are used to cope with TT&C operations with MEO and LEO constellations are described below.

Telemetry: Providing Health and Status Updates for the Satellite

It is the task of the command segment located on the ground to provide the commands that will keep the satellite operating as required. However, before any commands can be issued or even chosen, the team on the ground must know what the status of the satellite is at any given time. Telemetry is the collection of

measurements and onboard instrument readings required to deduce the health and status of all of the subsystems on the satellite. The TT&C subsystem must collect, process, and transmit this data from the satellite to the ground.

The first step in providing status updates to the ground is the collection of the measurements required by the command segment. Measurements related to the health and status of the satellite include:

- The status of resources (e.g., propellant supply and the health and charging status of batteries)
- The attitude of the satellite (e.g., the readings from sun and star trackers or RF tracking systems)
- The mode of operation for each subsystem (e.g., the on/off state of a heater)
- The health of each subsystem (e.g., output from the solar panels)

These measurements are not only necessary for the spacecraft bus, but also for assessing the health of the payload. On a communications satellite, the telemetry data would include information such as the switching configuration for the routing of signals, power output of the transponders, the direction the antenna is pointed in, or the health and status of imaging systems. All of these measurements are collected with various sensors such as thermometers, accelerometers, and transducers that provide outputs in such forms as measured resistance, capacitance, current, or voltage. The design of a spacecraft for the collection of data from such physical sensors and the associated wiring required for gathering this information concerning the health and status of the spacecraft and payload can lead to a noticeable mass and cost penalty. For example, some larger communications or remote sensing satellites can have up to 500 temperature sensors onboard the spacecraft. The collection of data from such a large number of sensors can lead to an extensive wiring harness (500 temperature sensors $\times$ 2 wires $=$ 1,000 wires). This has led to ongoing research related to new alternatives for collecting information such as the European Space Agency's Fiber Optic Satellite (FOSAT) project which uses fiber optics rather than conventional wiring in order to more efficiently gather data on the health of a satellite (Fig. 41.2)[1] (Ecke et al.).

The use of these sensors to gather the required measurements is only the first step in providing telemetry from the satellite to the command team on the ground. The second step is the processing of the measurements. This processing includes the conversion of analog measurements to digital information as well as formatting of all the measurements for effective and, if required redundant, transmission to Earth. The processing of telemetry data involves two key factors. These two factors involve the nature of the automation patterns of the spacecraft and the data storage algorithms. These will typically be different for each particular application satellite network. These will be based on the specific applications and mission parameters for each application satellite system.

Automation refers specifically to the ability of the satellite to interpret and respond to the telemetry measurements without interaction with the command segment. This allows the satellite to issue commands to the subsystems directly versus receiving and processing the commands from the ground. Automation can typically be found in response to predicable or common faults in certain subsystems

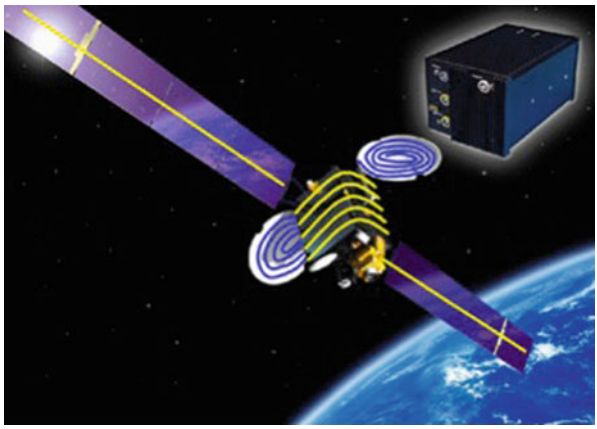


Fig. 41.2 Artist depiction of FOSAT using fiber optics for health monitoring (Courtesy of ESA)

that require actions such as placing components on standby when they operate outside of a range of parameters. Automation allows instantaneous actions to be taken onboard of a satellite and the degree of autonomous operation can vary widely among various applications satellites and the sophistication of their software. Automation can lead to three particular requirements for the TT&C subsystem as noted below (Pisacane 2005):

1. First and foremost is the ability of the onboard software to properly identify telemetry that indicates a subsystem is acting incorrectly and the corresponding ability to identify and process the correct response.
2. Closely associated with the first degree of automation is the related ability for the telemetry and the command components to communicate with each other and pass information back and forth on a nearly instantaneous basis.
3. Finally it is important for the software to have a diagnostic capability that allows the telemetry system to determine if an abnormal reading is caused by another subsystem or by errors in the TT&C subsystem itself.

Data storage of telemetry may also be required as transmissions down to the ground may not be feasible at any given moment. It is extremely costly (prohibitively so) to establish enough ground facilities for a global satellite system deployed in low earth orbit (LEO) to have constant contact with the command team unless there are inter-satellite links onboard all satellites that allows the relate of telemetry data to one or two central data processing locations. Due to this difficulty the satellite software and onboard computers must be able to process and store the data from the sensors on board while waiting for a viable window of communications to open up. The instrument readings and measurements are then transmitted along with other pertinent information such as when each measurement was collected. The cumulative data and exactly when it was collected is essential information for the ground team's data analysis particularly in case of anomalies.

An alternative to data storage is the use of a constellation of satellites that can relay telemetry from any given satellite to specific locations on Earth. One example is NASA's Tracking and Data Relay Satellite System (TDRSS) that consists of nine

on-orbit satellites that relay communications from any other LEO satellite to its ground segment known as the White Sands Complex. This system can also be used for providing uplinks necessary for issuing commands to a satellite. European and Japanese systems have developed similar capabilities.[2]

The final step the TT&C subsystem must perform in providing telemetry from the satellite to the command or ground segment is transmitting the data to the Earth. The principles, concepts, and hardware used for transmitting this type of data such as processing, commutation, multiplexing and antenna, and transponder design are described in ▶ Chaps. 12, "Regulatory Process for Communications Satellite Frequency Allocations," ▶ 13, "Satellite Radio Communications Fundamentals and Link Budgets," ▶ 14, "Satellite Communications Modulation and Multiplexing," ▶ 15, "Satellite Transmission, Reception and On-Board Processing Signaling and Switching," ▶ 16, "Satellite Communications Antenna Concepts and Engineering," ▶ 17, "Satellite Antenna Systems Design and Implementation Around the World," ▶ 18, "Satellite Earth Station Antenna Systems and System Design," ▶ 19, "Technical Challenges of Integration of Space and Terrestrial Systems," ▶ 32, "Digital Image Acquisition: Preprocessing and Data Reduction," and ▶ 40, "Overview of the Spacecraft Bus."

The communications system used for downlinking telemetry may be the same system as that used for communicating the payload data or it may be an independent system depending on the satellite's application. Typical frequencies for the telemetry system include: S-band (2.2–2.3 GHz), C-band (3.7–4.2 GHz), and Ku-band (11.7–12.2 GHz). Other frequency bands can also be employed for different types of application satellite systems. Telemetry communications tend to have a bit-error-rate of approximately 10^{-5}. Telemetry systems that utilize Ku-band frequencies need to make some allowance for rain attenuation in the design of the TT&C (Larson and Wertz 1999).

Tracking: Locating and Following the Satellite

In order to communicate with a satellite, whether it is to receive telemetry or send commands, the command segment must be able to locate, range, and track a satellite accurately. These ranging functions are part of the task of tracking which is performed by the TT&C subsystem. The satellite must first be able to locate and lock onto transmissions between the ground station and satellite. Once the satellite is locked on, the TT&C subsystem determines the range, or lineofsight distance between the satellite and the radial velocity of the satellite. This allows the command segment to know where the satellite is and where it is going.

The process of locating and locking onto a satellite from a ground station is known as carrier tracking. This is most commonly accomplished in application satellite operations by using a principle known as phase coherence. This involves creating what is called a "two-way-coherent." The typical operational mode in this respect involves establishing the downlink communication frequency at a predetermined ratio of the uplink frequency. This allows a synchronization of their phases. Initially the satellite searches and validates a connection to the uplink

frequency based on the predetermined parameters. The TT&C subsystem then implements commands to set the frequency of the downlink communications so it is related to the uplink frequency through a prespecified ratio. This not only allows a ground station to lock onto a satellite, but it allows it to do so quickly as the expected downlink frequency is already known. There are standards for this process. For instance, the transponders tracking setting for the NASA Ground Spaceflight Tracking and Data Network (GSTDN) which sets the downlink/uplink ratios 240/221 (NASA Space Communication: TDRSS https://www.spacecomm.nasa.gov/spacecomm/programs/tdrss/system_description.cfm).

Determining the range between a satellite and a ground station is typically achieved through the use of tones or pseudo-code. The tone or code is modulated to the uplink frequency and when the satellite recognizes it, the TT&C subsystem adds the same tone or code to the downlink. The command segment can then calculate the round-trip time required for that tone and use that information to calculate the distance between the ground station and satellite. With the range (i.e., the distance to the spacecraft) established, the actual location of the satellite can be determined by using the pointing information of the satellite to determine the satellite's azimuth and elevation angles (NASA Space Communication: TDRSS https://www.spacecomm.nasa.gov/spacecomm/programs/tdrss/system_description.cfm).

An alternative means for determining the range also allows for the radial velocity of the satellite to be determined. This method uses the Doppler shift of the frequencies of the uplink and downlink to determine the satellite's location and velocity. As discussed in earlier chapters, the Doppler effect is the change in frequency of the transmissions caused by the relative movement between the transmitter and the receiver. When the satellite is approaching a ground station, the frequency it receives is higher that the frequency transmitted. When the satellite is moving away from the ground station, the frequency it receives is lower than the frequency transmitted. This is also true for the frequency of the transmissions going from the satellite to the ground station. One issue with using the Doppler effect to determine the location of a satellite is that there are always two locations, the true or nominal location and the virtual or mirror location that are possible at any single point in time. In order to account for this the TT&C subsystem has to apply processing algorithms to determine which location is correct. Satellites, such as Argos, use two positioning algorithms: least squares analysis and Kalman filtering. The Doppler shift method is best applied to satellites in relatively low orbits (Fig. 41.3).[3]

Control: Commanding the Spacecraft Bus and Payload of the Satellite

Control or command is the third task of the TT&C subsystem and it is the act of ensuring the satellite's spacecraft bus and payload do what is necessary to meet the objectives of its particular mission. Allowing control of the satellite requires that the TT&C subsystem receive, process, and implement the commands required

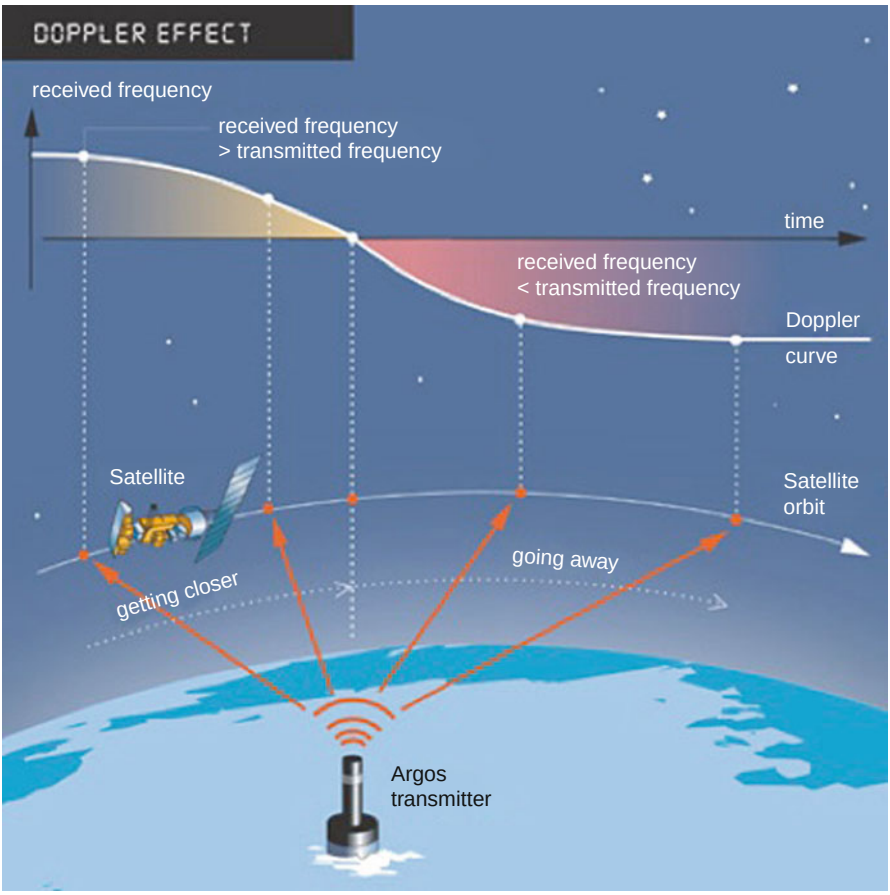


Fig. 41.3 Using the Doppler effect for tracking a satellite (Courtesy of http://www.argos-system.org/)

by the command segment on the ground. As discussed briefly earlier, some commands may be automated through the use of onboard software that implements predefined commands upon recognition of specific conditions. Some satellites designed for "autonomous operation" carry this degree of automation to very sophisticated levels.

Commands are used to reconfigure a satellite or its subsystems to respond to mission conditions. Commands may include switching subsystems and components between on and off states or changing the operating mode in another manner. The commands may be used to control the spacecraft guidance and attitude control or deploy structures such as solar arrays or antennas. Finally, commands may come in the form of software programs that are to be uploaded into the onboard computer to control components on an ongoing basis.

The first step of the command system is to receive the data from the ground through its communication system. This system uses the same principles and technologies

as described in ▶ Chaps. 12, "Regulatory Process for Communications Satellite Frequency Allocations," ▶ 13, "Satellite Radio Communications Fundamentals and Link Budgets," ▶ 14, "Satellite Communications Modulation and Multiplexing," ▶ 15, "Satellite Transmission, Reception and On-Board Processing Signaling and Switching," ▶ 16, "Satellite Communications Antenna Concepts and Engineering," ▶ 17, "Satellite Antenna Systems Design and Implementation Around the World," ▶ 18, "Satellite Earth Station Antenna Systems and System Design," and ▶ 19, "Technical Challenges of Integration of Space and Terrestrial Systems." Typical frequencies for the command system include: S-band (2.2–2.3 GHz), C-band (3.7–4.2 GHz), and Ku-band (11.7–12.2 GHz). Control communications tend to have a bit-error-rate of approximately 10^{-6} (Keesee 2003).

This rate is an order of magnitude less than that noted for the telemetry communications due to the importance of ensuring that the commands issued by the ground are recognized correctly by the TT&C subsystem. Typical data rates required for command systems range from 500 to 1,000 kb/s.

Once the satellite has received and demodulated the uplink command transmissions (or a command is produced by the onboard computer), the command system includes three additional segments: the command decoder, the command logic, and the interface circuitry. The decoder reproduces the command messages and produces the lock/enable and clock signals required. The command logic validates the command and rejects it if there is any uncertainty regarding its authenticity. This logic then gets implemented through the interface circuitry that connects to the other systems on the satellite. Some application satellite systems have sophisticated security processes in place. These might require that at least some ground commands be authenticated from another location in order to be implemented. There have been instances where spurious commands (whether intentional or unintentional) have disabled application satellite networks (Fig. 41.4).

The command decoder collects and processes all incoming commands from sources such as the ground and the onboard computer. The decoder includes an arbitration scheme that determines how each command is given priority in the processing queue.[4] Due to the criticality of the uplink commands, they are often encrypted and as noted above might also require authentication from another TT&C ground facility. Typical command messages include input checkerboard bits, synchronization bits, command bits, and error detection bits. The command itself includes the spacecraft's address and the command type. In virtually all operational application satellite systems both elements of the command must be verified. Command types include commands to flip a relay in a system, to pulse a piece of electronics, to change the output level of a component, or to request or send data to or from a component.

In some cases, the command may involve a whole sequence of events. In the case of the deployment of the Light Squared Land Mobile Satellite that was deployed in 2011, the 20 m antenna system did not normally deploy. In this case, a series of commands were sent to fire thruster jets in a sequence to "joggle" the spacecraft so that the antenna finally deployed.

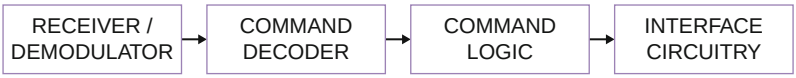


Fig. 41.4 Command system block diagram (Derived from information provided at http://ocw.mit.edu)

The command logic in the TT&C subsystem must verify and validate the command. This includes ensuring that the commands are being sent to the correct spacecraft or that the command itself is valid. Additionally, the timing of the command must be valid and the command itself must be authenticated. Once the logic is used to process the command, the TT&C subsystem activates the interface circuitry as necessary depending on the type of command. In the case of trouble shooting or failure-recovery operations, there may be a need to override constraints in the onboard software to allow higher risk commands to be executed.[5]

TT&C System Design Aspects

Satellites must be designed to operate correctly for the entire life of their mission. The TT&C subsystem is a critical part of ensuring that the satellite performs as required and can react to changes in conditions at the satellite (either internal or external). Because of this, the system must go through stringent quality control and testing of its components before they are used on a satellite. Additionally, key portions of the subsystem are designed to have redundant components to ensure that if one part fails; another is still available for use. If the satellite is designed for a 15 year time for instance, the TT&C subsystem might even be designed with a mean time to failure of 18–20 years.

One of the major trade-offs in designing a TT&C system is how complex the system must be to meet the goals of the mission. The more complex the system, it will be able to provide more telemetry and process more commands. The disadvantage is that this complexity typically leads to more components and therefore more mass and cost to the overall spacecraft. Most of the complexity of a TT&C subsystem is actually contained in the software that includes diagnostics to determine if a particular pathway in an onboard switch is somehow defective. The good news is that one can often update or reengineer software so that improved software can be uploaded after launch.

Connected to the concept of complexity, it the aforementioned use of automation. A designed must look at the requirements for the mission and determine how much telemetry and command should be dealt with onboard the satellite and how much should flow through the command segment on the ground. A thorough review of how the processing scheme for each command impacts the overall mission

reliability and overall cost should be performed prior to launch and reviewed periodically to see if new software could help insure improved and more reliable performance. In some cases, it may be more cost-effective to deal with situations through onboard computing, while other situations may be more cost-effectively dealt with through analysis on the ground.

Conclusion

The Telemetry, Tracking, and Control subsystem enables the critical connection between a satellite and the ground segment. Regardless of the application of the satellite in question, the TT&C subsystem must perform all of its tasks in order to have a successful mission. The three major tasks of the subsystem are:

1. *Telemetry*. The collection of information on the health and status of the entire satellite and its subsystems and the transmission of this data to the command segment on the ground in an accurate and consistently reliable manner.
2. *Tracking*. The act of locating and following the satellites to allow the command segment to know where the satellite is and where is going and with a high degree of precision. Geosynchronous satellites that can be 40 time further out in space than low earth orbit systems require more exacting methods to determine range because of the greater distances involved.
3. *Control*. The reception and processing of commands to allow the continuing operation of the satellite on an uninterrupted basis. Protection of the control commands to prevent spurious commands is just one of the elements that designers of satellite networks must take into account.

The next chapter discusses the important of lifetime testing, redundancy, and reliability. Each of these aspects should be taken into consideration when designing a TT&C subsystem.

Cross-References

- ▶ Common Elements versus Unique Requirements in Various Types of Satellite Application Systems
- ▶ Introduction to Satellite Navigation Systems
- ▶ Lifetime Testing, Redundancy, Reliability and Mean Time to Failure
- ▶ Overview of the Spacecraft Bus
- ▶ Satellite Communications and Space Telecommunications Frequencies
- ▶ Satellite Communications Antenna Concepts and Engineering
- ▶ Satellite Communications Modulation and Multiplexing
- ▶ Satellite Radio Communications Fundamentals and Link Budgets
- ▶ Satellite Transmission, Reception and On-Board Processing Signaling and Switching

Notes

1. ESA: FOSAT – Fiber Optic Sensing Subsystem for Spacecraft Health Monitoring in Telecom Satellites, http://telecom.esa.int/telecom/www/object/index.cfm?fobjectid=28652
2. NASA Space Communication: TDRSS https://www.spacecomm.nasa.gov/spacecomm/programs/tdrss/system_description.cfm
3. Argos User's Manual http://www.argos-system.org/manual/index.html#3-location/32_principle.htm
4. Op cit., Wiley
5. Op cit., Keesee

References

Army Space Reference Text. (US Army Space Institute, Fort Leavenworth, 1993), www.fas.org/spp/military/docops/army/ref_text/index.html

W. Ecke et al., Fiber optic sensor network for spacecraft health monitoring. Measure Sci. Technol. J. **12**(7), 974

C.J. Keesee, *Satellite Telemetry, Tracking, and Control Subsystems*. (Massachusetts Institute of Technology, 2003), http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-851-satellite-engineering-fall-2003/lecture-notes/l20_satellitettc.pdf

W. Larson, J. Wertz, *Space Mission Analysis and Design*, 3rd edn. (Microcosm Press, 1999)

V. Pisacane, *Fundamentals of Space Systems*. (Oxford University Press, New York, 2005)

Lifetime Testing, Redundancy, Reliability and Mean Time to Failure

42

Joseph N. Pelton

Contents

Introduction 1080
Satellite and Subsystem Lifetime Testing 1081
Hazards (Natural and Man-Made) 1081
Testing Strategies 1082
New Concepts 1083
Satellite Lifetime and Mean Time to Failure 1084
Key Components, Subsystems, and Lifetime Expectancies 1085
Optimum Lifetime Engineering and Testing 1086
Strategies to Extend Satellite System Lifetime 1087
Satellite Design and Redundancy 1089
Redundancy and Single Points of Failure 1091
TTC&M 1091
Autonomous Operation and In-Orbit Servicing 1092
Civilian Versus Military Satellite Design Strategies 1093
Cross-References 1094
Notes 1094
References 1094

Abstract

The environment of outer space is quite hostile to the many spacecraft that are now deployed in Earth orbit and beyond. There are many hazards in terms of severe thermal gradients, space weather from the sun and beyond as well as intense radiation from the Van Allen Belts as well as strong magnetic forces. Today, application satellites also must plan to cope with man-made hazards that arise from space debris, RF interference (RFI), and other possible hazards such as spurious commands. There are also risks associated with the launch and

J.N. Pelton
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: joepelton@verizon.net

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_70, © Springer Science+Business Media New York 2013

deployment of satellites since there are strong "g forces" during launch and difficulties that can arise from the deployment of solar arrays, antennas, and other systems that must be unfolded in space in response to remote command. This complex series of hazards requires extensive testing of application spacecraft that are deployed into Earth orbit with the hope of extended lifetime operation. These hazards and difficulties of space operations increase the importance of lifetime testing. It also demands the design of application satellites to be rugged and to have reasonable levels of redundancy so that service can be maintained if various components happen to fail. In the case of application satellites, rugged design, redundancy, and demanding lifetime testing of applications satellites and its subsystems and components are of utmost importance simply because there is little opportunity for repair or refurbishment operations in space. Without these precautions a very expensive application satellite that requires perhaps an even larger investment to launch it into space could be lost to the satellite operator and thus require replacement at very high cost either to the satellite operator or to the companies that have insured the launch and operation of the satellite. The following text discusses all of these strategies for coping with and minimizing risk for the satellite applications industry.

Keywords

Accelerated testing • Acoustical testing • Anechoic chamber testing • Autonomous operation • Deployment risks • Inclined orbit operation • Independent verification and validation (IV&V) • In-orbit incentives • Launch insurance • Launch risks • Lifetime testing • Mean time to failure • Pogo effects • Radiation • Redundancy • Reliability • RF interference • Satellite operational and launch insurance • Space debris • Subsystem testing • Thermal vacuum testing • Van Allen belts

Introduction

This chapter addresses the various strategies that operators of application satellite systems can undertake to increase the reliability of their space assets, ground systems, and overall system operations as well as to manage risk against various types of losses that can occur. These strategies include the following: (1) constant identification of and knowledge about various types of risks and hazards – natural and man-made (i.e., space debris, conjunction of other spacecraft, and RFI); (2) design of space and ground systems to withstand these risks, to have significant link and operating margins, or to have redundancy or spare components to restore failed systems; (3) design to eliminate as many single points of failure as possible; (4) carry out sophisticated testing strategies to identify weak or flawed elements against manufacturing mistakes, and to check space systems against a lack of tolerance to mechanical, vibrational, electronic, radiational, thermal, RF interference, power outages, as well as conduct deployment tests for antennas and solar arrays (this is done either through component, subsystem, or fully integrated

satellite testing); (5) provide for independent expert oversight of manufacturer design and testing; (6) write contractual provisions for incentives for reliable performance to manufacturers – including in-orbit incentives for successful operation; (7) provide for various forms of launch and operational insurance; (8) employ extensive computerized monitoring systems to track spacecraft health and operational parameters and where possible to evolve toward Autonomous Operation of spacecraft where onboard artificial intelligent or expert systems can anticipate problems and maintain maximum reliability of operation; (9) take evasive action and precautions such as avoidance of known space debris, powering down of spacecraft when major solar flares are ejected from the sun; (10) design of ground systems with a high level of redundancy, backup power, security codes against spurious commands, backup tracking, telemetry, command and monitoring (TTC&M) facilities and redundant areas of global coverage as well as constant training and education of satellite operators.

These efforts to provide for reliability, longer spacecraft and system lifetime, better management oversight, better trained personal, redundancy, and various forms of risk management including insurance coverage almost always add to cost. Commercial operators are thus constantly struggling against trying to add reliability, redundancy, and lifetime while minimizing cost. Different satellite system operators have thus not surprisingly devised different strategies to minimize risk and cost while maximizing reliability. Some have put the greatest stress on design, while others have pursued accelerated or altered forms of testing, while some others have relied more on various forms of insurance or contractual solutions such as having the contractor deliver the spacecraft in orbit after check-out and verification tests.

Satellite and Subsystem Lifetime Testing

Hazards (Natural and Man-Made)

The key element in trying to design in reliability for space assets is to understand as well as possible potential risks. In space the risks are many and they vary over time. Risks from space weather include various types of radiation and solar storms that follow an 11-year cycle that are the greatest risk during solar max. The Van Allen belts between low Earth Orbit and Medium Earth Orbit represent a major hazard to application satellites that operate in these types of orbit. Medium Earth Orbit satellites that operate just above these belts, for instance, often add additional glass coating on top of the satellite solar cell arrays to mitigate the deterioration over time of the solar cells effective power output. Even Geosynchronous orbit satellites although safety above the highest Van Allen belts must travel through these high-radiation zones during the launch operations. Military and defense satellite systems are often designed as so-called radiation-hardened (Rad-Hard) facilities. These "radiation-hardened" satellites have many design features to protect the electronics and power systems of these spacecraft. These designs are intended to provide protection against natural radiation as well as man-made

radiation. Some systems have been designed with a "Faraday Cage" to protect against the radiation and electromagnetic pulse (EMP) of a nuclear blast in outer space.

There are also systems that monitor the sun for evidence of intense solar flares. Since it takes 8 min for light to travel from the sun, but quite a bit longer for the particles from a solar storm (i.e. typically in the range from 24 to 56 hours), there is often ample time to "power down" satellites and to go into "safe mode" when such a violent solar blast occurs. Most application satellites are thus designed to provide at least some degree of protection against radiation, electromagnetic disturbance, thermal extremes of hot and cold (through thermal reflectivity, heat pipes, etc.). They are also tested against the violent shaking and vibrations that occur during rocket launches, including the so-called pogo-effect or oscillations that can occur between the various stages of a launch vehicle and the fairing structure that contains the payload. The design is also constructed so that the antennas of the spacecraft can be tested on the ground to verify their gain and beam shaping capabilities and their electronics and filters tested to screen out unwanted RF interference and background noise. The most recent hazard that has become of concern to application satellite operators is the increasing amount of orbital debris – particularly in low Earth orbit. Currently the US Space Command is monitoring on the order of 20,000 pieces of orbital debris which is of the size of a human fist or larger and there is an estimated 500,000 pieces of orbital debris of the size of 1 cm or more. Such debris, moving at super high orbital speeds can still do significant damage. Efforts have now been made to create data centers whereby satellite operators in Geo-orbit such as Intelsat, SES Global, and Inmarsat can share data about their satellites' orbits and times of close conjunction of their spacecraft. This coordination process continues to expand and soon it will cover spacecraft operation in the lower orbits.

Testing Strategies

The initial testing strategies for verifying the reliability and projected lifetime of application satellites were those techniques developed by the early Intelsat system and the manufacturers of their spacecraft. Intelsat, and their early System Manager, Comsat, decided to rely on a combination of techniques rather than a single strategy. Components, subsystems, and fully integrated satellites were subjected to vibration and acoustical testing, thermal vacuum tests, accelerated lifetime testing of units such as spacecraft batteries, etc. These tests were carried out by satellite manufacturers, but Intelsat engineers monitored the tests and provided oversight of the design, engineering, and testing process. Tests were also carried out to test the performance of antennas and RF electronic systems in anechoic chambers and testing range sites. Intelsat also provided for increasingly more comprehensive launch insurance against launch failure. It also structured its contracts so that the manufacturer was paid so much against delivery milestones, but it reserved a large final incentive payment after the satellite was launched and had

performed successfully in orbit. The reasoning was that the manufacturer thus had not only an encouragement to build the spacecraft but to see that it operated successfully in space. Experience over time showed that if a spacecraft could operate successfully over the first month in orbit it would likely remain successfully in operation until worn-out components begin to fail toward the end of the satellite's life. This lifetime projection curve is sometimes call the "bath tub curve" because failures occurred swiftly at the beginning, then there was a smooth steady state for years, followed by a rapid increase of failures at the end of life – much like the shape of a traditional bath tub.

The first satellites had only limited lifetime expectancies in the range of 18 months to 5 years. Over time as the satellites became larger, more sophisticated and with lifetimes up to 15 years and longer, more elaborate and demanding tests were developed. The testing process often comprised 20–35% of the total spacecraft manufacturing cost. With accumulated experience the satellite industry and the satellite manufacturers began to think of different strategies for lifetime testing and for reliability engineering.

New Concepts

As new satellite operators began to deploy satellite systems the quest to find ways to deploy reliable satellites while containing costs as much as possible, began in earnest. Some satellite operators deserted their own research and development programs and abandoned the use of oversight engineers and thus left all reliability concerns to the manufacturers. Some moved to buying their satellites based on a proven spacecraft bus series. In this process they benefited from the economies of scale since the initial nonrecurring design and engineering costs had already been recovered. Further they thought that if they were buying not the first three of a satellite series, but instead buying three satellites very much like the ones that had already been manufactured more than a dozen times, then weaknesses of design or component manufactured would have already been corrected.

Other concepts were more daring. The Iridium Satellite System which was a large constellation with some 66 operational satellites, a dozen spares, and with a final production run of some one hundred satellites decided that it could streamline its engineering, manufacturing, and manufacturing process by designing in the quality of its component and component testing and then carry out an accelerated testing program. At the end of the Iridium manufacturing process, a complete satellite was being produced in less than 5 days. This stood in contrast to large Geosynchronous satellites that might be in production and testing for well over a year and sometimes over 2 years. Although there were a number of early satellite failures due to mechanical and electronic failures as well as operator errors, the final Iridium constellation was able to achieve a combined network lifetime reliability record that exceeded over 500 satellite-years in orbit.

Increasingly the satellite industry has come to rely on the satellite launch and operational insurance industry to provide a key element of its risk management

strategy. Initially the insurance and risk management companies were reluctant to insure satellite launches. This was because launchers were much less reliable than today, the cost of the launchers and satellites were high, and the industry had little experience with this type of high risk coverage. Initially, insurance coverage only applied to two launches in succession and the premiums were high. By the 1980s, however, the launchers were more reliable and the insurance industry had become more familiar with space industry practices. Thus the reinsurance process was able to spread the risk over many different companies so that the exposure by any one company was comparatively low. Organizations with lots of launches, good technical oversight, and a good track record were able to get launch insurance for any particular launch for as low as around seven or eight percent of the risk exposure. Smaller organizations with fewer launches without a known track record of course paid higher rates. Today a typical launch operation requires satellite operators to pay in the range of 15–20% of amount seen as the "total risk exposure" for an application satellite launch.

Organizations insure not only against a launch failure and the loss of a spacecraft, but they also take on liability insurance against some form of catastrophic loss such as an event where a launcher goes off course and lands in a city with a huge loss of life and property. In many cases governments that host launch operations provide some level of liability coverage against such a catastrophic loss above a level such as $100 million. Some organizations today simply plan that they will pay a premium of up to 20% for insurance against a launch failure and loss of a spacecraft. This is seen as particularly prudent in cases where the entire satellite network consists of only one, two, or three spacecraft. A few organizations such as Intelsat have felt when premiums seem at their highest to self-insure against a launch failure. It is possible to insure against other risks such as a "crash between satellites" or orbital debris. The cost of this type of insurance was once quite modest such as $50,000 per satellite per year, but these premiums have risen sharply in recent years.

Satellite Lifetime and Mean Time to Failure

The world of satellite reliability largely operates in the domain of collective probabilities. The projected "mean time to failure" (MTTF) comes from combining multiple risk factors based on assessed risk of components and subsystems. If a satellite has 1,000 parts each with a 0.999 assessed reliability for 7 years, this might appear to constitute a very high overall reliability. The combined risk, however, falls significantly when the 1,000 component parts are combined to calculate a cumulative rate of failure. The actual risk assessment is far more complicated because some components are more reliable than others and have a proven track record in space. Some other elements may constitute a single point of failure, while others may be backed up by redundant components.

In such cases it may be straightforward to replace a failed power amplifier by simply switching over to a backup amplifier to continue reliable operation. In the case of a solid state power amplifier that is very light in mass, redundancy is

a reasonable design choice. In the case of a momentum wheel for stabilization of a spacecraft, the cost, weight, and complexity of design lead to this key subsystem to be a single point of failure. The process of reliability assessment based on component and subsystem design specifications produces a bell curve of projected times of failure.

The mean time of these various projections is when the satellite is seen as most likely to fail. A satellite with a mean time to failure of 15 years will have sufficient batteries, solar cells, station-keeping fuel, and backup components to last much longer than its MTTF. In fact the satellite may last for 5, 7, 10, 15, 18, or 20 years due to millions of different factors, but the 15 years MTTF will be the best projection the satellite designers can make on the basis of data available at the time of launch.

Key Components, Subsystems, and Lifetime Expectancies

The design of a space system for reliability has several key elements. One of the hardest challenges is to recognize those elements that represent a potential single point of failure on a spacecraft and to ensure the optimum design for those parts and effective test of those components prior to launch. Another key element is to provide for redundancy of parts or elements where it is feasible and cost effective to do so. Yet another element is to provide for sufficient margin in expenditures or for systems that wear out or deteriorate in performance over time. Although these concepts seem straightforward in terms of designing for reliability they prove difficult in practice.

In the case of single points of failure some elements such as the TT&C system, the deployment of antennas, the stabilization and pointing system are obvious and are given a great deal of attention in the design, engineering, and testing process. Where things can go wrong in this process is unfortunately very numerous. Some tasks such as electronic power converters can seem very easy because it has been done many times before. Thus, such tasks can be given to junior engineers. But if the power system fails, the entire satellite fails as was the case with power converters on the Intelsat V series. Thus one key rule is to let single point of failure "trump" assumed ease of design and lack of engineering complexity. Design reviews should consider single points of failure and not skip over simple elements like power converters just because they do not seem to present a challenge. Also it is important to recognize that two or more failures can combine to create a catastrophic failure.

The more complex a design is and the more subsystems or components included in a satellite or launch vehicle the more likely that multiple failures can combine to create a total system failure. In the case of a US military satellite, a low gain omni TT&C antenna was eliminated from a satellite for budgetary reasons. This satellite because of a problem due to fuel sloshing combined with satellite commands went into a flat spin (rather than a vertical spin). In this mode the satellite antennas were spinning around instead of pointing constantly toward Earth. With only the high gain antenna available for commands, it was not possible to command the firing of jets to allow the satellite to be recovered from flat spin. This particular combination of problems and design changes resulted in a catastrophic failure of the satellite.

The addition of redundant parts that can be called into service by a command to the satellite sounds like a very good idea. The problem is that every addition to the satellite increases its volume, mass, and complexity. The challenge is to decide where the best and wisest investment in redundancy makes sense in terms of prolonging the useful life of a satellite without unduly increasing its mass and associated launch costs. Redundancy in the crucial communications electronics of a communications or navigation satellite or having backup components in the key sensors of a meteorological or remote sensing satellite makes sense in that these are key to mission's performance and electronics and computer chips are lightweight and small in volume.

The "sparing philosophy" for application satellites is closely tied to probabilities analysis. At one time designers opted for one redundancy in crucial electronics. Later, based on in-orbit experience they began to opt for one backup repeater on communications satellites for every two devices with the ability to switch to the backup device when either one of the two repeaters failed. In the case of remote sensing satellites, space navigation and meteorological satellites the sparing philosophy is more complicated in that there tends to be one sensor of each type rather than multiple transponders. Nevertheless, one can still provide for redundancy within the electronics associated with these sensing or transmission devices.

Finally there is the issue of providing additional performance margin in elements of the satellite that deteriorate over time or involve expendables. In the case of power systems, it is not possible to provide completely redundant systems but one can provide additional solar cells to support peak power requirements for the projected end of life. Likewise, it is possible to provide additional battery capacity to support power needs at the projected end of life. Also it is possible to add fuel to stabilization and positioning systems. Additional fuel can be added to let a satellite to be maintained in orbit more accurately and for longer periods of time. When a satellite is constructed it is designed against a set of mass and volume constraints with some margin as needed as it is finally engineered and manufactured. At the time of launch there is usually some mass margin left and at that stage additional fuel is often added up to the lift capacity of the launching system.

In the case of Geo satellites there are also operating strategies that can be employed to extend lifetime. The gravitational pulls on a satellite in Geo-orbit are ten times stronger in the North–South (i.e., latitude) direction as opposed to the East–West or longitudinal direction. Some operators as a satellite moves toward its end of life allowed a GEO spacecraft to go into a slightly "inclined orbit" that moves in an "S-shaped curve" above and below the equator while maintaining the specified and assigned longitude position. This allowed lifetime extension at quite low cost. As long as the inclination did not build up to over 5° off of the equator this added only minor risk to the continuity of service.

Optimum Lifetime Engineering and Testing

The world of applications satellite engineering is quite different from the world of human space flight. In the case of application satellites, the objective is to design

and manufacture satellites of high reliability over reasonably projected lifetime without incurring excessive costs. In contrast, the "man-rating" of a spacecraft has typically involved adding demanding margins on top of demanding margins and almost endless testing. In the case of the Lunar Excursion Model, the main contractor, Grumman, was required by NASA to construct ten major test articles of the complete craft. In fact, Grumman, in its zeal to produce a no-fault vehicle ended up constructing 29 test articles including complete structure, thermal and electrical models (Williamson 2006).

In the world of manufacturing application satellites, the spacecraft designers and manufacturers strive for extremely high quality and engage in extensive testing, but to do so in a cost-effective manner with only a reasonable number of tests. Current strategies used by manufacturers include the use of standardized platforms for various types of application satellites sized to different launchers and fairing enclosures. This not only reduces the engineering and manufacturing costs but also allows reduced testing after a particular platform design has been tested on the ground – and in space – a number of times.

The same is true for components such as batteries, heat pipes, thrusters, solar cells, application specific integrated circuits (ASICs) as well as their substrates, etc. Such component parts are increasingly standardized in design and manufacturing techniques for similar reasons. The problem standardization of space qualified and proven components in terms of their design, manufacture, and testing is that this tends to block design innovations. Once a component or subsystem has been standardized and proven in space in a number of missions it becomes "locked in" and design upgrades become difficult. One finds anomalies such as several generations old computer chips on spacecraft that have been "space qualified" some time ago. Thus one can find contemporary spacecraft with very slow processors even though one can find off-the-shelf processors commercially available that can work at much faster speeds and purchased at much lower costs.

The design of a reliable spacecraft thus involves a number of judgment calls. Does one use an older and well qualified ASIC or perhaps use a redundant new ASICs that can process information four times faster? How much glass coating should be applied to solar cells to slow the effects of space-based radiation? Does one specify multiple and redundant testing of components, subsystems, and a fully integrated satellite or can some tests be skipped? If one is deploying a large-scale constellation of LEO satellites with scores of satellites in the production process can one eliminate qualification tests if earlier satellites have passed with flying colors? The logic of such an accelerated testing program in such a case could be that there are multiple spares being launched and thus if one or two satellites should fail in orbit, spares can quickly replace the failed spacecraft.

Strategies to Extend Satellite System Lifetime

As noted in the previous section, different types of systems can be designed for different levels of reliability and projected lifetimes based on the orbit used, the

type of network deployed, and how it is operated. Geo satellites are much more expensive to launch; can be designed for longer life because much less gravitation pull degrades their orbit, in contrast to Leos; and far fewer of them are needed to deploy a viable operating network. In fact, one Geo can create a fully operational system for a country, a region, or even transoceanic service. For all of these reasons these satellites tend to be designed for the highest levels of reliability and the longest lifetime. Key strategies that are available for lifetime extension include: (1) inclined orbit operation; (2) sparing philosophies; and (3) constellation deployment schemes.

Inclined orbit operation was briefly described above as a way to extend the life of Geo satellites and does not apply to Meo or Leo constellations. This mode of operation is simply a strategy to conserve onboard station-keeping fuel. This is a viable strategy for extended lifetime operation only if the satellite is otherwise functional in terms of its power systems, antenna and electronic capabilities and TT&C systems. Operators of Geo satellite networks are assigned locations in the orbital arc between 0° to 360° in the equatorial plane and they are expected to maintain their satellites within a half degree East or West of their assigned location. Excursions in the North or South direction (i.e., latitude) are not as rigidly controlled.

Operators are normally expected to maintain their satellites within 5° North or South of the Equator. The tilt of the Earth's axis and the gravitational forces of the Sun and Moon, however, make it challenging to keep a satellite exactly on the Earth's Equator. Relaxing the "box" within which one seeks to keep a Geo satellite and allowing inclination to build up can add years to the end of life for the spacecraft. The alignment problem for Earth Stations pointed to such a Geo satellite becomes most difficult for those in the subsatellite location in equatorial countries, especially where narrow spot beams are in use.

There is a simple mechanical device that can be added to the pointing mechanism on all affected Earth Stations. This mechanism allows "tracking" of the satellite as it moves above and below the equatorial plane on a 24-h-a-day cycle. Organizations such as Intelsat, SES Global, etc., have used this technique to extend the lifetime of their Geo satellite by 1, 2, or 3 years. Usually other components of the satellite die out, such as the power system, and thus end the spacecraft life. Nevertheless, these types of inclined orbit satellites can be used as emergency spares or to provide service until replacement satellites can be launched. With today's increased concern with orbital space debris there is currently increased concern that Geo satellites retain at least a sufficient amount of fuel in their tanks to raise them out of the Geo-orbital arc.

Sparing philosophy addresses not the reliability of a satellite, but the ability to restore service in the case a particular satellite should fail. In this regard different types of applications have different degrees of service standards and Leo and Meo constellations have different requirements than Geo systems. Telecommunications satellites and space navigation satellites have the highest requirements for continuity of service in that even the slightest moments of outages can have the most significant requirements. Real-time communication satellite networks seek to achieve Integrated Service Digital Network (ISDN) standards of 99.98% reliability

that allow for only about 100 min of outage in a year's time.[1] Space navigation satellite systems seek an even higher standard in that systems such as the Global Positioning Satellite system is used for such vital functions as assisting with the takeoff and landing of aircraft. Telecommunications and Space Navigation satellites have live spare satellites in orbit with the ability for fast and hot switch over from an operational satellite to a spare. There are also Mutual Aid Working Group (MAWG) procedures that allow for rapid transfer of telecommunications traffic from fiber optic systems to satellites and vice versa. Soon there will be a varied of space navigation satellite service (SNSS) networks in orbit that can mutually reinforce one another for vital navigational services. In essence these various networks such as the US GPS, the Russian GLONASS, the Indian Regional Space Navigation System, the Compass/Beidou System of China, the Quasi-Zenith System of Japan, and the Egnos/Galileo System of Europe will, in time, work to mutually reinforce one another (although each will have their own sparing capabilities and sparing philosophy). These attempts to coordinate the interoperability of these various systems are being carried out through the International Committee on Global Navigation Satellite Systems (ICGNSS). These systems, working in tandem will essentially allow the entire collection of space navigation satellites to achieve 100% availability.[2]

Remote sensing, earth observation, and meteorological satellite systems, of course, also seek a high degree of continuity of service, but because of cloud cover, the need for processing and "ground truthing" of data, the demand for absolute and total system availability, and thus the need for extensive "live sparing" of satellites in orbit are less than the case with telecommunications and space navigation systems. The various systems that exist around the world help to supplement one another and provide a reasonable degree of backup, especially in the case of a satellite failure among governmentally operated systems. In short remote sensing and meteorological satellites tend to be launched on a scheduled replacement timetable. If there are short periods of time where a system is not fully populated then sharing of data among various international systems tends to provide a reasonable degree of backup. The demands for more and more extensive data monitoring related to climate change are altering this picture and an ever larger number of climate change monitoring satellites are planned to be launched in the next few years.

Satellite Design and Redundancy

Crucial systems engineering and optimization decisions are made at the outset when various types of telecommunications, space navigation, remote sensing, and meteorological satellites are being designed. Initial decisions are made as to the expected performance and capabilities the satellite will have and the anticipated mean time to failure objective set. Geo-orbit satellites today tend to have lifetimes in the 12–18 years range. Meo satellite constellations are roughly in the 10–15 years range and Leo satellites because of gravitation pull and atmospheric drag have the shortest lifetime of only 8–10 years. There are exceptions to the above lifetime

ranges for application satellites. Nevertheless, these are "normal" lifetime expectancies based on the design of power systems, fuel for thrusters, and other systems that degrade in performance over time. Only small incremental investments in terms of batteries, solar array cell size, fuel for orientation and station-keeping thrusters, and redundant transponders or electronics are needed to extend the lifetime of a satellite. One might, therefore, assume that operators of application satellite systems would automatically design for the longest possible lifetime. This, however, is not the case. The reasons "against lifetime extension" are severalfold. Succinctly stated these reasons are: (1) *Obsolete Technology*. Satellites are very much like specialized digital computers in orbit. The resolution of sensors improves. ASIC chip technology races ahead. Solar cells become more efficient. In light of rapid technical innovation in this field a satellite can become obsolete in less than a decade. Lifetime extension in such conditions becomes a questionable proposition. (2) *Unanticipated failure modes*. One can add a number of elements that can potentially add years of life to a satellite. The problem is that there can be unanticipated failures due to a violent solar flare, an electric power connection to a momentum wheel, etc. that turns into a single point of failure that disables the entire satellite. The addition of fuel, longer-lived power systems, and redundant transponders do not guarantee longevity in the harsh environment of space. (3) *Higher Launch and Satellite Costs*. Additional redundancy or add-ons to extend satellite lifetime can lead to higher launch and satellite costs. Although one might add fuel up to the launcher's capacity or one might add redundant lightweight computer chips, but that is all some operators feel is prudent. The investment in additional batteries or solar cell arrays is seen as adding unnecessarily to the mass and cost of the satellite and thus also adding to the launch cost. System designers look at features such as lifetime extenders or capabilities such as inter-satellite links as an "opportunity cost." In short one can invest in features such as more batteries, solar cells, or inter-satellite links but this prevents the "opportunity" to have a bigger payload for remote sensing, meteorological imaging, telecommunication services, or higher power space navigation signals.

Some believe that, particularly with low earth orbit constellations – where many dozens of satellites are launched to complete a system – the key to extending system lifetime is in the sparing philosophy. The concept is to launch a number of operational spares and to have more on-ground spares that can be launched as needed. In this scenario fuel can be added but the lifetime extension is largely accomplished through sparing philosophy. This, in any event, was largely the approach utilized with the Iridium and Global space mobile satellite systems in the cases of their Leo constellations. The point is that in designing satellite systems and planning for their reliability, redundancy, sparing philosophy, and lifetime extension concepts a wide range of options are considered against cost optimization formulas. This economic optimization involves not only the design of the satellite themselves in terms of capacity, performance, reliability, etc., but also the optimization of the orbital configuration that is used in terms of orbital height, number of satellites and spares, and special requirements such as inter-satellite link (i.e., cross-link) communication.

Redundancy and Single Points of Failure

After first level of system design and optimization is completed that results in an initial satellite and system network design, the second level of analysis focuses on what are called "single points of failure." A three-axis-body-stabilized spacecraft has a single momentum or inertia wheel that spins at thousands of revolutions per minute to keep the satellite constantly pointed toward Earth. Redundancy of this massive system is not cost effective to consider and thus great pains are taken to ensure that this system is highly reliable, tested on the ground and on previous satellite missions, and engineered so that it never loses power. The deployment of the satellites' solar arrays and the communications and/or TTC&M antennas, the release of the satellite from the launch vehicle, and a number of other possible failure modes are identified as ways in which a single mishap can completely end the mission. The experience of launching application satellite over the past half century lend important insight and are highly instructive in identifying these critical failure modes. Electronic power converters, exploding batteries, stuck bearings, solar storms, antennas, or solar cell arrays that will not deploy, loss of a command channel to a satellite, over heating or freezing of a satellite's electronics, even miscalculation as to whether a launcher is deploying one, two, or many satellites can represent the difference between success and mission failure. These various critical mission systems, components, or operation are thus given intensive attention at the design, engineering, manufacturing, qualification testing, and deployment and operation stages. One of the key elements of independent verification and validation is the sharing of data through Knowledge Based Information (KBI) networks so that lessons learned from earlier failures or problem recovers can be shared. Thus standards for design and testing come from experience gained from application satellite programs around the world. Each time a satellite is manufactured, launched, and operated it adds to the knowledge about vulnerabilities and ways to make future satellites more reliable.

TTC&M

The payload that provides communications, space navigation, remote system, or meteorological imaging is, of course, why an applications satellite is launched. Nevertheless, it is the spacecraft bus that supports the successful operation of the spacecraft payload. In terms of an analogy, the platform or spacecraft bus is like the "school bus" that allows the delivery of its passengers to the right location in good health. If the "bus" breaks down or malfunctions the "critical services" do not get delivered. The key means by which the "bus" is able to operate successfully is via the tracking, telemetry, and command systems that allow the operators on the ground to know where the satellite is, whether the spacecraft is oriented in the correct position, whether the power and thermal systems are performing correctly, or whether there are any subsystem anomalies or failures. The Command system

allows thrusters to be fired or backup components to be activated in case of electronics or mechanical failures.

In many satellites there is also a "Monitoring" function that checks that the payload (as opposed to the platform) is performing correctly and is not experiencing interference, system overload, or pointing errors. Some of the most advanced satellites have been designed with expert systems and/or artificial intelligence to allow the satellite to engage in "autonomous or semiautonomous operation." Such satellites are thus more independent of ground control and troubleshooting capabilities. For most application satellites, however, the loss of TTC&M function – even for a short while – can spell the death of a satellite. This is why the design and operation of the TTC&M subsystems to have continuous access from the ground control centers to the spacecraft's TTC&M antennas remains so critical. This required to always stay connected – either to ground controls or computer monitors – which is why there are higher gain TTC&M antenna systems plus backup "omni antennas" that can be commanded from any angle in order to help recover from outage with the higher gain antenna systems.

There is another type of concern involving satellite commands. This is the problem of "hackers" or other organizations or people with hostile intent sending spurious commands to application satellites that would temporarily or even permanently disable them. To protect against unintentional or intentional sabotage of satellites most operators have not only special codes that must be employed, but also require that the properly encoded commands be sent from at least two rather than a single location. There have been recorded cases of "hacker attacks" against commercial application and military satellites to send them commands that resulted in their loss of pointing orientation or to power down their operation. Fortunately, these have resulted in temporary loss of service and no permanent satellite failures.

Autonomous Operation and In-Orbit Servicing

The planners of next generation satellite systems have given increased focus to the possibility of autonomous operation and in-orbit servicing. In the case of "autonomous operation" there are dual objectives of reducing operating costs while at the same time increasing reliability. The truth of the matter is that a number of satellite failures have been due to operator errors when a number of operations are required in a condensed period of time.

The thought has thus been to automate as much of the TTC&M operation as possible and to have onboard computers (or their backup processors) to be able to manage the operation of the spacecraft power, thermal, and pointing functions as possible. These systems would work to keep onboard systems always within a specified range of parameters and to operate with prearranged recovery procedures based on years of operator experience.

Although there can be emergency operation and ground-based overrides, it is thought that computer-driven satellite operations supported by millions of lines of code can allow a very great reduction in the need for large ground crews working 24/7

shifts while at the same time actually decreasing operator error. This would seem to be particularly true with regard to the operating of very large-scale Leo satellite constellations where 60 or more satellites are under active management at all times. The same objectives for autonomous operation also apply to Meo and Geo systems.

Civilian Versus Military Satellite Design Strategies

Today, there are a large number of military communications, remote sensing, and space navigation systems that in many ways resemble commercial and civilian application satellites, but they are different in their design, engineering, and manufacture in many ways. The military systems tend to be significantly more expensive. This is because of efforts to make the military systems more robust in the case of attack and where possible to make them more reliable. Some efforts such as radiation hardening of military systems represent clear-cut differences. These satellites also operate in different frequency bands and are often required to interface with telecommunications or other facilities on the ground, on the seas, and in the skies. Some satellites are equipped with Faraday Cages to protect against Electromagnetic Pulse (EMP) and associated nuclear explosive effects. The commercial alternative is to provide for more spares (i.e., both in-orbit and on the ground) as well as to rely on insurance coverage as a way to protect against various types of risk factors. Consequently the commercial systems without the protection and special redundancy are generally less costly.

To date there have not been specific attacks on commercial application satellites consistent with the nonmilitary uses of space as recorded in the Outer Space Treaty of 1967 and the five resolutions of the United Nations Committee on the Peaceful Uses of Outer Space (COPUOS) (Jakhu et al. 2009). Nevertheless, there continue to be concerns about the various tests of antisatellite weaponry that have been carried out by the United States, Russia, and China to verify their ability to destroy in-orbit assets. Recent tests by China and the United States have increased concerns not only about the military use of antisatellite weapons, but also concerns that the use of such techniques can greatly expand the problem of orbital debris and the sustainability of space. These issues are being addressed in the context of United Nations discussions regarding the demilitarization of space in Geneva, Switzerland.

Conclusion

Reliability is absolutely key in the operation of applications satellites because they are providing a vital service. Telecommunications satellites and space navigation satellites do not tolerate even brief outages due to the types of real-time services that they provide. Remote sensing, Earth Observation and meteorological satellites can tolerate somewhat more extended outages but even these satellites need to be as absolutely reliable as possible since there are today no available means to do carry out in-orbit repairs, refueling or installing of new batteries or solar cell systems. Although such in-orbit serving has been discussed as a concept, such capabilities are years away. The harsh environment of space

includes hazards like radiation, sharp thermal gradients, a nearly complete vacuum, and shifting orbital mechanics. These conditions plus the remoteness of the satellite from Earth demands that all application satellites be designed to the highest levels of reliability, be tested under extreme conditions, and provided with critical backup components where and when possible. It also requires a tracking, telemetry, command and monitoring system that allows operators to know exactly where the satellite is at all times, how it is oriented, how the satellite components are functioning, and to be able to command the satellite subsystems. Remote commands can activate backup subsystems, fire jets to reorient the satellite, recharge batteries, reorient solar arrays, and carry out a wide range of activities to recondition the satellite or rescue it from a variety of hazards. Satellites can be powered down during solar flares or coronal mass ejections. Backup transponders or ASIC components can be activated or a fault switch located. Once an application satellite is deployed and tested out it has a remarkably good chance of operating for its full lifetime. A part of its capabilities such as a solar array, a battery, or a transponder or one of its sensors might be reduced in performance but the odds are, based on 50 years of experience, that the satellite will have over a 90% chance of achieving its projected lifetime. Many satellites continue to function well past their mean time to failure dates.

For the future, autonomous operation, in-orbit servicing, and improved component and subsystem design will likely assist in extending lifetime, improving reliability, and otherwise enhancing the cost-effectiveness of service.

Cross-References

- ▶ Common Elements Versus Unique Requirements in Various Types of Satellite Application Systems
- ▶ Overview of the Spacecraft Bus
- ▶ Telemetry, Tracking, and Command (TT&C)

Notes

1. D. Kegel: About ISDN, alumni.caltech.edu/~dank/isdn/isdn_ai.html
2. United Nations Organization for Outer Space Affairs: Report on current and planned global and regional navigation satellite systems and satellite–based augmentation systems, http://www.oosa.unvienna.org/oosa/SAP/gnss/icg.html

References

R. Jakhu, J. Logsdon, J. Pelton, Space policy, law and security, in *The Farthest Shore: A 21st Century Guide To Space*, ed. by J.N. Pelton, A.P. Bukley (Apogee Press, Burlington, 2009), pp. 306–314

M. Williamson, *Spacecraft Technology*: The Early Years (IEEE, London, 2006), p. 312

Ground Systems for Satellite Application Systems for Navigation, Remote Sensing, and Meteorology

43

Scott Madry, Joseph N. Pelton, and Sergio Camacho-Lara

Contents

Introduction 1097
TT&C Ground Systems 1097
Ground Systems for Satellite Navigation 1099
Ground Systems for Meteorological and Remote Sensing Satellites 1105
Future Trends 1107
Cross-References 1109
Notes 1110
References 1110

Abstract

The technology, the applications, and the economic forces that have driven the design, functionality, and performance of ground systems for satellite communications have been very closely mirrored in the other major application satellite services. It is for this reason that this chapter combines consideration of the ground systems for satellite navigation, remote sensing, and meteorology. In essence, all the ground systems for the various applications are communications

S. Madry (✉)
International Space University, Chapel Hill, North Carolina, USA
e-mail: Scottmadry@mindspring.com

J.N. Pelton
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: peltonjoe@gmail.com

S. Camacho-Lara
Centro Regional de Enseñanza de Ciencia y Tecnología del Espacio para América Latina y el Caribe (CRECTEALC), Luís Enrique Erro 1, Santa María Tonantzintla, Puebla, México
e-mail: sergio.camacho@inaoep.mx

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_11, © Springer Science+Business Media New York 2013

systems. Although the radio frequencies, modulation, and multiplexing methods and encryption schemes utilized vary for a variety of reasons – including defense and military-related consideration, all application satellites employ satellite communications between the spacecraft and the ground system. Some systems are broader or narrower in bandwidth and some only involve down links while other are more interactive with up and down links.

The common elements that range across the ground systems for all application satellites include the following:

- All application satellites have become higher in power, more accurate in their stabilization and pointing of their onboard antennas and better able to deploy higher gain and larger aperture reflectors. This has allowed ground systems to be smaller, more compact, lower in power, lower in cost, and more widely distributed.
- Down-linked information is often encrypted to protect the integrity of information and data relayed from the satellite – particularly if there is a proprietary or defense-related application for the down-linked information.
- Solid-state digital technology associated with integrated circuitry, application-specific integrated circuits (ASICs), and monolithic devices that have allowed the ground systems to be more highly distributed.
- There are essentially two tracks in ground systems development – one where geosynchronous satellites are involved and the ground system can be constantly pointed toward a single fixed point in the sky and the other where the ground system must have the ability to receive signals across the horizon and capture signals from a satellite that moves across the sky. Both types of ground receivers suited to "fixed" or "non-fixed" signal reception are needed in satellite communications, remote sensing, meteorological satellites, and satellite navigation.
- In addition to the user terms associated with different types of applications, there is a need for a tracking, telemetry, and command system to ensure the safe operation of the application satellite.

Despite these elements of commonality, there are indeed differences in the ground systems, the antenna characteristics, their tracking capabilities, the frequency utilized, the degree to which the data is protected by encryption, and the need for expert analysis of the data received from the spacecraft.

Keywords

Application-specific integrated circuit • Autonomous control • BeiDou space navigation system • Compass space navigation system • Encryption • Galileo space navigation system • Glonass space navigation system • GPS receiver • Indian regional space navigation system • Monitoring • Monolithic devices • Navstar • Omni-antennas • Quasi-Zenith space navigation system • Spacecraft performance monitoring • Squinted beam antenna • Telemetry and command • Tracking

Introduction

The first type of commercial satellite applications was for satellite communications. The types of ground systems evolved as the satellite communications industry evolved to cover fixed satellite services, broadcast satellite services, mobile satellite services, and store and forward data relay services. The use of satellites for remote sensing and meteorological services followed fairly closely in time, but these operations remained largely governmentally operated because there was not an established commercial market for these services. Only in the last few years have commercial remote sensing satellites evolved but they remain heavily dependent on governmental and defense-related clients. The last of the satellite applications to evolve is that of satellite navigation.

In recent years, the number of types of application satellites has continued to multiply rapidly. In all types of satellite applications, the evolution of ground systems for users has been remarkably parallel. Satellite systems have become larger and more powerful and thus space sensors and payloads have become more capable. The bottom line is that ground systems for satellite application users have become smaller, lower in cost, more accessible, and allowed the range of applications and services provided by these ground systems to diversify and thrive (Pelton 2012). The need for expert analysis of remote sensing data and meteorological satellite data has lessened the trend in these areas, but an ever-increasing number of new digital applications is fueling this trend even in these areas. Despite this overall trend toward moving satellite applications to the "edge" with smaller and lower cost user terminals in play, each type of application still has its own types of user devices. The one thing that all application satellites truly have in common on the ground is the Tracking, Telemetry, and Command (TT&C) systems.

TT&C Ground Systems

The most vital part of any application satellite network is its Tracking, Telemetry and Command (TT&C) system. This is simply because without a functioning TT&C system the satellite is lost in space and essentially not able to function. One can think of the TT&C system as both the brains and guidance control capability of the satellite. As noted in ▶ Chap. 41, "Telemetry, Tracking, and Command (TT&C)" "the three major tasks that the TT&C subsystem performs to ensure the successful operation of an application satellite are: (1) the monitoring of the health and status of the satellite through the collection, processing, and transmission of data from the various spacecraft subsystems, (2) the determination of the satellite's exact location through the reception, processing, and transmitting of ranging signals, and (3) the proper control of the satellite through the reception, processing, and implementation of commands transmitted from the ground." These functions remain constant regardless of what type of application the satellite might be involved.

There is another function, however, that is very much dependent on the type of application. This is what is called the "spacecraft performance monitoring function" and this capability allows ground controllers to operate the satellite correctly, but the electronic monitoring can also allow operators to determine if the satellite is actually delivering the service for which the satellite was intended in a proper fashion or if there is radio frequency interference that is impairing the service. Thus, for communications satellite, monitoring would determine if the quality as measured in bit error rate was appropriate and if there was undue interference occurring. In the case of remote sensing and meteorological satellites, the performance of the onboard sensors would be measured and calibrated. In the case of satellite navigation, the transmit signal would be measured and calibrated and interference detected.

The TT&C design for various types of application satellites is governed by only a few factors. These are:

- *Global, Regional, or Domestic Network*: If there is a global network then TT&C facilities need to be available around the world. If the network is polar orbit and sun synchronous then the TT&C facilities need to be located differently than if the network is geosynchronous. When application satellites were first deployed, each system operator had to arrange for or build their own TT&C network of expensive ground stations at appropriate locations around the globe. In time, however, as more and more systems were deployed, arrangements were made so that TT&C facilities could carry out these operations for multiple systems under an appropriate fee basis.
- *Assigned Frequencies for TT&C operations*: A variety of frequencies were assigned to carry out TT&C operations. Most of these were in the Very High Frequency (VHF) and Ultra High Frequency bands (UHF) since these did not require broadband or high data rate to carry out tracking, ranging, telemetry, or commands functions.
- *Encryption*: Encrypted signals in special frequency bands are used for military and defense-related satellites, but as application satellites have gotten larger, more powerful, and broadband, application satellites of all types have moved to more and more secure modes of operation. Today, TT&C operations for all types of satellites are typically carried out with encrypted signals and commands may not only be digitally encrypted but there may be special security techniques employed such as requiring all commands to be confirmed from a separate TT&C site.

The high cost of TT&C operations is a matter of concern for all satellite application systems. This issue of cost has been addressed by automating many aspects of the TT&C operation with alarms sounding when a certain set parameter has been exceeded. The further step is to move toward what is called "autonomous operations." This means that an artificially intelligent (AI) onboard computer assumes control of the satellite and responds to routine technical issues based on expert systems designed to adjust satellite settings or employ backup components and subsystems. With autonomous control, a satellite's operation can become

largely "autonomous" and ground control is reassumed only in the case of major difficulties.

Ground Systems for Satellite Navigation

In a rapidly expanding field of applications of the signals provided by the GNSS, it is essential that the user community and receiver-producing industry have a clear and consistent description of the global and regional systems that are currently operating and that which will operate in the future. To this end, the United Nations Office for Outer Space Affairs, in its role as Executive Secretariat for the International Committee on GNSS (ICG), prepared a publication on the planned or existing systems and on the policies and procedures that govern the service they provide (United Nations Office for Outer Space Affairs 2010). To reflect future changes the publication will be updated and will be available on the website of the ICG. The following paragraphs describe the ground segments for the existing and planned GNSS and their augmentations.

Because of the large number of satellites that comprise the global and regional navigation satellite systems and their augmentations (see ▶ Chap. 25, "Current and Future GNSS and Their Augmentation Systems"), their ground segments involve a larger number of antennas and monitoring stations. Figure 43.1 shows a typical GNSS antenna.

The Global Positioning System (GPS) of the US operational control segment consists of four major subsystems: a master control station, an alternate master control station, a network of four ground antennas, and a network of globally distributed monitor stations. The master control station is located at Schriever Air Force Base, in Colorado, USA, and is the central control node for the GPS constellation. The master control station is responsible for all aspects of constellation command and control, including routine satellite bus and payload status monitoring, satellite maintenance and anomaly resolution, management of signal-in-space performance, navigation message data upload operations, and detecting and responding to GPS signal-in-space failures.

In September 2007, the GPS operational control segment was modernized by turning to a distributed system resulting in increased capacity for monitoring GPS signals, from 96.4% to 100% worldwide coverage with double coverage over 99.8% of the world. All the current GPS Interface Control Documents can be downloaded from GPS.gov, the official US Government webpage for GPS.

The ground segment of the Wide-Area Augmentation System (WAAS) is operated by the Federal Aviation Administration (FAA) of the USA There are 38 wide-area reference stations throughout North America (in Canada, Mexico, the USA, and Puerto Rico). The FAA plans to upgrade the wide-area reference stations with receivers capable of processing the new GPS L5 signal.

The Local-Area Augmentation System is a ground-based augmentation system that was developed to provide precision-approach capability for categories I, II, and



Fig. 43.1 Unified state ground control network GLONASS receiving station (Courtesy of Russian Space Agency Roscosmos)

III approach procedures. It is designed to provide multiple runway coverage at an airport for three-dimensional required navigation performance procedures and navigation for parallel runways with little space between them and "super-density" operations.

The Nationwide Differential Global Positioning System is a national positioning, navigation and timing utility operated and managed by the US Coast Guard. It consists of 50 maritime sites, 29 inland sites, and 9 waterway sites. The System provides terrestrial services to 92% of the continental USA with 65% receiving dual coverage. The System is used in surface and maritime transportation, agriculture; environmental and natural resource management; weather forecasting; and precise positioning applications.

The national network of continuously operating reference stations, coordinated by the National Geodetic Survey and tied to the National Spatial Reference System, consists of more than 1,300 sites operated by over 200 public and private entities, including academic institutions. Each site provides GPS carrier phase and code range measurements in support of three-dimensional centimeter-level positioning activities throughout the USA and its territories.

The Global Navigation Satellite System (GLONASS) of the Russian Federation ground segment consists of a system control center; a network of five TT&C

centers; the central clock; three upload stations; two satellite laser ranging stations; and a network of four monitoring and measuring stations, distributed over the territory of the Russian Federation. Six additional monitoring and measurement stations are to start operating on the territory of the Russian Federation and the Commonwealth of Independent States in the near future.

The Unified State Ground Control Network[1] (USGCN) is designed to control automated spacecrafts, manned spacecrafts, and space stations. USGCN solves movement control problems for spacecrafts of different purposes during all flight and descent phases, control of operation of all their equipment and systems; scientific, meteorological, communication, television, navigation, and topogeodesic data reception; manned spacecrafts crew radio communication, and carrier vehicle launch measurements. USGCN is a combination of hardware components and facilities located at the Head test center of spacecraft testing and control (HTCTC SF) of Ministry of Defense (MoD) and at ground-based detached command and measurements complexes (DCMC). These facilities are connected by data and control communication channels into the unified automated control complex. Facilities inside USGCN designed to control single spacecrafts or constellations of similar spacecrafts form ground control complexes (GCC), which together with onboard control complexes comprise an automated control network. The document that defines requirements related to the interface between the space segment and the navigation user segment is the Interface Control Document (version 5.1, 2008).[2]

The Galileo ground segment controls the Galileo satellite constellation of the European Union, monitoring the health status of the satellites, providing core functions of the navigation mission (satellite orbit determination, clock synchronization), determining the navigation messages and providing integrity information (warning alerts within time-to-alarm requirements) at the global level, and uploading those navigation data for subsequent broadcast to users. The key elements of those data, clock synchronization and orbit ephemeris, will be calculated from measurements made by a worldwide network of reference sensor stations. The current design of the system includes 30–40 sensor stations, 5 tracking and command centers, and 9 mission uplink stations. The present Galileo Open Service Signal-in-Space Interface Control Document (OS SIS ICD) Issue 1[3] contains the publicly available information on the Galileo Signal In Space. It is intended for use by the Galileo user community and it specifies the interface between the Galileo Space Segment and the Galileo User Segment. As the Galileo constellation is placed In orbit, the interface document is subject to evolution, and the information contained in it may change.

The EGNOS ground segment is mainly composed of a network of ranging integrity monitoring stations, four mission control centers, six navigation land Earth stations, and the EGNOS wide-area network, which provides the communication network for all the components of the ground segment. Two additional facilities, the performance assessment and system checkout facility and the application-specific qualification facility, are also deployed as part of the ground segment to support system operations and service provision.

The ground segment of the Compass/BeiDou Navigation Satellite System of China consists of one master control station, upload stations, and monitor stations. COMPASS/BeiDou user terminals are intended to be "compatible" with GPS, GLONASS, and Galileo receivers.

The ground segment of Japan's Quas-Zenith Satellite System (QZSS) performs the tracking, computation, updating and monitoring functions needed to control all of the satellites in the system on a daily basis. It consists of a master control station in Japan, where all data processing is performed, and some widely deployed monitor stations in the area that are visible from the space segment. The monitoring stations passively track all satellites in view and measure ranging and Doppler data. These data are processed at the master control station so that the satellite's ephemerides, clock offsets, clock drifts, and propagation delay can be calculated, and are then used to generate upload messages. This updated information is transmitted to the satellites via TT&C and to the navigation message uplink station at Okinawa for subsequent transmission by the satellites as part of the navigation messages to the users. The Interface Specification (IS-QZSS) document[4] defines the interface between the Space Segment (SS) provided by the Quasi-Zenith Satellites and the User Segment of the QZSS.

The ground segment of the augmentation system MSAS consists of two master control stations (one at Kobe and one at Hitachioota), two monitoring and ranging stations (one in Australia and one in Hawaii), and four ground-monitoring stations (at Sapporo, Tokyo, Fukuoka, and Naha). The master control stations generate augmentation information based on the GPS and MTSAT signals received at the ground-monitoring stations and the monitoring and ranging stations. The ground-monitoring stations monitor GPS satellite signals and transfer the information to the monitoring and ranging stations.

The ground segment of the Indian Regional Navigation Satellite System (IRNSS) is responsible for the maintenance and operation of the constellation. This segment comprises nine IRNSS TT&C stations, two spacecraft control centers, two IRNSS navigation centers, 17 IRNSS range and integrity monitoring stations, two IRNSS timing centers, six CDMA ranging stations, and two data communication links. As part of the ground segment, 15 Indian reference stations for monitoring and collecting the data, two master control centers and three uplink stations are planned for the GPS-aided GEO-Augmented Navigation (GAGAN) System.

The range and diversity of satellite navigation units that are available today is quite large. Among the units dedicated to use with the US Navstar system, GPS alone is staggering huge. There are units optimized for truckers, for boaters, for hikers, and even golfers. There is a unit that is loaded with the layout of thousands of golf courses. These are low in cost and these consumer-oriented units can cost in the range of US$100 (in used condition) to up to US$1,000. The price variation depends on the software for visual display, storage capability, touch screen capability, etc (see Fig. 43.2).

There are then much higher end GPS units that are designed for aviation applications and can provide 3-axis orientation and real-time display for pilots that shows the location with respect to the Earth (see Fig. 43.3). There are also units



Fig. 43.2 Garmin© Montana GPS receiver for hikers (Graphic courtesy of Garmin©)

that are designed to operate in the frequencies for the GPS network as well as the GLONASS system. These are larger and more expensive (see Fig. 43.4). Finally, there are units that can access not only GPS and GLONASS but are designed to flexibly access other satellite navigational systems as they are deployed such as the European Galileo, the Chinese Beidou/Compass system, the Japanese Quasi-Zenith system, and the Indian Regional Satellite Navigational system (see Fig. 43.5). These multiuse units that can utilize signals from all the current and planned satellite systems cost in the range of about US$10,000–US$12,000. As the operators of the GNSS move toward interoperability (see ▶ Chap. 24, "▶ International Committee on GNSS") the receiver systems should become less expensive.

The above are only some of the ground receivers for space navigation systems now broadly available. There are even more sophisticated GPS receivers that are designed for space experiments and for activities or experiments where greater 3-axis spatial accuracy is required. In such cases, for instance, a group of four GPS receivers might be configured in close proximity to establish greater range accuracy. Such a configuration might be used for aeronautical or space applications.[5] At the other end of the spectrum there are quite low cost ARGOS omni receivers available that are used in such applications where there is not a requirement for



Fig. 43.3 The Helm X650 "True Map" GPS that provides a 3-axes display for aircraft applications (Graphic courtesy of Helm©)



Fig. 43.4 The LaiPac Tech dual mode receiver for both GLONASS and GPS signals (Graphic courtesy of LaiPai Tech)

quite high spatial or geographic accuracy. These applications using Argos receivers might include activities such as locating ocean buoys associated with ocean-based experiments.[6] The number and types of space navigation receivers will doubtlessly increase as more satellite navigation systems such as Galileo, Compass, the Quasi-Zenith, and the Indian Regional Space Navigation Satellite Systems continue to be deployed in coming years.



Fig. 43.5 Tokay professional satellite navigation receiver is equipped for all Sat Nav frequency bands (Graphic courtesy of Tokay)

The ground units for space navigation must be able to receive signals from orbiting satellites that are typically in quite high medium earth orbit, or in the case of the Quasi-Zenith network (a Geo orbit that is tilted 45° to the equator). All of the ground receivers now in operation are essentially very low gain omni or squinted-beam omni devices. These terminals, since they do not have an active tracking capability are designed to capture transmitted signals from medium earth orbit satellites anywhere above the horizon. The key to these devices, from a technical performance perspective, are highly specialized application-specific integrated circuitry (ASIC) that allows digital processing algorithms to augment the ground unit's ability to receive a very low level signal from the navigation satellite that is orbiting many thousands of kilometers above.

Ground Systems for Meteorological and Remote Sensing Satellites

The ground systems for meteorological and remote sensing satellites are, for the most part, quite different from satellite navigation satellites. This is because most of these ground systems are designed to receive signals from meteorological or remote sensing satellites at special facilities designed for data analysis by trained specialists. Essentially, meteorological satellites are just a specialized form of remote sensing satellite, and only geosynchronous meteorological satellites are a separate case. In some cases data from a remote sensing satellite can be relayed via a data relay satellite to provide the data for analysis on an accelerated basis. This could be in the case of a military or defense-related application or in the case of a hurricane, monsoon, or typhoon in order to mitigate a disaster or provide accurate warning notices.

In essence, virtually all ground stations supporting remote sensing or meteorological activities are not intended for mass consumer use although there are lower cost "hobbyist ground systems" that can be acquired at rather reasonable cost that will be discussed below. The number of professional or ground receivers located at universities for training purposes are much fewer than is the case with space

navigation receivers. In short, while there are tens of millions of GPS, GLONASS – and soon other types of space navigation terminals – there are only thousands of professional high-gain meteorological or remote sensing receiving ground stations with rapid tracking capabilities. For polar orbiting meteorological and remote sensing satellite operations, these will typically have rapid tracking capabilities since the relatively low orbits involve passes over the ground station and data processing center in a matter of just a few minutes.

All types of these ground antennas for meteorological and remote sensing operations (including Geo and polar orbiting systems) will have increased gain and thus relatively antenna apertures. As remote sensing and meteorological satellites have been equipped with more and more types of sensors – spectral, hyperspectral, infrared, radar, monitors of lightning strikes, etc. – the ground systems have been upgraded to receive more data more efficiently via enhanced digital transmission capabilities. During a single satellite pass, many gigabytes of data might be gathered at a single processing center. The initial ground systems have huge antenna systems since the transmission capabilities of early satellites were limited. The early Tiros ground systems represent some of the largest ground systems constructed during the 1960s (see Fig. 43.6).

The design of some remote sensing systems such the so-called "Mission to Planet Earth" system was scaled down in terms of the amount of data collected via various sensors not because of the ability of ground stations to collect the data, but due to the ability to accurately process the many terabytes of data that could be collected by a network of scores of remote sensing and meteorological satellites. Today, the ground systems for collecting data from remote sensing systems such as Spot Image, GeoEye, and other remote sensing satellite networks are still quite capable of rapid tracking across the sky but nevertheless much smaller than the giant Tiros stations shown below (see Fig. 43.7).

The first step in processing data at a ground receiving complex is "unpack" the incoming data and store it for processing. The data processing of remote sensing data is often divided into "preliminary processing" that is then followed by "thematic processing." The preliminary processing involves "unpacking." This involves converting the raw data, received by the ground station, into products suitable for storage and further thematic processing. This "preliminary processing" can include a number of steps. These can include radiometrical calibration, geolocation, and geometric correction of images.[7]

Later thematic processing involved the detailed interpretation that is carried out to accomplish specific tasks associated with agriculture, forestry, fishing, mining, urban planning, or even disaster recovery. There is today a wide range of ground stations capable of reception and preliminary processing of data and these vary in their characteristics if they are receiving from Geo- or polar-orbiting satellites in terms of their tracking capability. Also, there are some differences as to the receiving data rates in terms of the governmental, military, or commercial applications and especially with regard to the resolution of the sensing. Obviously, the higher the resolution, the more data that is captured to be transmitted to the ground.

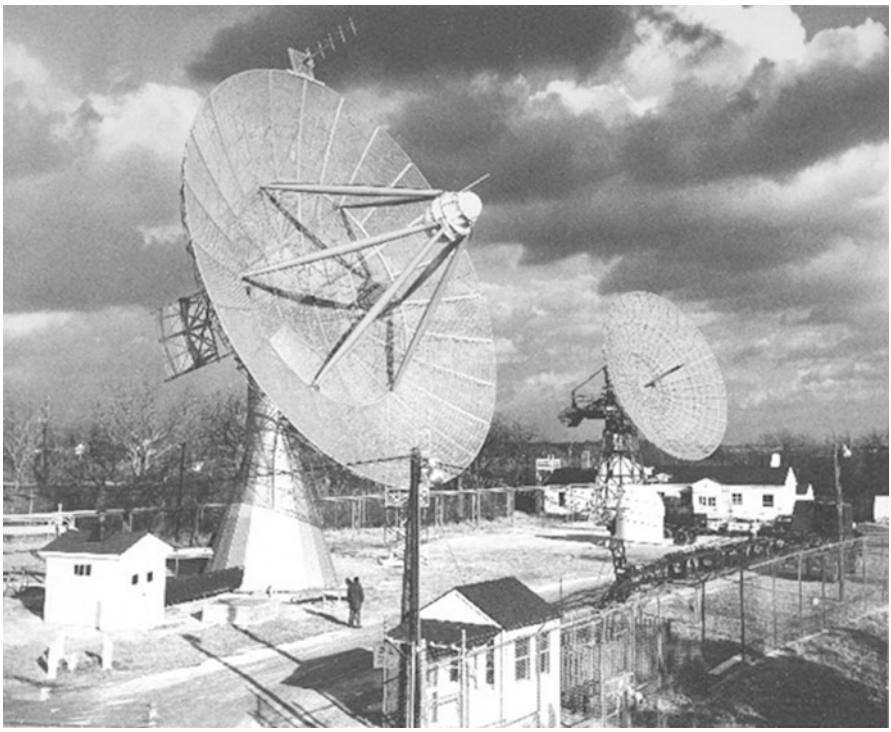


Fig. 43.6 The Tiros ground stations from the 1960s (Photo by Frank Vosk)

There are much simpler, non-tracking receiver antennas that can capture far less data, but still could be of interest to the hobbyist who is interested in space-based meteorology or remote sensing. The following diagram outlines a schematic for such a low-resolution ground receiver that could be purchased or constructed from components by a hobbyist for a few hundred dollars (US)[8] (see Fig. 43.8).

Future Trends

The future trends for ground systems for all application satellites will tend to follow the pattern seen over the past 50 years. This trend will be for smaller, less expensive, and more widely available and user-friendly ground systems. The fields of satellite communications and satellite navigation have seen this development quite successfully accomplished since software and ASIC chips have allowed the consumer to utilize the transmission directly from the satellite with increasing ease. As the volume of receiver terminals in these areas have risen to many millions, the cost of these devices have continued to drop so that consumer TVROs for direct



Fig. 43.7 A spot image remote sensing receiving station (Spot Image Receiving Station http://www.astrium-geo.com/files/pmedia/public/r2184_9_spot_receiving_station_antenna.jpg) (Graphic courtesy of Astrium)

broadcast television and GPS receivers are now in the hundreds of dollars (US) range.

Today, meteorological and remote sensing satellites still require a significant amount of professional formatting and analysis (preliminary processing and thematic processing) that creates the need for larger receiving stations with relatively complex operation and maintenance requirements. Pressures to develop software so that remote sensing and meteorological data can be brought closer to the "edge" – particularly to support warfighters in field – has continued the trend toward decentralization and the use of expert systems in mobile computers and even personal data assistants for weather and remote sensing data assessment. The complexity of thematic analysis will prevent the complete decentralization seen in satellite navigation and satellite communications, but the future trends toward smaller, simpler, and lower cost receivers seem to be universal throughout the industry.

Conclusion

The ground systems for all types of application satellite systems will be increasingly smaller, simpler, and highly distributed with communications satellite and space navigation units continuing to lead the way. There is speculation that such devices might someday evolve into wearable antennas that might take the shape of wristwatches or be embedded in a shirt or jacket. The evolution of meteorological and remote sensing ground systems will take a slower and more diverse

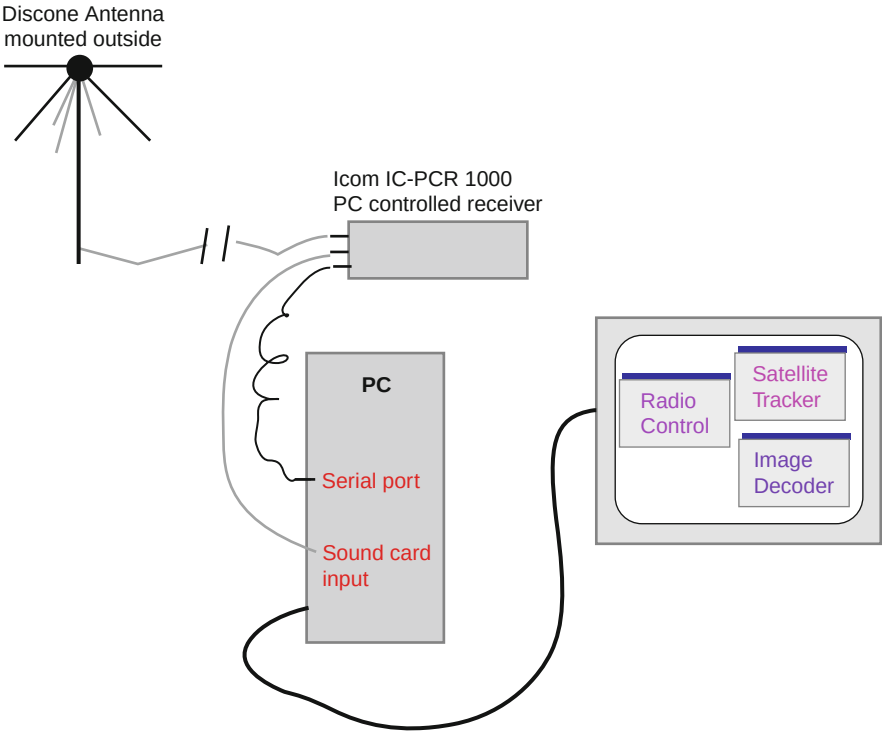


Fig. 43.8 Diagram for a hobbyist version of a simple ground station to receive remote sensing or meteorological data (Graphic Courtesy of Hobbyspace)

route. This could be a full spectrum of ground systems with complex antennas continuing to be linked to analysis centers for more complex operations related to map making, resource prospecting, and high-value remote sensing applications, but other simpler applications being much more distributed through the use of artificially intelligent or expert systems software allows data going directly from satellites to hand held or lower cost mobile units.

All of the applications will continue to need tracking, telemetry, and command plus performance monitoring to operate the application satellites in the sky. Indeed, to allow the ground units to shrink in size and cost, the satellites will need to be higher in power and increasingly capable. This means that the satellites and the TT&C facilities will need to be more capable to compensate for the simpler units on the ground or onboard vehicles (including aircraft).

Cross-References

► Current and Future GNSS and Their Augmentation Systems
► International Committee on GNSS

► Introduction to Satellite Navigation Systems
► Satellite Antenna Systems Design and Implementation Around the World
► Telemetry, Tracking, and Command (TT&C)

Notes

1. http://www.spacecorp.ru
2. http://www.spacecorp.ru/en/directions/glonass/control_document/index.php?sphrase_id=3633
3. http://ec.europa.eu/enterprise/policies/satnav/galileo/files/galileo-os-sis-icd-issue1-revision1_en.pdf
4. http://qz-vision.jaxa.jp/USE/is-qzss/DOCS/IS-QZSS_14D_E.pdf
5. G. Lachapelle, H. Sun, M.E. Cannon, G. Lu, Precise aircraft-to-aircraft positioning using a multiple receiver configuration. http://webone.novatel.ca/assets/Documents/Papers/File19.pdf
6. Argos Receiving Antenna. http://www.telonics.com/products/argosReceivers/
7. V.I. Gershenzon, A.A. Kucheiko, *Remote Sensing Data*. http://www.scanex.ru/en/publications/pdf/publication4.pdf
8. Building a Weather Satellite Station. http://www.hobbyspace.com/Radio/WeatherSatStation/

References

J.N. Pelton, *Satellite Communications* (Springer, New York, 2012)
United Nations Office for Outer Space Affairs, *Current and planned global and regional navigation satellite systems and satellite-based augmentation systems*, ST/SPACE/50 (United Nations, New York, 2010). http//:www.icgsecretariat.org

Common Elements versus Unique Requirements in Various Types of Satellite Application Systems

44

Joseph N. Pelton and Scott Madry

Contents

Introduction 1112
Common Technical Elements in Application Satellite Programs 1113
Spacecraft Structures and Bus Platforms 1114
Power Systems 1116
Tracking, Telemetry, and Command 1118
Ground and User Systems 1119
Launch Services 1120
Common Operational and Regulatory Aspects of Application Satellite Programs 1121
Common Market and Business Considerations in Application Satellite Programs 1123
Dissimilar and Unique Requirements in Application Satellites 1124
Differences That Stem from the Differences in the Various Satellite Markets 1124
Design and Engineering of Satellite Payloads: Communications Antenna Systems, Multispectral Sensors, and Radar Systems 1126
Cross-References 1127
Notes 1127
References 1128

Abstract

The concept of developing a Handbook on Satellite Applications is based on the concept that all of the commercial and practical applications of space have many elements in common. In fact very similar power systems, spacecraft platforms, stabilization and positioning systems, and tracking, telemetry, and command systems are used for the various types of application satellites. It is the payloads

J.N. Pelton (✉)
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: joepelton@verizon.net

S. Madry
International Space University, Chapel Hill, North Carolina, USA
e-mail: Scottmadry@mindspring.com

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_71, © Springer Science+Business Media New York 2013

that tend to be quite specialized. Even in the field of telecommunication satellites quite different antenna systems and communications subsystems are now developed and deployed for various satellite systems for satellite broadcasting, fixed or mobile services. It is equally true that different types of remote sensing, space navigation, and meteorological satellites can and do have different payload designs. The purpose of this chapter is to contrast and compare different types of application satellites to note major areas of similarities as well as how and why differences occur. Such an analysis is useful to understand where the most promising common elements lie in order to aid identifying new potential synergies for future research and development in order to seek out improved methods for common forms of reliability testing, sparing and redundancy strategies as well as lifetime extension and reduced operating and monitoring costs.

Keywords

3-axis stabilization • AC/DC current • Antenna systems • Ariane launch vehicle • Carbon/epoxy composites • De-spun platform • Earth and sun sensors • Energy density • Heat pipe • Intellectual property • Launch services • Lifetime testing • Lithium ion batteries • Payload design • Power systems • Solar cell • Space agencies • Spacecraft bus • Spacecraft structures • Thermal control • Thrusters • Tracking, telemetry, and command (TT&C) • Vernier jets • World Trade Organization (WTO)

Introduction

There are clear parallels between the various types and classes of application satellites. The various sections of this book are organized to indicate first the various ways that telecommunication satellites, remote sensing satellites, space navigation, and meteorological satellites are different in terms of payloads and antenna systems as well as market trends and structure. This is followed by noting the various elements that tend to be common in terms of spacecraft structures, power systems, orientation and positioning systems, and launch arrangements.

This chapter has two main objections. The first of these objectives is to seek to provide an analysis of the various technical, operational, market, and business aspects of the overall satellite applications field in a manner to show those elements that are quite parallel and common and which would benefit from future technical R&D and management strategies that allow mutual benefit across the entire field. Examples of this would be things like improved and longer life batteries or fuel cells, improved inertial or momentum wheels, lower cost and more precise atomic clocks, improved tracking, telemetry and control systems, improved storage and buffering systems, better orientation and positioning systems, improved data relay systems, low cost and more reliable launch systems, etc.

The second of these objectives is to clearly identify areas within the overall field of satellite applications where individual and unique requirements exist and

separate approaches are needed to develop appropriate and improved new technology, operating capabilities, and/or business and management techniques. Examples hear including extremely large aperture antennas for mobile satellite services, techniques to overcome rain attenuation in Ka-band systems and above, synthetic aperture radar imaging, and indeed most developments related to specific payloads or antenna systems.

There is a temptation to assume that this sort of dichotomy with regard to common elements and unique elements can quickly be identified and sorted out from one another. Unfortunately this is not always the case. If one takes the example of onboard autonomous control, there may well be common elements that could reduce the cost of operating the various types of application satellites. It is also true that a technique developed within an artificial intelligence or expert system software might be very well applied to satellite telecommunications, for instance, but might also produce an unintended or harmful result in a space navigation system or remote sensing satellite.

The main point to emphasize is that even if there is a presumed benefit from a new capability developed for one type of application satellite, there must be very careful consideration as to how or why it might be applied to another type of system. This caution even applies to a particular subfield of satellite applications. A technique that is appropriate for broadcasting (or a one-way service where transmission delay is not a key issue) may not work well for interactive systems that provide mobile satellite communications, fixed satellite services, or defense-related applications. This is, in short, a much more difficult process than one might at first think. It is much harder than the children's game of "which of these things is not like the other" that is played on education television. Nevertheless, the following analysis seeks to create an analytic framework for identifying the most promising areas for future development where application satellites might have quite common goals for improved performance, reliability, or cost reduction.

Common Technical Elements in Application Satellite Programs

There are many aerospace manufacturers around the world and their expertise and research capabilities are central to the supply of quality application satellites – past, present, and future. A number of space agencies around the world add some research funds and expertise in the application satellite field, although support for application satellite research has waned as industrial and commercial capabilities in these areas have strengthened over the years. Today there is still active and meaningful support for research in the application satellite area. Notable R&D programs are pursued by the Brazilian Space Agency (INPE), the Canadian Space Agency (CSA), the Chinese National Space Agency (CNSA), the European Space Agency (ESA) (especially the TIA-ARTES research program), the French Space Agency (CNES), the Germany Space Agency (DLR), the Indian Space Research Organization (ISRO), the Italian Space Agency (ASI), the Japanese Space Agency (JAXA) (especially the Engineering Test Satellite (ETS) research

satellite series), the Russian Space Agency (Roscomos), and even the US Space Agency (NASA) (which was one time active with regard to telecommunications satellite research but now only carry out programs regard to remote sensing and meteorological satellites). These various governmentally -sponsored research programs – and even some defense agency research support – add quite useful supplemental research funding and technology development. Nevertheless it is the major aerospace industries that are critical to the future development of application satellite technology. Major firms such as Ball Aerospace, Boeing, Lockheed Martin, Northrop Grumman, Orbital Sciences, Space Systems/Loral (in the USA); Alcatel, BAE Systems, EADS/Astrium, Paradigm, Siemens, and Thales Alenia (in Europe); NEC, Toshiba, and Mitsubishi Electronics Company (MELCO) (in Japan); the Chinese Aerospace Science and Technology Corporation and the Great Wall Industry Corporation (in China); Comdev, MacDonald Dettwiler, and MDA Space Missions (in Canada); and Hyundai (in Korea) represent some of the most important suppliers of the world's application satellites.

The above listing of aerospace manufacturers is, of course, only intended as partial listing of the entirety of global application satellite system suppliers. This is nevertheless indicative of the largest suppliers. There are important emerging suppliers in Brazil, Israel, India, and Russia, as well as entities like Surrey Space Technologies in the United Kingdom and several US suppliers that specialize in micro-satellite design and manufacture. The appendices to this Handbook provide a more comprehensive listing of launch vehicle suppliers and satellite manufacturers.

The basic point is these commercial spacecraft manufacturers actually design and manufacture the overwhelming number of telecommunications, remote sensing, space navigation, and meteorological satellites. Any transfer of technology-related technical, operational, or business systems for application satellites must take into account the pool of satellite system suppliers and intellectual property protections and patents that would be central to this process (Ippolito and Pelton 2004).

Spacecraft Structures and Bus Platforms

A very high percentage of application satellites today are built on "bus platforms" equipped with high-speed, magnetically suspended momentum wheels that allows the entire system to be stabilized in all three axes and very accurately pointed. The "boxed shaped" bus is capable of supporting a wide range of antenna structures and deployable solar power arrays that can be oriented to achieve maximum sun exposure. In the very early days of satellite design and deployment satellites came in a wide range of shapes and sizes and many satellites from start to finish were either one of kind or a few of a kind. This rather chaotic design environment tended to drive up sharply the initial design and engineering costs, i.e., the nonrecurring cost component, for each new satellite or satellite series. In the age of the "despun satellite design" there was a move toward standardization of different classes of satellites, i.e., small, medium, and large configurations (Pelton 2006).

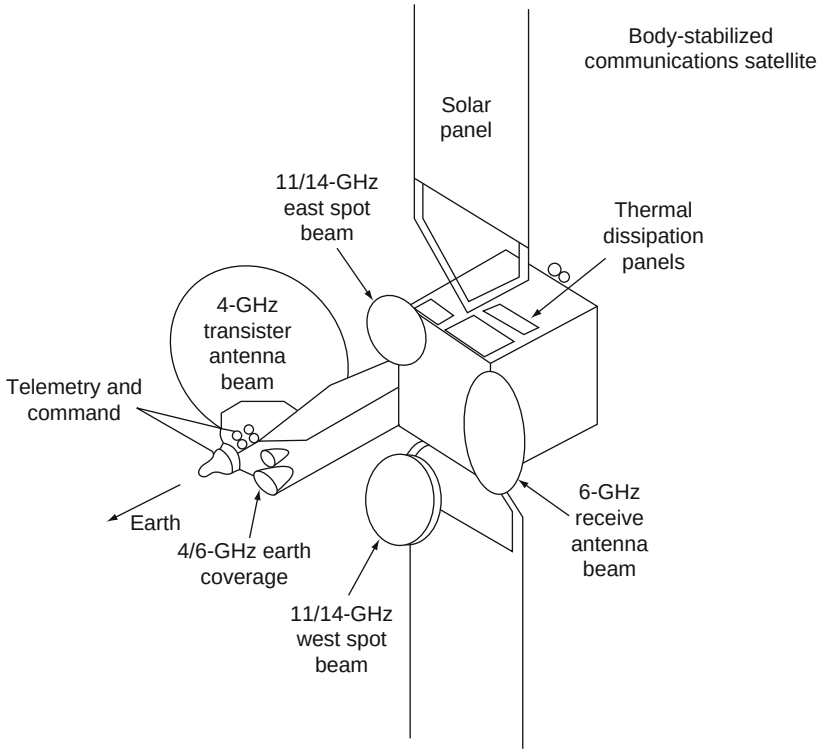


Fig. 44.1 The 3-axis body stabilized application satellite with a box-like platform is now the industry standard for virtually all commercial spacecraft (Graphics Courtesy of J. N. Pelton)

The advent of the three-axis body stabilized spacecraft that used high-speed momentum or inertia wheel to achieve much more precise pointing of space systems in Earth orbit revolutionized the space applications industry. This approach led to a typically standardized "box design" for the basis spacecraft unit from which antennas, solar power arrays, and imaging systems were deployed. Once this concept of how to design a satellite that could precisely point antennas or imaging devices and allow solar arrays to be deployed with maximum efficiency, the wisdom of developing specific classes of "platforms rather naturally evolved to even more sophisticated levels" (Fig. 44.1).

Just as automobile manufacturers tend to have just a few standardized chases (or platforms) on which to design cars and trucks, the satellite industry found scale economy and more rapid build and test efficiency to have only a few platforms to support telecommunications, remote sensing, space navigation, and meteorological satellites. In fact these same platforms might indeed be used as the basis of exploratory, scientific, or even defense-related satellites as well.

Other common elements in spacecraft bus design evolved as well. Most spacecraft buses and "masts" for antennas or sensors were built from ultrastrong but quite

lightweight carbon/expoxy composite and acrylic fiber structures. The electrical systems used within the spacecraft bus also tended to be similar "direct current (DC)" based designs, although highly specialized systems in rare instances also use "alternating current." In general AC to DC conversion tends to be minimized where possible. Passive and active thermal control systems from reflective exterior materials to heat pipes could be used over again with confidence since common design allowed for use of design that had been actually tested in space.

Tracking, telemetry, and command (TT&C) systems were often attached to the bus in a similar fashion. Sensors for detecting the spacecraft orientation were likewise common as well as the thrusters and Vernier jets that maintained the proper position accuracy to the Earth below (Pelton 2006).

All of these similarities in spacecraft bus design for application satellites provide powerful economies of scale. The creation of just a few different sized bus platforms not only allows for economies in terms of design, engineering, and production, but it also contributes to reliability in terms of being able to eliminate design flaws, reduce manufacturing errors, and improve testing techniques. The use of common batteries, momentum wheels, and thruster jets allows improved lifetime testing practices.

Power Systems

Satellite power systems, because they are crucial to performance, and because they vary greatly in terms of requirements from satellite to satellite – even from different telecommunication remote sensing, space navigation, or meteorological satellite deserve particular consideration. Power failures are a very key issue for application satellites. This is because both in space and on the ground, this is one of the most frequently experienced problems. While active components fail, they can often be backup by redundant units that allow rapid restoration of service. If a solar power array fails to deploy or if a battery system fails the mission is essentially a loss. Some telecommunication satellites (especially direct broadcast or mobile satellite systems requires quite high power level in the range of 10 kW or above) as to active remote sensing systems such as radar satellites that must generate power to be reflected back to the satellite. Other micro satellites for communications or remote sensing as designed by the Surrey Space Technology Center can have relatively low power requirements. Thus power systems for application satellites constitute a particular area of focused research. There are several key factors of commonality.

- *Solar cell performance*: Launch service costs are quite high. It may well cost as much to launch a satellite and insure against its failure than to manufacture it in the first place. Since launch costs are high one desires solar cells that are quite efficient in converting solar energy from photons into electrical energy. Thus low efficiency by low cost amorphous silicon cells are typically not used on application satellites. There is often a choice between lower cost structured silicon solar cells (around 15% or efficiency) and higher cost gallium arsenide solar cells (around 20–25% efficiency). Clearly if one could develop much

higher efficiency solar cells while keeping costs low this would be a boon to all forms of application satellites. (Indeed such developments could spin off to Earth-based solar energy generating systems.) This is why there is a good deal of research in new high valence materials for solar cells as well as multi-juncture solar cells (i.e., the violet solar cell and/or the "rainbow" solar cell). So-called quantum dot technology that could be up to three times more efficient than solar cells is subject to intensive R&D. Others believe that reflective surfaces or solar concentrators that allow a solar cell to "see" the equivalent of three or indeed many suns might provide an important pathway forward.

- *Battery and fuel cell systems*: One of the common problem experienced by application satellites is that the sun's energy is at one time or another blocked by the sun. The Earth blocking the sun from illuminating solar arrays is the most common problem. In order to provide power during eclipses or other short-term outages that may occur, one must have an alternative power supply and energy storage supply. There has been steady improvement in battery performance in terms of energy density and reliability. Lithium ion batteries are now the most common satellite energy storage and power supply system. Progress continues to develop unitized and regenerative fuel cells that could be used on satellites or even Earth-based applications. Some shorter-lived satellites deployed in LEO orbits might have much different requirements from longer-lived satellites in GEO orbits. Satellites in polar orbit that have maybe 35–40 min of eclipse out of a 90 min orbit will have different battery requirements from a GEO satellite that only experience only seasonal eclipses and even then for a maximum of no more than a hour out of a 24 h day.
- Despite these various differences, all types of application satellites could benefit from improved battery or fuel cell development and/or increased lifetime and reliability. Some engineers believe that in time batteries, fuel cell, or even nuclear power sources may evolve to such an improved state in terms of performance, cost, and reliability that one might even deploy a satellite without solar power arrays and rely exclusively on one of these onboard power sources that does not require the risk of deployment of a solar array and does not entail the considerable mass of having both solar and power storage systems onboard the satellite. Technology to support such a design concept is still in the future.
- *Protecting satellite power systems from failure*: There are ongoing research and development (R&D) programs directed toward enhancing the reliability of satellite power systems. Satellites that fly through the van Allen belt and especially medium earth orbit satellites typically have their solar cells coated with a silicon veneer to protect the lifetime effectives of these devices. There are efforts to find solar cells that are more resistant to radiation as well as to find lightweight coating systems to protect the solar arrays. There are also new designs to deploy solar arrays more reliably. Instead of "accordion-like" extension systems some of the newer designs have solar arrays that can be rolled out to full extension.[1]

For the reasons noted above there is ongoing research and development to extend the performance, reliability of satellite power systems as well as efforts to

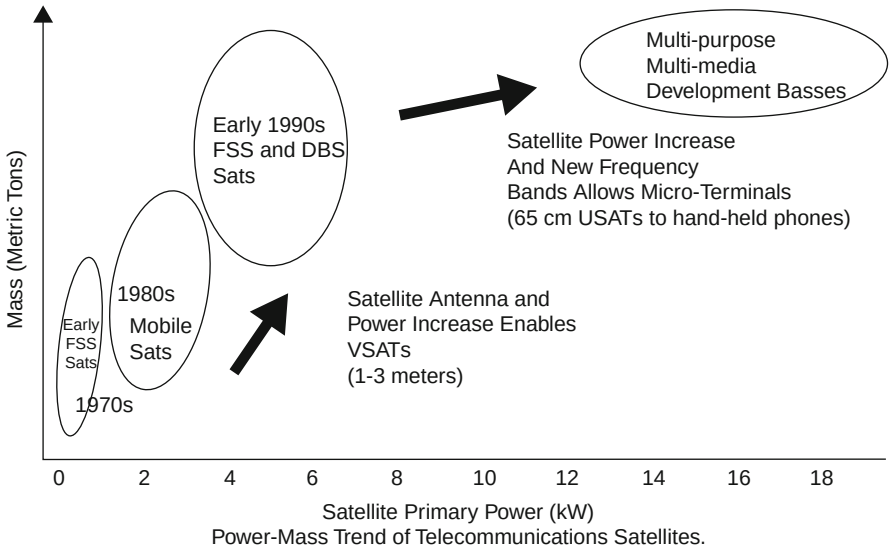


Fig. 44.2 The continuing trend toward more powerful application satellite (Graphic Courtesy of J. N. Pelton)

reduce the cost and mass of the various sources of power supply. The satellites that require the most power such as direct broadcast satellites, mobile satellite communications, and radar sensing systems, because they represent the greatest challenges will likely lead the way forward. As can be seen in Fig. 44.2, the trend line has been and continues to be larger and more powerful satellites can beam information to smaller and more mobile user terminals.

Tracking, Telemetry, and Command

All forms of application satellites require a tracking, telemetry, and command (TT&C) system. The different types of applications satellite systems for various services have different requirements and challenges in terms of TT&C. The precise ranging and tracking methods used to detect the precise position of a GEO satellite that is one tenth of the way to the Moon are much more challenging than carrying out this function for a LEO satellite that is in an orbit that is 40 times closer to Earth. On the other hand, a LEO constellation with perhaps 50, 60, or more operational satellites in a global network plus spares makes tracking of such a large number of spacecraft much difficult to "see" from the ground. This is simply because the Earth blocks the ability to track the satellite from a particular location except for a brief period of time.

The problem of tracking a LEO constellation (as well as collecting telemetry data or sending it commands) breaks down into three different solutions. All of these are rather expensive. One solution is to have a large number of TT&C

facilities on the ground to carry out these functions. This could require as many as 100 locations with up to three TT&C facilities ability to track in different directions at each location. (In practice the actually number is quite a bit less because of the geography of the oceans, polar caps, etc. The Global star system, for instance, has an extensive terrestrially-based TT&C system to track its constellation that extends from 55° north latitude to 55° south latitude.) The second solution is to equip a LEO constellation with inter-satellite links such as the iridium system that has each satellite with links to four other satellites in a highly symmetrical constellation. In this case the links are to the two satellites that are either ahead or behind and another two that are to satellites that are across. This allows the tracking, telemetry, and command functions to be carried out at only one site with another backup site. The third solution, which has been used by a number of the space agencies such as NASA, ESA and JAXA, has been to deploy in GEO orbit a tracking and data relay satellite network. Three such satellites can collect information from the LEO or MEO satellites and then relay the information to a global command and control center. Again the satellites in LEO or MEO orbit would have to be equipped with TT&C antenna systems to accommodate such a space-based solution.

The bottom line is that TT&C systems are technically challenging and expensive to engineer and operate, but they are nevertheless essential to reliable functioning of any satellite application network. Despite the differences between orbits and mission requirements here is a great deal of common technology with regard to TT&C that can be shared.

Ground and User Systems

The general trend in telecommunication satellites with regard to user antennas, handheld transceivers, and receive-only terminals has been to migrate from centralized systems utilizing very large and expensive antenna systems to more and more decentralized units that consumers can purchase and operate themselves. This same trend has now obviously continued with regard to space navigation systems. In the area of remote sensing and meteorological satellite systems, this same decentralization process is now also beginning to occur. In the case of remote sensing and meteorological systems, the practice for decades has been to relay the incoming data from satellites to processing centers that subsequent release data to various institutions and the public. The advent of expert system and artificial intelligence has led to new more immediately distributed applications.

For instance, research is ongoing that would allow airline pilots to receive information in the cockpit during flights in “near real time” about ambient weather conditions that is relayed from meteorological satellites. Remote sensing data and space navigation information can now be increasingly acquired directly by defense forces operating in the field via deployable mobile receiving stations. The overall trend seems to be to move voice, video, and data acquired directly from all types of application satellites and relayed to the end user at the “edge” of various satellite distribution or broadcast networks. This means that compact, transportable, and

low-cost transceivers will be increasingly in demand. Critical to this development will be lightweight, high energy density batteries that can be easily recharged as well as more and more capable application-specific integrated circuits and monolithic devices that allow more and more miniaturization. Improved antenna systems, both on application satellites as well as on user terminals will also be a critical development path forward. In summary, there are key parallel trends across the applications satellite industry to develop better user terminals that are lower in cost, more transportable, equipped with improved batteries and antenna systems that allow higher throughput rates and higher transmission speeds.

Launch Services

Launch services, in terms of cost, reliability, and flexible availability, are critical to the success of the applications satellite industry. For many years it was the demand for telecommunications satellite launches that drove the commercial launch services industry. For the years from 2005 through 2008, Ariane-5 launch vehicles placed 40 telecommunications satellites in orbit and only two non-telecom payloads. Prior to 2005, of the 155 satellites successfully launched by Ariane-4 in the course of its operation, 139 were telecommunications satellites. Other providers of launch services in China, India, Japan, Russia, the Ukraine, and the United States likewise were heavily oriented toward the launch of communications satellites. Although the growing diversity of commercial markets for applications will continue to expand the demand for commercial launches, telecommunications satellites will likely predominantly drive the market for years to come.[2]

This is to say that the sizing of payloads to accommodate various classes of launch vehicles and the design of future direct broadcast and mobile satellite communications systems will drive the design of the largest spacecraft platforms and their sizing to meet the "fairing dimensions" with the largest launch vehicle available. At the other end of the spectrum the platforms used on microsatellites will likely be designed for telecommunications satellites and then adapted for use by other satellite applications. Launches of smaller satellites could be multiple launches at the same time. For instance, multiple Orbcomm satellites were simultaneously launched by the Orbital Sciences Corporation's Taurus launch vehicle and Globalstar unsuccessfully attempted to launch seven of its satellites on a Zenit launcher. In other cases such launches might be "piggyback" launches accomplished on larger vehicles with special configurations to launch one or two major payloads plus a several other microsatellites at the same time. Regardless of the configuration and the payload, the launch industry is most likely to design lift capabilities optimized by demand from the communications satellite industry because of the relative size of the market demand.

The history of launch vehicle development has been dominated for the last 50 years by chemically fueled launchers – using either liquid or solid fuels. This development has led to more and more reliable lift capabilities but only modest reduction in launch services costs. Although launchers have become more and more

reliable, launch insurance with premiums in the range of 12–20% of the cost of the insured risk have also remained relatively high.

Currently there is a great deal of research to develop new launch or orbital insertion technologies. Much of this focus is on higher and higher performance ion engines and electrical propulsion. These efforts are aimed at adding reliability and cost reduction, but today these systems only have limited thrust capabilities suitable for positioning, station-keeping, or orbital insertion after liftoff via a chemically fueled rocket. Other lift capabilities such as nuclear propulsion, use of tethers for orbital elevation, and higher performance electrical propulsion systems are under R&D investigation. Other unconventional approaches such as "rail-guns," "space elevators," lighter-than-air "dark sky" stations as liftoff sites for ion engine propulsion, etc. are also under study. Any new lift capability that would increase reliability, decrease cost, lessen the cost of launch insurance, and perhaps also reduce the polluting effects of chemically fueled launches – particularly solid fueled systems – would be a boon to the entire satellite applications industry.

Common Operational and Regulatory Aspects of Application Satellite Programs

All application satellites must have some degree of operational management and control to maintain the satellite's health. There must be conditioning of batteries, monitoring of solar array performance and orientation, firing of jets to maintain proper orbits, monitoring of thermal control systems to see that temperatures are maintained within appropriate limits, etc. Most importantly there is an ongoing review of payload performance to ensure that the desired information is flowing to and from the satellite without significant interference and that key operational components are performing correctly.

Most tracking, telemetry, command, and operational monitoring functions are automated with alarms sounding if a fixed parameter limit happens to be exceeded. In such cases alarms sound and trouble-shooting efforts begin to identify corrective measures. This might be to reorient the space craft in the case of thermal difficulties or to command a switch to a back up sensing device or communications transponder. Since application satellites are constantly subject to in orbit hazards such as solar flare, a component failure, or even a possible collision with space debris or another satellite, operators must have a 24/7 capability to respond to any difficulty that may arise. In the case of geosynchronous satellites that are well above the Van Allen belts, have relative stable orbits, have less risk of physical collision and continuous TT&C links with a single facility on the ground, satellite operations are generally less complicated than is the case with a very large constellation of low-orbit satellites with multiple inter-satellite links. In the initial deployment of such systems there can be so-called cockpit errors when satellites are being tested, initial parameters checked and multiple anomalies being addressed at once.

The increasing complexity of satellites, particularly in terms of communications satellites that may require the instant interconnection of literally hundreds of spot

beams is driving the operation of such networks toward increasing automation of the operating process as well as sophisticated self-diagnostic capabilities to identify system faults. For instance, in the case of the Intelsat 4 satellite, where there are on the order of 240 global, zonal, and spot beams to interconnect in real time, the number of potential pathway interconnections is $n/2(n-1)$ or $120 \times 239 = 16{,}780$ pathways. Without digital processing support a rapid determination of a pathway fault would be almost impossible.

The most ambitious operational objective is what is called autonomous operation. This is to design the software on the satellite with as much "intelligence" as possible. The idea is not only add onboard switching and signaling capability but also design the satellite so that it can operate as independently as possible with very little ground-based monitoring and control. Experience with experimental satellites designed by ESA, NASA, and JAXA have shown that truly "smart" and virtually totally automated satellites are still some years away.

Another serious operational concern that is common to all application satellites is the possibility of hackers or even terrorist organizations obtaining unauthorized access to an operational commercial satellite and commanding it to move out of the correct orientation or position, discontinue service, fire its jets to put the satellite into a descent trajectory that would remove it from orbit, or to otherwise send spurious commands that would disable the satellite. There are several instances of attempts by hackers to gain access to satellites and send spurious commands. Most operational satellite systems have sophisticated "firewalls" and command controls to prevent such occurrences. Most large-scale systems with global operations have sophisticated codes that must accompany commands. In some cases there is an additional "failsafe" requirement that another TT&C station must send a confirmation code.

There is an ongoing effort within the satellite applications industries to use artificial intelligence and expert system software to continuously monitor the health and performance of each satellite. The number of lines of code associated with the TTC&M functions (i.e., tracking, telemetry, command, and monitoring) can now run to millions of lines of code for the most sophisticated satellites. The trend for the field is thus toward greater onboard intelligence, autonomous operations, and operational security.

The regulatory aspects of application satellites are an area where there may be more dissimilarities than there are common elements. The field of telecommunications satellites has often served as the major shaper of international, regional, or national regulatory policies. This logically follows from the following key points: (1) Satellite communications represent the earliest major satellite applications and thus first shaped regulatory policy and decisions and process related to radio frequency allocations; (2) Satellite communications services represent by far the largest market for commercial satellite applications; (3) Satellite communications services (to a much greater extent than other satellite applications) tend to represent a competitive service with respect to terrestrial communications networks, domestic industries, and program service providers; and (4) Trade regulations, landing licenses, intellectual property right protections, competitive access to local markets,

particularly those as addressed by the World Trade Organization have largely concentrated on satellite communications in rule making processes (largely for the reasons indicated in 1, 2, and 3 above (Pelton 2005).

A company or entity wishing to operate a satellite application business within a particular country will thus often find that the rules for being able to operate, sell services, have "landing rights" or protect their intellectual property will find a set of regulations largely defined within the context of satellite communications. The same is true with regard to how the World Trade Organization (WTO) addresses issues of opening of national markets to competition and what national regulatory practices are considered acceptable and valid. As the transition is made from "voluntarily declared" open competition policies for services to mandatory plans that are overseen by the WTO with the authority to impose fines for noncompliance, the rules for market entry and competitive access to commercial space markets will likely become more contentious for all types of satellite applications.

Common Market and Business Considerations in Application Satellite Programs

The global commercial markets for satellite applications, as noted above are becoming increasingly competitive on a global basis. This may not only make the regulatory aspects of the business even more challenging in terms of national landing rights, requirements to maintain local offices equipped with marketing, regulatory, and technical staff, but will also tend to make market competitiveness even more important. Satellite operators that have the advantage of economies of scale will likely be able to negotiate reduced prices or obtain price advantage through competitive contract awards for satellite purchases, launch services, and launch insurance arrangements. These larger organizations will also be able to obtain TTC&M services at lower net costs simply because they have more satellites in orbit and they can spread these costs for these services more efficiently in terms of net operational cost per satellite. The same type of economy of scale also applies to global marketing efforts as well. For those commercial satellite organizations that are selling their products globally the market and business elements will likely see

- A very highly competitive market
- Economies of scale continuing to be extremely important
- The challenges of being compliant with all regulatory requirements ever more difficult at the national, regional, and international level
- Issues related to regulatory licensing and landing rights at the national level more laborious with increasing likelihood of these requirements serving as "nontariff barriers to competitive entry"
- Access to necessary frequency allocations more challenging and intersystem coordination more difficulty – in large part due to the expanding number of commercial satellites
- Market domination by the largest of the suppliers

Certain innovations can aid commercial business planning. One of the most important innovations can be lifetime extension that allows operators to do more without needing to make large capital investments simply because their satellites can last longer in orbit. Innovation in autonomous operation can serve to reduce operating costs. One of the biggest challenges could be the development of alternative technology. There is continuing R&D to develop High-altitude platform systems that might be able to provide wide areas of telecommunications, broadcasting, remote sensing, and surveillance as well as even meteorological services. There could also be development of multipurpose, large-scale spacecraft buses that could offer a variety of services that might cut across lines of the traditional commercial satellite application industry. Satellite with multi-payloads could offer telecommunications, broadcasting, and/or remote sensing services. At this point such developments can and should be considered long shot possibilities, but one should always be attuned to the fact that technologies, system economics, and markets often change when innovations emerge.

Dissimilar and Unique Requirements in Application Satellites

The analysis up to this point has largely focused on commonalities among the various application satellite technologies, the various operating systems, and satellite sparing practices as well as the TT&C services. It has also been noted that regulatory and business practices and considerations are often common as well. Although there are many common trends and parallel patterns for the various types of application satellites, it is perhaps equally important to note the important dissimilarities that also do exist.

Differences That Stem from the Differences in the Various Satellite Markets

Clearly there are different service and availability requirements that impact the technical design of satellites and also their operational management. All the big three types of telecommunication services (broadcasting, fixed, and mobile) all involve real time communications and thus exacting availability standards. The Integrated Services Digital Network (ISDN) standards only allow for about an hour of outage in an entire year. Data relay (or machine to machine (M2M)) satellites can perform to a lesser standard but there are also high availability expectations for this type of satellite as well. Likewise, space navigation satellites that are now use to support aircraft takeoff and landing must have the highest availability service of any type of application satellites. This is why there is extensive deployment of spare satellites in both telecommunications and space navigation satellite systems. Likewise there is often redundancy of critical components within these satellites.

This is not to say that in the case of meteorological and remote sensing satellites there is not a need for reliability and continuous system availability. Nevertheless

the "sparing philosophy" is not as exacting for these systems. There can be, for instance, a reliance on the meteorological and remote sensing satellite systems of other countries in the case of unexpected outages.

Since space navigation satellites are used for strategic and defense-related purposes and since these satellites are also use for a number of applications that are human life dependent, these satellites are designed to the highest standards in terms of radiation hardening, redundancy, and system sparing. This is closely followed by telecommunications satellites, with defense-related satellites designed to the highest system availability standards and telecommunications satellites designed to achieve 99.98% availability or better in most instances. This is a particular challenge for satellites that operate in the Ka-band frequencies and higher due to loss of signal from high rates of rain attenuation. This requires a high level of link margins for satellite beams that experience intensive rainfall.

The remote sensing and meteorological satellites are also designed to the highest standards with redundancy of critical components. Nevertheless because of sharing of data and other precautions, service availability at such a high rate as 99.98% is not an engineering requirement. In some instance spare satellites are maintained on the ground and not launched until needed. The same approach can also be used in some cases by national communications satellite operators, particularly when there is spare capacity that might be obtained from other satellite or terrestrial sources. Finally the tracking, telemetry, and command and monitoring systems for space navigation and satellite communications must be engineered in a more exacting fashion to be able to monitor operations on a 24/7 basis and to be able to respond to anomalies or interference on a virtually instantaneous basis. In the case of meteorological systems the same degree of global interconnectivity for TTC&M systems is not seen to be as critical.

Another very obvious difference is that the commercial satellite telecommunications markets worldwide are significantly larger than other commercial satellite applications. This difference does not have a technical impact, but it does have a business and operational impact. The commercial satellite market offers the largest operators opportunities to achieve economies of scale in many elements of their procurements and operations. At least as far as spacecraft procurement goes for the rest of the industry they can benefit from the use of similar spacecraft platforms, solar arrays, battery systems, and perhaps TT&C systems, but the rest of the spacecraft is likely to be highly specialized and with high engineering and nonrecurring costs.

The commercial satellite industry is also the largest by far of all the commercial application satellite services. The business models and the nature of "customer relations" are, not too surprisingly, different from those of the entities that own and operate space navigation, remote sensing, and especially meteorological satellite systems. The satellite communications markets, especially for direct broadcast satellite services and broadband Internet access, have made the greatest transition toward selling "retail services" directly to consumers. In the case of space navigation, remote sensing, and meteorological satellites, the "marketplace" for their

services remains predominately at the "wholesale" level in terms of the size of the revenues that support the operation of the satellite networks. Individuals may pursue space navigation units and farmers may purchase remote sensing data, but these "retail purchases" provide modest or no support for the operation of these satellite networks.

It is largely governmental agencies, large corporations, or international organizations that make the major purchases that support the operation of the "other" application satellite networks. It was only when direct broadcast satellite companies were able to sell on a retail basis their services to consumers that these industries' revenues increased sharply in value. The same is likely true for the other parts of the commercial satellite industry. Thus it is only when successful business models that allow application satellite providers to sell their services directly to individuals on a "retail basis" that these industries can be expected to experience exponential growth of revenues. There are currently plans for the deployment of new types of application satellites to beam back electrical energy from space-based solar power satellites. Most of the business models for these new projects involve selling their solar-derived energy directly to energy companies. Only if these new companies can find a business model whereby they could sell their services directly to consumers, can they expect to achieve true commercial success and high levels of profitability.

Design and Engineering of Satellite Payloads: Communications Antenna Systems, Multispectral Sensors, and Radar Systems

The companies that contract for the integration and delivery of application satellites listed earlier such as Alcatel, Alenia Thales, Astra, BAE, Boeing, EADS, Lockheed Martin, Mitsubishi Electric Company, Motorola, Northrop Grumann, Orbital Sciences, and so on essentially design and manufacture the main elements of the spacecraft bus but rely on specialized companies to design and build specialized hardware such as an 18-m deployable antenna for mobile satellite communications, an active radar system or multispectral sensor for a remote sensing satellite, or an atomic clock for a space navigation satellite. It is the "payload" of an application satellite that defines its purpose and may represent a third or more of the cost of the entire satellite project. While there is considerable synergy in the design, engineering, test, and operation of application satellites, there is often little to none when it comes to the actual payload of a particular satellite.

It is the objective of commercial satellite manufacturers to make the design, engineering, manufacture, and test of the building blocks of an application satellite (i.e., bus structure, power system, TTC&M, stabilization and orientation system, thermal control system, etc.) as common as possible for various platforms that could be used for communications, remote sensing, space navigation, and meteorological observation. This, as previously observed, helps reduce costs, aids reliability, and speeds production and testing. The "non-common elements" of the

payload represent the largest technical challenge and this highly specialized part of the spacecraft is typically designed and built by other contracts that specialize in particular payload that is required.

Conclusion

Application satellites are now a key part of the world economic structure. Billions of consumers rely on satellites to receive their daily news, see sporting events, to access the Internet, to know of current weather conditions, and especially to learn of threatening storms. Passengers on airplanes, on ships, and in cars and buses rely on space navigation systems for safe passage every day. The aerospace industry has learned to standardize many technologies related to the "spacecraft platforms" that are fundamental to the manufacturing of reliable and cost-effective application satellites. There are many technologies still to be explored to design better and lower cost application satellite platforms, but progress continues to be made. Promising areas of future research and development (R&D) that can extend performance, increase reliability, and/or reduce costs are discussed within this chapter. One of the areas where there is considerable interest is finding even more reliable and cost-effective ways to deploy application satellites to orbit and make the launch services more competitive. One of the biggest differences among various satellite systems in terms of cost, operations, TT&C, and launch cost is whether the network employs LEO, MEO, and/or GEO satellites. Here there may or may not be commonality between various application satellite systems, but innovations do transfer from one market to another as better designs are found or improved ways are found to operate networks. This is particularly true with regard to LEO and MEO constellations. This is, however, largely due to the fact that there is now almost 50 years of experience with GEO satellite operations and the learning curve for these satellites is quite simply much longer.

Cross-References

- ▶ Lifetime Testing, Redundancy, Reliability and Mean Time to Failure
- ▶ Overview of the Spacecraft Bus
- ▶ Telemetry, Tracking, and Command (TT&C)

Notes

1. D.K. Sachdev, Three growth engines for satellite communications, in *AIAA 20th International Communications Satellite Systems Conference*, Montreal, Canada
2. European Space Agency (ESA) Telecommunications: The Satellite Market http://www.telecom.esa.int/telecom/www/object/index.cfm?fobjectid

References

L. Ippolito, J.N. Pelton, Satellite technology: the evolution of satellite systems. *Communications Satellites: Global Change Agents* (Lawrence Erlbaum, Mahwah, 2004), pp. 33–54

J.N. Pelton, *Future Trends in Satellite Communications: Markets and Services* (International Engineering Consortium, Chicago, 2005). Chapter 6, pp. 73–95

J.N. Pelton, *The Basic of Satellite Communications* 2nd edn. (The International Engineering Consortium, Chicago, 2006), Chapter 6, pp. 119–140

Section 6

Launch Systems and Launch-Related Issues

Launch Vehicles and Launch Sites

45

Joseph N. Pelton

Contents

Introduction 1132
Early History of Rocket Technology and Systems 1133
Development of Solid-Fueled Missile Systems 1134
Liquid-Fueled Launchers 1136
Avionics and Guidance Systems 1136
Launch Options for Commercial Application Satellites 1138
Cost of Launches and New Commercial Options 1140
Launch Sites 1141
Station-Keeping, Spacecraft Operations, and End of Life 1142
Notes 1144
References 1144

Abstract

The Handbook of Satellite Applications focuses on the practical applications of satellites. This means that the handbook addresses the many uses that are made of communications, remote sensing, satellite navigation, and meteorological systems as well as the spacecraft, the ground systems, and tracking, telemetry, and command systems that make these networks possible. There are also chapters that address regulatory issues, economic and insurance issues, and even threats to the future operation of application satellites. This chapter addresses the remaining critical areas that are critical to the successful operation of application satellite systems.

All type of applications satellites could not carry out their function unless they were first launched into the right orbit. Even after successful launch they must also be properly maintained there through necessary station-keeping

J.N. Pelton
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: joepelton@verizon.net

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_72, © Springer Science+Business Media New York 2013

operations. This chapter addresses the history of rocket and launch vehicle development and explains the basic technical capabilities that allow applications satellites to be placed into orbit with greater and greater reliability. This chapter also briefly addresses in-orbit operations that allow spacecraft to be maintained in orbit and to operate over increasingly long practical lifetimes. Over the past 60 years of the space age an expanding variety of different propulsion systems and launch systems have been developed to carry out the important tasks of launch, station-keeping, and deorbit or removal of spacecraft to a graveyard orbital location.

One of the key elements of success for applications satellites of all types is the fact that gradually the reliability and the lift capability of launch vehicles have improved over time. It has been hoped for many years that new technology could allow the cost of launches to be significantly reduced, but to date such breakthroughs in the economics of launch systems have not yet been achieved. The precision thruster systems that allow spacecraft to be pointed with ever greater precision and to maintain crucial station-keeping have quite successfully continued to evolve. This has allowed application satellites to operate for much longer lifetimes and with greater pointing accuracy that has increased their functionality.

Keywords

Avionics • Celestial mechanics • Computer guidance • Fairing • Guidance • International Telecommunication Union (ITU) • Ion engine • Launch vehicle • Launch sites • Liquid fuel • Propulsion • Rocket stages • Solid fuel • Station-keeping • Thrusters

Introduction

The history of rocketry and propulsion is actually quite long and complex, but it has only been in the last 60 years that practical launch systems have been developed and implemented. The development of rocket systems has, in many ways, been a dual pathway forward. One pathway has been the improvement of missile systems developed by military organizations as weapons for the delivery of bombs, and this approach has largely focused on solid rocket systems that use the controlled force of solid explosives to deliver a payload to a particular location against an adversary. The other pathway has been that of civilian space programs that have tended to focus more on liquid propellant systems to launch applications or scientific satellites to orbit or even to launch payloads with humans aboard into space. Liquid-fueled rockets cannot be fired instantaneously like solid-fueled rockets that, in effect, can be launched essentially by the push of button that is akin to lighting the fuse on a bomb. Both types of launch systems have paid an important part in space activities over the past 60 years.

The purpose of this chapter is to explore what types of launch capabilities exist to place application satellites into orbit and the newly evolving capabilities that

allow this to be done more effectively, more reliably, and at lower cost. This chapter will also address the importance of launch sites and launch range safety in the successful launch of applications satellites. Finally the future of launch systems will be briefly explored. There are many interesting and in-depth books on launch systems that provide greater detail on the rocketry and launch sites that will be indicated in the endnotes for those who wish more information on this subject (Michael Lennick 2006; Travis 2007). One of the most important features of the Handbook on Application Satellites are the two appendices provided at the end of this book that report in detail about all of the many launcher systems that are available from around the world today as well as the many launch sites. The new growth and development of commercial launch systems that characterizes the world of rocketry today will undoubtedly bring a wider range of capabilities to the owners and operators of application satellites in the decades ahead.

Early History of Rocket Technology and Systems

The concept of rocket propulsion is actually thousands of years old. Archytas of Tarentum (428–347 BC), who was friend of Plato and an outstanding mathematician and scientist, discovered the principles of propulsion almost 2,500 years ago. He devised a wooden pigeon that used steam-powered jet propulsion to fly around Archytas' home in ancient Greece. Ancient Chinese speculated about the use of gunpowder-powered rockets to fly into space. Through the ages writers of science fiction suggested the use of rockets to fly to the Moon and beyond. Everett Edward Hale in the nineteenth century actually wrote of launching an application satellite – a Brick Moon – into polar orbit for the purpose of navigation and communications (Pelton and Global Talk 1981).

With Newton's discovery of the gravitational effects of the Earth, and his explicit description as to how a projectile fired with enough speed could escape the Earth's "gravity well," however, eventually gave rise to serious thought about rockets and space travel. The Russian scientist Tsiolkowsky conceived of rocket designs that could actually carry people into space and others such as Willy Lev began a systematic study of space and rocketry. The American Robert Goddard is considered by many to be the modern-time father of rocketry. Beginning in the 1920s, he developed a series of increasingly powerful prototype rockets with liquid fuel propulsion and elaborate stabilization systems. Although mocked for his experiments at the time by the *New York Times* as the "Moon Man," in part because his rockets had limited range of his experiments, he persevered and developed more and more capable launchers. Goddard's work was increasingly taken seriously and his efforts rather directly led Wehrner von Braun and his team in Germany to develop the V-2 rockets as bomb delivery systems during the Second World War.

After the Second World War Soviet scientists and engineers continued work in rocketry to develop missile weapons systems as well as rockets that could achieve orbit as did the United States but with much less funding and concerted governmental support behind the American efforts. This all changed in October 1957

when the Soviet Union launched the Sputnik I satellite into Earth orbit. The United States formed the National Aeronautics and Space Administration (NASA) in 1958 and gave significant new funding to develop missile technology within the US military as well as support to Dr. von Braun's team in Huntsville, Alabama to develop civilian rocketry capability.

The years that followed gave rise to a major "missile race" between the United States and the Soviet Union. The US election for President in 1960 won by President John F. Kennedy was largely focused on what has called the "missile gap" between the United States and the Soviet Union. The United States invested heavily in developing civilian rocket technology and even more money was invested in military rocket and missile systems. The 1960s culminated with the Apollo 11 mission that sent three astronauts on a lunar exploration mission that allowed two astronauts to land and explore the Moon's surface in July 1969. Since the start of the Space Age over 500 people have gone into space via rocket launches and a number of space stations have been launched to sustain astronauts and cosmonauts in orbit for sustained periods of time.

In the new age of commercial space travel it is possible that a much larger number of people will be able to travel into space. Today not only are commercial space planes under serious development for flights starting in 2013, but private space stations are planned for launch as well. The idea of people being able to travel into space on commercial space planes was quite vividly envisioned in the Stanley Kubrick and Arthur C. Clarke movie *2001: A Space Odyssey* and today that prospect seems to transcend the jump from science fiction to science fact.

Development of Solid-Fueled Missile Systems

As noted earlier there are two basic kinds of chemical propellants for missiles and rocketry systems. These are liquid- and solid-fueled launch systems. In both cases the fuel is "oxidized" or ignited to create a powerful discharge of gas in order to provide the needed chemical propulsion or "thrust" to lift the rocket skyward. More recently electrical propulsion systems have also been developed. These systems (i.e., ion thrusters) have much lower thrust but can operate for much longer periods of time and provide more net thrust over time. In the case of electrical propulsion ions are expelled at very high velocities for guidance and station-keeping systems. These systems can be more reliable and provide higher thrust to mass ratios than chemical propulsion systems but do not have sufficient thrust to provide lift off from the Earth's surface. These thrusters, however, can be used for station-keeping and final orbital positioning. In time nuclear-fueled systems or other more exotic capabilities such as tether-lift systems, rail guns, or even the so-called space elevator may be used to provide access to Earth orbit but today chemically powered rockets are still the exclusive way for application satellites to attain orbital access.

Solid- and liquid-fueled rockets both have advantages and disadvantages. Liquid propellants tend to require complicated piping and very high performance pumps to feed the rocket engines with a large and steady stream of rocket fuel. Liquid-fueled

rockets can provide greater net propulsive thrust over their full period of operation, but they require time to fuel and thus quick liftoff is not possible. Further elaborate storage, handling, and fueling systems are required for this type of rocket. Also from a safety perspective it is possible to throttle and precisely control the flow of fuel to the combustion chamber in the case of the liquid-fueled rocket. This means that a liquid-fueled engine can be immediately shut down by closing a valve that shuts off the flow of propellant to the engine. A liquid-fueled rocket is more complicated in design and slower to fuel and launch. The liquid-fueled rocket is also slower to build up thrust because of the pumping of the fuel into the combustion chamber.

Solid rockets do not require complicated engines, pumps, or plumbing. Instead the thrust of these rockets depend on the explosive power of the solid-rocket fuel and require stronger casings to endure the great pressures that come from the exhaust thrusts. Since they are essentially a controlled bomb they can be ignited much more rapidly and their initial acceleration at liftoff is nearly instantaneous. From a safety viewpoint, however, most solid fuel systems (except for the hybrid systems that use nitrous oxide as an oxidizer and neoprene rubber as the fuel) cannot be throttled or controlled once ignition has been achieved.

The above factors are some of considerations that are taken in account in terms of turning to solid fuels for missile systems as well as decisions to utilize liquid-fueled systems for systems involving astronauts. Such flexibility as to what propulsion system to use was not readily available at the start of the space age. This was simply because liquid-fueled systems became available first – starting with the Robert Goddard developments and the V-2 which were liquid-fueled systems.

Thus the V-2 in Germany, the Atlas, the Thor, and the Jupiter in the United States, and the R-7 ICBM in the Soviet Union were all liquid-fueled systems. These systems took a fair amount of time to launch. They had to be loaded not only with a fuel but also with the oxidizer and this loading process was not only time consuming but also quite dangerous because a spark could set off a very dangerous explosion.[1]

As noted above the first impetus to design and build launchers was in the context of war and weapon systems. Both the United States and the Soviet Union in the post–Second World War time period proceeded to build missiles, equipped with atomic bombs and in time with hydrogen bombs. These weapons systems served as primary deterrence against attack during the Cold War that existed between the United States and the Soviet Union for the decades that followed.

For the reasons noted above there was a concerted effort to develop solid-fueled systems for quick, easy to fuel, and less hazardous fueling and launching operations. In the United States, the limited capacity Scout rocket, the Polaris and the Minuteman systems were developed using solid-fuel propellants. The Minuteman – named for its quite launch capability – was designed to be suitable for instantaneous launch from land-based missile silos and the Polaris was developed so that it could be launched from submarines. Clearly a liquid-fueled system would be extremely difficult to deploy from a submarine, although the Soviet Union managed to equip their largest submarines with liquid-fueled

systems for a number of years because of problems in developing solid-fueled systems (Spaceflight: rockets and missiles. http://centenialofflight.gov/essay/SPACEFLIGHT/solids/SP13.htm).

In the Soviet Union the liquid-fueled R-7 was used not only to launch Sputnik but also employed for the manned missions that followed – starting with Yuri Gagarin's flight. The U.S.S.R. continued to rely on the R-7 for a decade even though there were serious problems with loading of fuel on this missile and it was very slow to prepare for launch. It was not until 1971 – almost a year after the United States began deploying solid-fueled systems that the Soviet Union deployed the first RT-2 system. In fact the liquid-fueled systems remained in service as Soviet weapons systems for another decade. The Soviet Union and now Russia continue to rely primarily on liquid-fueled systems for its civilian space program.

Today the number of countries with some launch capability continues to increase as can be seen in Appendix 2; those countries now include China, Europe, India, Iran, Israel, Japan, the Republic of Korea, Russia, the Ukraine, and the United States. A number of other countries, or commercial companies based in other countries, are also developing sounding rockets and launchers with modest lift capabilities. These include Argentina, Australia, Brazil, Canada, and North Korea.

Liquid-Fueled Launchers

Although some small satellites have been launched by the solid-fueled Scout vehicle, most of the commercial applications today are launched by liquid-fueled vehicles. These liquid-fueled systems almost always fall into the three categories of petroleum-based fuels (most typically a highly refined kerosene known as RP-1), cryogenic fuels (most typically liquid hydrogen) or hypergolic fuels (most typically some form of hydrazine). The Chinese Long March vehicle uses a form of hydrazine (known as UDMH) in its first two stages with nitric acid acting as the oxidizer plus some Russian vehicles use hydrazine as well, but virtually all other major launch vehicles rely on either liquid hydrogen (LH_2) and liquid oxygen(LO_2) or on RP-1 rocket fuel with an oxidizer (typically LO_2). In some instances there are solid fuel strap-ons to supplement liftoff capabilities. The following table adapted from data prepared by Robert A. Braeunig indicates the specific impulse of various types of types of rocket fuels when working with their various oxidizers (Braeuni 2008) (Table 45.1).

Avionics and Guidance Systems

Many people think of rocket launcher systems as powerful explosive systems that power a rocket to orbit and do not necessarily consider that there must not only be great thrust achieved through the rocket engines, but that there must also be a very accurate guidance systems that steers the rocket and its payload on the exact pathway

Table 45.1 Rocket propellant performance (Adapted and simplified from a chart prepared by Robert Braeunig)

Combustion chamber pressure, $P_c = 68$ atm (1,000 PSI). Nozzle exit pressure, $P_e = 1$ atm				
Oxidizer	Fuel	Hypergolic	Mixture ratio	Specific impulse (s, sea level)
Liquid oxygen	Liquid hydrogen	No	5.00	381
	Liquid methane	No	2.77	299
	Ethanol + 25% water	No	1.29	269
	Kerosene	No	2.29	289
	Hydrazine	No	0.74	303
	MMH	No	1.15	300
	UDMH	No	1.38	297
	50-50	No	1.06	300
Liquid fluorine	Liquid hydrogen	Yes	6.00	400
	Hydrazine	Yes	1.82	338
FLOX-70	Kerosene	Yes	3.80	320
Nitrogen tetroxide	Kerosene	No	3.53	267
	Hydrazine	Yes	1.08	286
	MMH	Yes	1.73	280
	UDMH	Yes	2.10	277
	50-50	Yes	1.59	280
Red-fuming nitric acid (14% N_2O_4)	Kerosene	No	4.42	256
	Hydrazine	Yes	1.28	276
	MMH	Yes	2.13	269
	UDMH	Yes	2.60	266
	50-50	Yes	1.94	270
Hydrogen peroxide (85% concentration)	Kerosene	No	7.84	258
	Hydrazine	Yes	2.15	269
Nitrous oxide	HTPB (solid)	No	6.48	248
Chlorine pentafluoride	Hydrazine	Yes	2.12	297
Ammonium perchlorate (solid)	Aluminum + HTPB (a)	No	2.12	266
	Aluminum + PBAN (b)	No	2.33	267

Notes:
Specific impulses are theoretical maximum assuming 100% efficiency; actual performance will be less.
All mixture ratios are optimum for the operating pressures indicated, unless otherwise noted.
LO_2/LH_2 and LF_2/LH_2 mixture ratios are higher than optimum to improve density impulse.
FLOX-70 is a mixture of 70% liquid fluorine and 30% liquid oxygen.
Where kerosene is indicated, the calculations are based on n-dodecane.
Solid propellant formulation (a): 68% AP + 18% Al + 14% HTPB.
Solid propellant formulation (b): 70% AP + 16% Al + 12% PBAN + 2% epoxy curing agent.

needed to acquire the proper orbit. The celestial mechanics needed to calculate the proper trajectory to get to the desired orbit is quite challenging indeed. Without very fast computing and guidance systems known as the avionics system the launch of modern application satellites would not be possible. Gyroscopic systems onboard the

launcher send back to command centers data as to the rocket's exact location to the Earth so that the thrust vectors of the rocket and the gimbals on the rocket motors can be corrected to keep the launcher on the exact orbital trajectory. Millions of data points and calculations are made at extremely rapid speeds to keep the rocket exactly on course in all three x, y, and z reference planes (i.e., of roll, pitch, and yaw).

Completely different trajectories must be calculated for launches to polar, LEO, MEO, or GEO orbits. There are even different calculations and trajectories created if a rocket is launching a single payload or multiple payloads. Most modern launchers are multiple-stage rockets that must separate between firings of the stages and these elements of the launch must be accommodated by the avionics system and the tracking data to make sure there is a successful launch. Even the separation of the rocket's nose fairings at high altitudes and the spinning of the spacecraft at the time of release from the rocket are parts of the launch operation that need to be carefully planned and monitored in real time via telemetry sent down to the launch command center and tracking stations around the world that monitor and control every aspect of the launch.

This constant telemetry monitoring operation is not only critical to the successful launch operation, but has another important element as well. Especially trained range safety officers are also directly involved in the launch operations. In case the rocket should for some reason go off course there could be a need for the errant rocket to be destroyed in order to avoid loss of human life and perhaps buildings or even cities on the ground. Launch insurance arrangements typically include liability protection against an unsuccessful launch and to protect in particular for a rocket that might go off course and thus lead to substantial damages or lead to causalities. The possibility that a rocket could actually land in a city – such as a launch from the Kennedy Space Center landing in Miami, Florida – could lead to an incredible level of destruction and loss of human life, and thus special liability arrangements have been made in such a case that when commercial insurance liability limits are reached an additional layer of liability insurance is actually provided by the U.S. Government – as do the governments of other countries that support launch operations. In this case, however, the security precautions that are in effect and the reliability of today's launch vehicles make such a possibility extremely remote. The insurance arrangements and costs, however, do strongly motivate the site location process for launch sites and space ports to be situated in remote areas with launch operations being conducted either over the ocean or in extremely sparsely populated desert locations.

Launch Options for Commercial Application Satellites

The dominant provider of commercial launches today is Arianespace with their Ariane 5 vehicle. The current dominance of Arianespace in providing launch services to commercial satellite application providers is based on a combination of reliability, launch efficiency (i.e., the Ariane 5 launch site in Kourou, Guyana is



Fig. 45.1 Ariane 5 launch from Kourou, Guyana launch site (Graphic courtesy of Ariane space)

essentially right on the equator and thus provides the maximum Earth-assisted boost to GEO orbit), plus effective pricing.

Certainly the Ariane 5′s large lift capability provides economies of scale. This allows the launch of very massive satellites or the joint launching of several satellites to GEO at the same time. Recently a new launch facility has been constructed at the Kourou launch site for the launch of Russian designed and built Soyuz and Soyuz-2 launch vehicles. This new capability at the Guyana launch facility will provide a wider range of launch options that can now be available from this equatorial site (Fig. 45.1).[2]

There are sufficient launch options available in the global market place to ensure competitive pricing. There are over a dozen competitive launcher entities offering launch services today and literally hundreds of launch options to meet needs for launch to low earth, polar, medium earth, highly elliptical or GEO orbits as is obvious by a review of the information provided in Appendix 2.

The launch options and launch site opportunities are constantly changing. It is now possible to arrange for a Soyuz-2 launch from the Kourou launch site.

The Atlas V and Delta II now provide relatively high lift capabilities and more attractive prices than in the past. It is possible that the Atlas, Delta, or other US vehicle might be upgraded to "man-rated capabilities" and the new NASA Space Launch System (SLS) that will launch the Orion capsule into Deep Space may give rise to new launch options in future years.[3] Currently both the Orbital Sciences Corporation and Space X are seeking to develop commercial launch capability that could safely fly cargo to the International Space Station and perhaps in time even fly astronauts to orbit as well. These new capabilities could also ultimately help to serve future launch needs for application satellite service providers (Lindemoyer 2011).

Perhaps most significant in terms of changing launch options are the Pegasus and Taurus launch capabilities from Orbital Sciences and especially the new Falcon launch vehicle that are truly "commercial vehicles" that appear able to offer increasingly cost effective new launch options.[4]

Cost of Launches and New Commercial Options

It is not possible to cite a specific launch cost for a particular class of vehicle. First of all the cost of a low earth orbit launch is much less than that of a GEO launch or one can lift much more mass to LEO for the equivalent cost of a GEO launch. A larger satellite with greater mass can often attain a more cost effective rate on a per kilogram basis. The interface requirements, nose fairing, and other special requirements can also affect cost. In today's market a cost of $10,000–20,000/kg for a satellite launch to GEO orbit is not uncommon. In generally these costs are expected to come down, particularly driven by new commercial launch vehicles services that offer lower costs in coming years.

Also most organizations will take out launch insurance that is often equivalent to 15–20% of the total mission cost, even though this also can vary to higher or lower levels depending on the mission. Some organizations choose to self-insure.

The series of launch vehicles developed as a total new launch vehicle and a strictly commercial venture are promising even more competitive pricing and more economical service without giving up reliability. The key to the Space X approach has been complete vertical integration and thus designing and manufacturing all elements of the launcher. A number of organizations, perhaps most notably the US Air Force, has signed up for a number of Falcon launches. If these launchers indeed prove to be reliable and the costs are significantly less expensive than other launch options, this then will clearly impact the global market for these services (Fig. 45.2) (Falcon launch vehicle. http://www.spacex.com/falcon1.php).

Some governments, particularly the United States, have significant restrictions as to what technology can be shared or released to other countries. The so-called ITAR (International Trade in Arms Restrictions) prohibit or restrict the launch of



Fig. 45.2 The launch of a Falcon 1 vehicle from its Kwajalein launch site (Graphic courtesy of spaceX)

some satellites with "sensitive technology" to be launched by certain other countries because of concerns about unauthorized transfer of technology.

Launch Sites

Appendix 1 provides a listing of available launch sites around the world. These are distributed among many countries around the world. Many of these sites are in reasonably close proximity to the Equator. This is because the Earth's rotation speed at the Equator which is in excess of 1,600 km/h provides significant assistance to a satellite launch into GEO orbit which is located in the equatorial plane. This means that launch sites essentially located at the Equator such as Kourou, Guyana, or the Sea Launch, which provides the optimum location for a GEO launch. The launch sites that are further away from the Equator such as the sites of China, Japan and the United States thus can be on the order of a 20% disadvantage in relation to the locations exactly on the equatorial plane.

There are many other sites that launch into a polar orbit, LEO, MEO or highly elliptical orbit. These sites are much less constrained as to their geographic location. A prime consideration for a launch site in all cases is to have a location

that is considered as safe as possible. This would involve the possibility of a flight path that would not cover populated areas and would allow for safe aborted launches in the case the range safety officer believed termination of flight vehicle was necessary. Today most of the launch sites are those designed for vertical lift off of chemically fueled vehicles. Under the regulations of the United States and some other countries there is a need for an environmental impact statement to be issued prior to each launch.

There are a number of commercial "spaceports" now in planning or approved. These sites vary a great deal in terms of accommodating different types of flight options. Some spaceports include provision for only horizontal take-off and landing with winged vehicles and these facilities greatly resemble a conventional airport. Other commercial spaceports are in many cases collocated with a governmentally licensed or owned launch site and are designed to accommodate both vertically and horizontally launched vehicles. These facilities are designed not only with launch pads but with specially designed fuel storage facilities and specially designed buildings and observation towers for space range officers.

Station-Keeping, Spacecraft Operations, and End of Life

The chemical rockets that are used to launch a spacecraft into geosynchronous orbit or to position satellites within a constellation of satellites in medium or low earth orbit is actually just the start of a process of orbital operations that can last for many years. The lifetime of an application satellite varies for a variety of factors. Orbital altitude is one important factor. Low earth orbit satellites, in particular, because of their lower altitude are subject to atmospheric drag that lessens their useful lifetime. Most of these satellites at the end of life are subject to a controlled descent and splash down into one of the earth's oceans. Satellites launched into MEO have intermediate lifetimes but pose the largest challenge at end of life. These satellites are not high enough to be put into so-called graveyard orbits and not low enough to be easily deorbited in a controlled manner. Fully 40% of the station-keeping-fuel needed for orbital operations must be devoted to proper end-of-life deorbit.

The Geosynchronous orbit, because it is one tenth of the way to the Moon does not present an atmosphere drag issue. Since it is so far removed from the earth's primary gravity well, GEO satellites present perhaps the easiest condition for in-orbit operations and end-of-life operation. At the end-of-life, if fuel remains operators merely push the satellite into a higher orbit where it will remain for millions of years in a so-called graveyard orbit. Even for the GEO orbits there are tradeoff considerations. It is at least ten times easier to maintain a GEO satellite within its "station-keeping box" in terms of East-West excursions as compared to North-South deviations above or below the GEO arc. As a satellite in GEO orbit nears its end-of-life operators often relax their maintenance of North-South station-keeping so that it can drift North and South of the equatorial plane in an "figure 8" shaped orbit in order to save fuel. Under International Telecommunication

Union (ITU) regulations a satellite is considered in GEO orbit as long as it remains within 5° of the equatorial plane.

There are thrusters onboard application satellites that can be fired to either reposition a satellite from one orbital position to another or to maintain it in its registered location under the official filing with the ITU.

In the earlier days of application satellites most station-keeping thruster systems used a hypergolic fuel system (and most typically hydrazine thrusters) to maintain satellites in the desired orbit. More recently there has been more and more common use of ion engines to maintain application satellites in-orbit. This is because electrical propulsion systems provide lower impulse thrust levels than chemical systems in a single burn, but they nevertheless allow for a longer operational life since they provide higher net thrust capability per kilogram of fuel. Another consideration that is increasingly coming into consideration is that hydrazine fuel is quite noxious and decaying satellites with fuel can explode and create orbital debris. Thus, even though ion-engine control systems are more expensive than hydrazine or bi-propellant systems that have been used in the past, these thruster systems are becoming increasingly common for virtually all forms of application satellites.

One of the latest developments in the operation of application satellites is the idea of creating a space tug and refueling and maintenance system in orbit. Such a device could possibly add new batteries and refuel the tanks of application satellites so that they could have an extended "second life" of perhaps many more years. Although this concept is still at an early stage, there are active experimental programs underway involving MacDonald Dettwiler Aerospace (MDA) (the designer and builder of the robotic Canadarm device) and Intelsat to see if a satellite could be captured by a robot arm and refueled.[5]

Conclusion

Rocket propulsion has evolved a long way in the last 60 years. Launchers that are 98% reliable in terms of successful lift to orbit have now been achieved. New commercial space ventures are seeking to develop systems that could far more reliable still. Efforts to create launch systems that are dramatically more cost effective have eluded rocket developers to date. Launcher systems, however, have become increasingly dependable and certainly have increased their launch to orbit capabilities by orders of magnitude. Thus while launchers are not greatly more cost effective in terms of the cost of a lifting a kilogram to orbit, their expanded ability to lift larger satellites to orbit has allowed the satellites themselves to become more capable and cost effective due to economies of scale. More and more countries and companies have developed launch capabilities. The most dynamic new element in this regard is the Falcon class launchers developed by the Space eXploration Technologies Corporation known as SpaceX that is developing not only lower cost launch capabilities for the orbiting of application satellites, but perhaps ultimately seeking to develop a commercial option to lift astronauts to the International Space Station (ISS). Today as indicated in Appendices 1 and 2 there are rich options in terms of launch

sites and launch systems to support the launch operations of commercial, governmental and military application satellite systems.

Research continues to develop better launch capabilities, new technologies to lift satellites to earth orbit, new capabilities to maintain satellites in orbit, carry out extended station-keeping, and even in time possibly to allow the re-fuelling of satellites in orbit. Ultimately new capabilities to place satellites in orbit that are even more reliable and cost effective will evolve. There are yet other challenges to be faced such as dealing with orbital debris and new launch and robotic capabilities may possibly be able to address these problems as well.

Notes

1. Spaceflight: rockets and missiles. http://centenialofflight.gov/essay/SPACEFLIGHT/solids/SP13.htm
2. Soyuz launch complex in Kourou Guyana. www.russianspaceweb.com/kourou_els.html
3. The NASA space launch system. http://www.nasa.gov/exploration/systems/sls/sls1.html
4. Falcon launch vehicle. http://www.spacex.com/falcon1.php
5. "MDA in-provide in-orbit operations and maintenance support". http://sm.mdacorporation.com/

References

R.A. Braeuni, Basics of spaceflight: rocket propellants (2008). http://www.braeunig.us/space/propel.htm

A. Lindemoyer, COTS status: NASA advisory council (2011). http://www.nasa.gov/pdf/580727main_4%20-%20Lindenmoyer%20COTS%20Status_508.pdf

Michael Lennick, Launch Vehicles Pocket Space Guide: Heritage of the Space Race (Mar 1, 2006) (Pocket Space Guides) New York

J.N. Pelton, Global Talk (Sitjhoff and Noordhoff International, Alphen aan den Rijn, 1981) pp. 13–20

Travis S. Taylor, Introduction to Rocket Science and Engineering (2007) CRC Books, New York

Orbital Debris and Sustainability of Space Operations

46

Heiner Klinkrad

Contents

Introduction 1146
Space Debris and Their Effect on Space Applications 1148
Conclusion 1172
Cross-References 1173
Notes 1174
References 1174

Abstract

The orbital particle environment around the Earth is dominated by man-made space objects, except for a limited particle size regime below 1 mm, where meteoroids provide a significant contribution, or may even prevail in some orbit regions. The mass of man-made objects in Earth orbits is on the order of 6,300 t, of which more than 99% is concentrated in trackable, cataloged objects larger than typically 10 cm. The mass of meteoroids within the regime of Earth orbits is only on the order of 2–3 t, with most probable sizes around 200 μm. As a consequence of their size spectrum and associated mass man-made space objects, in contrast with meteoroids, represent a considerable risk potential for space assets in Earth orbits. To assess related risk levels a good understanding of the space debris environment is essential, both at catalog sizes and sub-catalog sizes. The derivation process and the key elements of today's debris environment models will be outlined, and results in terms of spatial densities and impact flux levels will be sketched for those orbit regions that are most relevant for space applications.

To cope with the existing space debris environment spacecraft can actively mitigate the risk of collisions with large-size, trackable space objects through

H. Klinkrad
European Space Agency ESA/ESOC, Robert-Bosch-Str. 5, 64293, Darmstadt, Germany
e-mail: heiner.klinkrad@esa.int

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_77, © Springer Science+Business Media New York 2013

evasive maneuvers. Alternatively, or in addition, the risk of mission-critical impacts by non-trackable objects can be reduced through shielding, in combination with protective arrangements of critical spacecraft subsystems. With a view on the future debris environment international consensus has been reached on a core set of space debris mitigation measures. These measures, which will be explained in more detail hereafter, are suited to reduce the debris growth rate. However, even if they are rigorously applied they are found to be inadequate to stabilize the debris environment. Long-term debris environment projections indicate that even a complete halt of launch activities cannot prevent the onset of a collisional run-away situation in some LEO altitude regimes. The only way of controlling this progressive increase of catastrophic collisions is through space debris environment remediation, with active mass removal, focused on retired spacecraft and spent orbital stages.

Keywords

Collision avoidance • Collision risk assessment • Debris collision flux • Debris environment models • Debris environment projections • Debris environment remediation • Debris mitigation • Evasive maneuvers • Impact protection • Inter-Agency Space Debris Coordination Committee (IADC) • Orbital debris • Space debris • Sustainability of space activities • US Space Surveillance Network (SSN)

Introduction

More than half a century of space flight activities since the launch of Sputnik-1, in 1957, has generated a significant man-made particle environment in Earth orbits that is referred to as "space debris." According to a definition by the Inter-Agency Space Debris Coordination Committee (IADC) "space debris are all man-made objects including fragments and elements thereof, in Earth orbit or reentering the atmosphere, that are non-functional." This sizeable population of space debris must be considered in the payload and mission designs to ensure successful space operations with an acceptable, low risk of losing or degrading a mission, or of suffering casualties during human space flight. Likewise, payloads and orbital stages must be designed, operated, and disposed of such that they do not further deteriorate the space debris environment, or pose an unacceptable risk to the ground population or air traffic during reentries.

Throughout this chapter, a snapshot of the orbital population of space objects close to 2010 (±1 year) will serve as a reference. The orbital debris environment in 2010 is the product of more than 4,700 launches and 245 on-orbit breakups that led to about 16,000 objects which are accessible through the unclassified catalog of the US Space Surveillance Network (SSN) (see Fig. 46.1). Another 6,000 objects were systematically tracked, but were either classified, or they were not yet correlated with a launch or deployment event. All SSN catalog objects combined represent some 6,300 t of on-orbit mass. Several tens of tons of further material from different sources are expected to exist at sub-catalog sizes, below diameters of 10 cm. Only 6% of the catalog entries

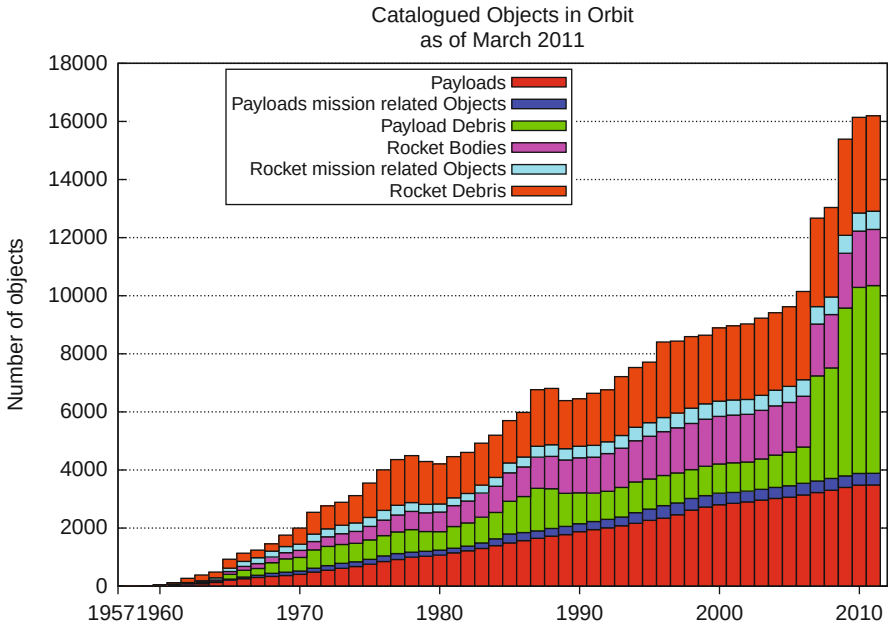


Fig. 46.1 Historic evolution of the US SSN catalog of trackable space objects

are operational spacecraft (slightly more than 1,000), while 30% are nonfunctional but intact objects, and 64% are fragments, mainly resulting from explosions, but also from recent in-orbit collisions. 77% of the catalog objects are in low Earth orbits (LEO), 6% are in or near-geostationary orbits (GEO), and 17% are in highly eccentric orbits (HEO), medium Earth orbits (MEO), or other orbit classes. Since 2007 the SSN catalog has experienced two significant step increases: on January 11, 2007, the Chinese Feng Yun 1 C satellite was intercepted in an ASAT (Anti-Satellite) test, generating 2,900 catalog objects of which 2,800 were still in orbit 4 years later; and on February 10, 2009, the first accidental hypervelocity collision between two intact catalog objects (Iridium 33 and Cosmos 2,251) generated some 1,800 cataloged fragments in two separate clouds, of which 1,700 were still in orbit 2 years later. Both of these events have produced a long-lasting increase in spatial object densities, and hence in collision risk, at altitudes between 750 and 900 km.

The risk of collision-induced catastrophic fragmentations or mission-terminating impacts is the highest in the low Earth orbit (LEO) regime. It exceeds the risks in other orbit regions, including the geostationary orbit (GEO) by at least three orders of magnitude. As a consequence, the following analysis will concentrate on the collision risk levels for the International Space Station (ISS), as an example of a manned LEO platform, and on the collision risk levels for a typical remote sensing spacecraft, on a sun-synchronous orbit, as an example of a robotic LEO platform. The concepts of active protection (shielding) and passive protection measures (avoidance maneuvers), and their effectiveness as a function of debris size will be discussed as possible risk mitigation measures for the specific debris environment of given operational orbits at

360 km altitude and 51.6° inclination for the ISS, and at 780 km altitude and 98.5° inclination for an Earth observation mission.

Roughly 40% of the entire mass in orbit is concentrated in the LEO regime, within just 0.3% of the operationally used volume from LEO up to super-GEO altitudes. Debris risk mitigation through collision avoidance, passive protection, and end-of-mission disposal turns out to be a necessary but insufficient condition to maintain an acceptable space debris environment. Long-term projections indicate that even drastic mitigation measures, such as an immediate, complete halt of launch and release activities, will not result in a stable LEO debris environment (see Liou and Johnson 2008a, b; Bastida et al. 2009; Liou 2011). Catastrophic collisions between existing space hardware of sufficient size will within a few decades start to dominate the debris population sources, and lead to a net increase of the space debris population, also at sizes which may cause further catastrophic collisions. A self-contained collisional cascading process in the LEO regime may hence ultimately lead to a run-away situation (the so-called Kessler syndrome), with no further possibility of control through human intervention. The only way to prevent the onset of collisional cascading is an active removal of mass from orbit. Since most of the LEO mass is concentrated in decommissioned though intact satellites and orbital stages, an effective mass removal operation must focus on this class of objects, and on preferred orbit classes for their mission deployments. Several operational concepts and physical principles have been explored to enable a space debris environment remediation through mass removal. Some of the most promising of these concepts suggest the use of electrodynamic or momentum-exchange tethers, space tugs, the deployment of drag augmentation devices or solar sails, or the release of large momentum-retarding surfaces. Such options will be reviewed hereafter.

Apart from the systematically trackable catalog population of space objects, there is a much larger population of sub-catalog debris objects than can disable or seriously degrade a space mission. The related objects can only be observed in a statistical manner, by means of research radars, telescopes, and in situ detectors. Based on orbital and physical characteristics of the observed debris, and based on ground test benchmark data, debris environment models can be established that compose an image of the current environment from a replicate of historic launch, release, and breakup events. One of the leading debris models, ESA's MASTER software (Meteoroid and Space Debris Terrestrial Environment Reference, see Oswald et al. 2005; Flegel et al. 2010), will be used in the following risk assessments. An in-depth technical discussion of underlying theories and analysis techniques is provided in (Klinkrad 2006) and will not be repeated here.

Space Debris and Their Effect on Space Applications

The resident mass in operationally used orbit regions around the Earth is to 99.95% dominated by man-made space debris, totaling approximately 6,300 metric tons in the year 2010. Only a few tons of additional material within the same reference volume originate from natural meteorites, with most probable sizes of about

Table 46.1 Orbital distribution of US Space Surveillance Network catalog objects in May 2010 (*MRO* mission-related objects, associated with payloads and rocket bodies, *LEO* low earth orbits, *MEO* medium earth orbits, *SSO* semi-synchronous orbits, *GEO* near-geostationary orbits, *GTO* GEO transfer orbits, *HEO* highly eccentric orbits, *HAO* high-altitude super-GEO orbits, *UDF* undefined)

	Payloads			Rocket bodies			
	Intact	MRO	Debris	Intact	MRO	Debris	Total
LEO	1,827	156	6,171	823	470	2,528	11,975
MEO	61	53	42	32	7	7	202
SSO	160	0	1	40	2	0	203
GEO	812	5	7	180	2	6	1,012
GTO	18	1	0	105	29	96	249
HEO	108	29	227	253	56	481	1,154
HAO	169	27	88	282	34	241	841
UDF	4	0	1	12	2	0	19
Total	3,159	271	6,537	1,727	602	3,359	15,655

200 μm. As a consequence, space debris dominate the risk for operational space missions and will be in the focus of the following discussion.

Within 1 decade after the first space launch the annual launch rates reached a level of more than 120 at the end of the 1960s. As a consequence of reduced Russian space activities at the end of the 1980s annual launch rates today have reached a stable level of about 70. By mid 2010 there were some 4,700 successful launches (out of 5,050 launch attempts) that deployed 3,159 payloads, 1,727 rocket stages, and another 873 mission-related objects (MRO) into orbit (see Table 46.1). These intact objects account for most of the in-orbit mass of about 6,300 t. However, they only account for 37% of the space object population that can be routinely tracked by operational surveillance networks. Out of 15,655 objects of the US Space Surveillance Network (SSN) catalog in May 2010, the dominant space debris population contributed 9,896 trackable objects (67%). With 11,975 objects (77%) the vast majority of the SSN catalog resides in low Earth orbits (LEO), below altitudes of 2,000 km, another 1,012 objects (6%) are in the vicinity of the geostationary ring (GEO), at altitudes of 35,786 $\pm$ 2,800 km and inclinations of $0° \leq i \leq 15°$, and the remaining 2,668 objects (17%) are distributed across medium Earth orbits (MEO), semi-synchronous orbits of navigation constellations (SSO), GEO transfer orbits (GTO), highly eccentric orbits (HEO), and high-altitude orbits beyond the GEO regime (HAO). Table 46.2 shows individual contributions to the SSN catalog according to launch nation and Fig. 46.1 shows the historic evolution of the catalog population.

The US Space Surveillance Network has a cataloging size threshold that ranges from about 10 cm in the LEO regime to about 1 m in the GEO ring. Related routine observations are performed by a network of radars or LEO and low MEO altitudes and by globally distributed electro-optical telescopes for the remaining part of MEO up to GEO altitudes. For the dominant LEO catalog population Fig. 46.2 shows the altitude distribution of objects, with a main maximum close to 800 km and a secondary maximum slightly below 1,500 km. Since the vast majority of catalog objects are on

Table 46.2 Status of the US space surveillance network catalog in May 2010 according to the NASA satellite situation report

		Objects in orbit			Objects decayed		
Launch Nation/organization	ISO	PL's	Debris	Total	PL's	Debris	Total
Algeria	DZ	1	0	1	0	0	0
Argentina	AR	9	0	9	2	0	2
Australia	AU	12	0	12	2	0	2
Bermuda	BM	6	0	6	0	0	0
Brazil	BR	13	0	13	0	0	0
Canada	CA	30	8	38	1	3	4
Chile	CL	1	0	1	0	0	0
China	CN	87	3,287	3,369	47	481	528
Colombia	CO	1	0	1	0	0	0
Czech Republic	CZ	1	0	1	0	0	0
Czechoslovakia	CS	4	0	4	1	0	1
Denmark	DK	4	0	4	0	0	0
Egypt	EG	3	0	3	0	0	0
Eumetsat	EUM	6	4	10	0	0	0
ESA	ESA	52	3	55	6	1	7
ESRO	ESR	0	0	0	7	0	7
Eutelsat	EUT	21	0	21	0	0	0
France	FR	57	466	523	9	671	680
France/Germany	F/D	2	0	2	0	0	0
Germany	DE	37	1	38	11	0	11
Greece	GR	1	0	1	0	0	0
Hongkong	HK	5	0	5	0	0	0
India	IN	40	134	174	9	270	279
Indonesia	ID	13	0	13	1	0	1
Inmarsat	INM	9	0	9	0	0	0
ISS Space Station	ISS	4	0	4	0	48	48
Intelsat	ITS	59	0	59	2	0	2
Iran	IR	1	0	1	1	1	2
Israel	IL	9	0	9	3	5	8
Italy	IT	17	1	18	9	1	10
Japan	JP	124	76	200	27	181	208
Kazakhstan	KZ	1	0	1	0	0	0
Korea, Rep. of	KR	10	0	10	0	1	1
Luxemburg	LU	17	0	17	1	0	1
Malaysia	MY	6	0	6	0	0	0
Mauritius	MU	1	0	1	0	0	0
Mexico	MX	7	0	7	0	0	0
Morocco	MA	1	0	1	0	0	0
Netherlands	NL	9	0	9	1	0	1
Nigeria	NG	2	0	2	0	0	0
NATO	NAT	8	0	8	0	0	0

(*continued*)

Table 46.2 (continued)

		Objects in orbit			Objects decayed		
Launch Nation/organization	ISO	PL's	Debris	Total	PL's	Debris	Total
Norway	NO	5	0	5	0	0	0
Pakistan	PK	1	0	1	1	0	1
Philippines	PH	2	0	2	0	0	0
Portugal	PT	1	0	1	0	0	0
Russian Fed.	RU	341	2,175	2,516	233	1,628	1,861
Saudi Arabia	SA	21	0	21	1	0	1
Singapore	SG	1	0	1	0	0	0
South Africa	ZA	2	0	2	0	0	0
Spain	ES	13	0	13	1	0	1
Sweden	SE	11	0	11	0	0	0
Switzerland	CH	1	0	1	0	0	0
Taiwan	TW	8	0	8	0	0	0
Thailand	TH	7	0	7	0	0	0
Turkey	TR	6	0	6	0	0	0
Ukraine	UA	0	0	0	2	0	2
USSR	SU	1,084	2,296	3,380	1,727	10,082	11,809
United Emirates	AE	4	0	4	0	0	0
United Kingdom	GB	35	1	36	9	1	10
United States	US	1,187	3,785	4,972	774	4,516	5,290
Venezuela	VE	1	0	1	0	0	0
Vietnam	VN	1	0	1	0	0	0
Column Totals		3,423	12,259	15,655	2,888	17,890	20,778
Overall Total	36,433			15,655			20,778

near-circular orbits (with more than 50% of the eccentricities smaller than 0.01), the depicted, resident probability-weighted, mean altitude distribution is very similar to the actual perigee and apogee altitude distributions. Figure 46.3 shows that the inclination distribution of LEO orbits is driven by mission and launch constraints, with distinct, preferred inclination bands around 65°, 75°, 82°, 90°, and 98°. Figure 46.4 illustrates how the altitude and inclination distributions of catalog objects are correlated.

Space debris caused by fragmentation events are the most important source of catalog objects, with a contribution of 63% to the trackable population in mid 2010. In the course of space history some 245 on-orbit fragmentation events were inferred from the detection of new objects and from the correlation of their determined orbits with a common source. The dominant breakup causes are believed to have been deliberate explosions or collisions (dominated by an ASAT test that destroyed Feng Yun 1C in January 2007), propulsion-related explosions, battery explosions, and four known accidental collisions (the Cosmos 1934 spacecraft with a Cosmos 926 MRO in December 1991, the Cerise spacecraft with an Ariane H-10 fragment in July 1996, a Thor stage with a CZ-4B stage fragment in January 2005, and Cosmos 2251 with Iridium 33 in February 2009). About a third of all breakups were of an unknown cause. With the exception of two known GEO explosion events (an Ekran-2 satellite

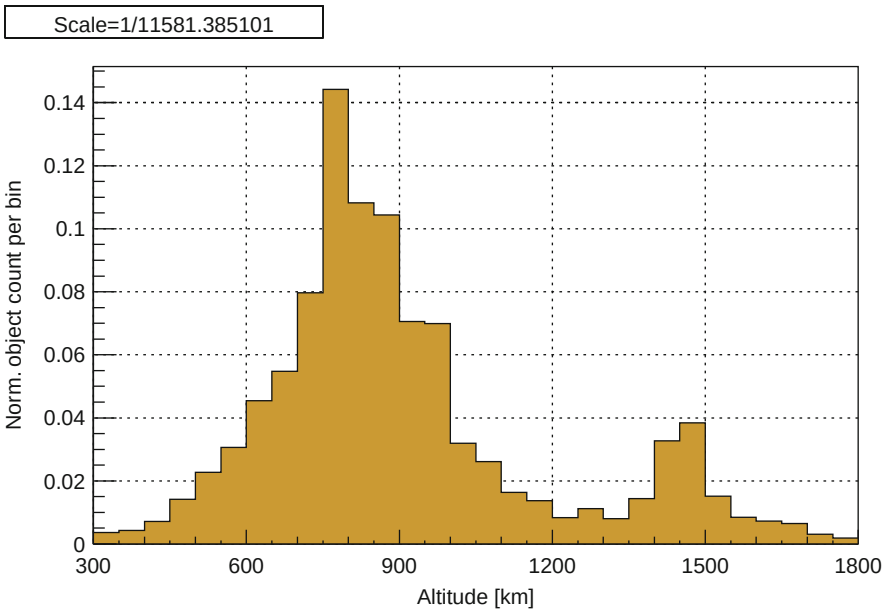


Fig. 46.2 Altitude distribution of catalog-size objects (>10 cm) in low earth orbit (LEO) in 2010. The normalized count is in fractions per 50 km altitude bin for a total of 11,581 objects

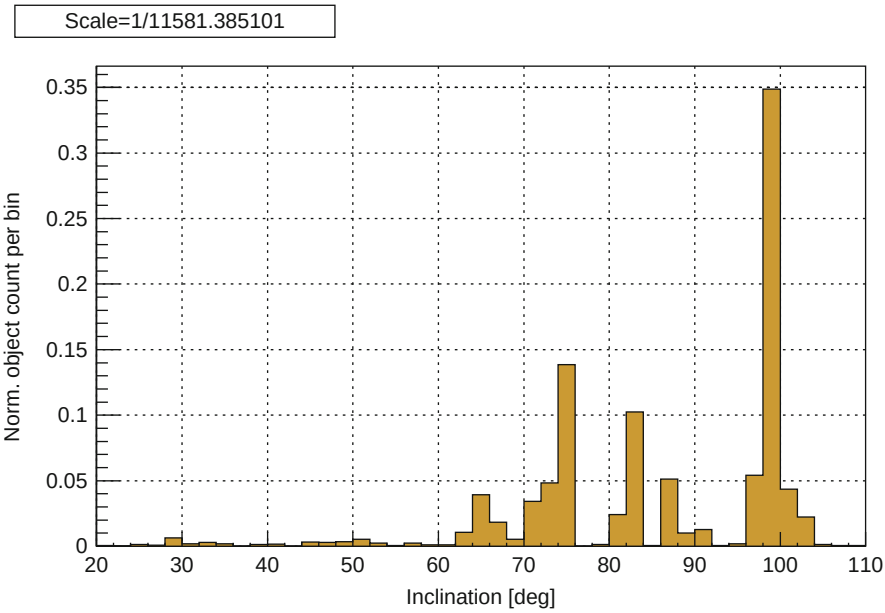


Fig. 46.3 Inclination distribution of catalog-size objects (>10 cm) in low earth orbit (LEO) in 2010. The normalized count is in fractions per 2° orbit inclination bin for a total of 11,581 objects

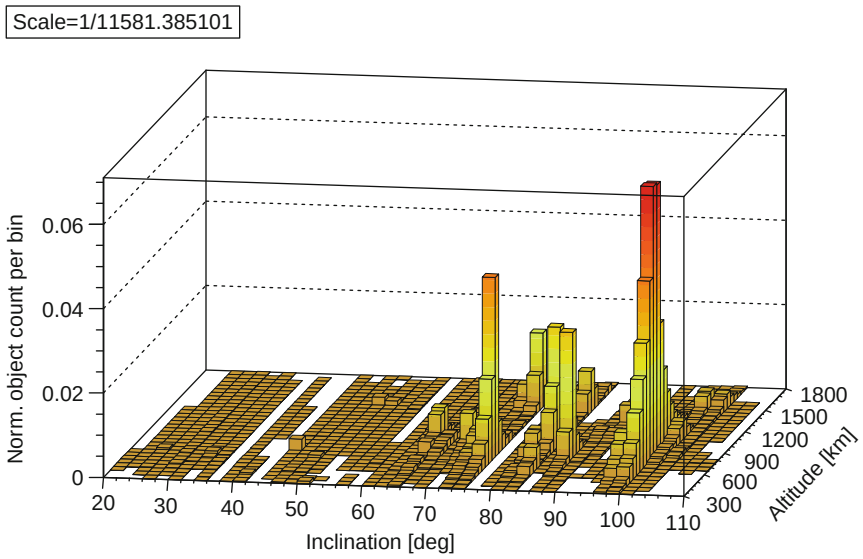


Fig. 46.4 Inclination and altitude distribution of catalog-size objects (>10 cm) in low earth orbit (LEO) in 2010. The normalized count is in fractions per bin of 2° × 50 km for a total of 11,581 objects

on June 22, 1978, and a Titan III-C Transtage on February 8, 1994), all known fragmentations occurred on orbits passing through LEO altitudes, with about 80% of the orbits entirely within LEO, and with 17% on highly eccentric trajectories passing through LEO (see Klinkrad 2006). Table 46.3 shows a list of the ten most significant in-orbit breakups, sorted by the number of cataloged fragments. Eight of these top ten events occurred on orbit inclinations of 90° ± 10°, mainly at altitudes of 800 ± 50 km. Since the orbit inclination is a very stable parameter, directly linked to the orbit momentum and only marginally affected by orbit perturbations, it strongly governs the latitude distribution of resulting spatial object densities. Figure 46.5 indicates that the highest concentration of catalog-size objects is at high latitudes δ, where $\delta \approx i$, with i being the inclinations of breakup orbits. As a consequence, catastrophic collisions between catalog objects are most likely at high latitudes in densely populated altitude bands. Fragmentation debris from in-orbit explosions and collisions dominate the space debris population down to the cm-size regime (see Table 46.4). The most significant breakup-related, relative increase of the catalog population occurred in 1961, when the first accidental explosion in space of an Ablestar injection stage more than tripled the catalog population from 110 to almost 400. The most significant absolute growth of the catalog so far occurred in January 2007, when the Feng Yun 1C kinetic ASAT test produced some 3,000 trackable fragments (+30%), and in February 2009, when the accidental collision between Cosmos 2251 and Iridium 33 generated another 2,800 fragments (+22%).

At sub-catalog sizes residues from solid rocket motor (SRM) firings become important. The number of solid rocket motor firings up to 2010 was on the order of 1,100, with peak rates of up to 47 events per year, and a mean annual rate of 23.5.

Table 46.3 On-orbit breakup events with highest counts of cataloged fragments, and their contributions in 2010

Object name	Launch date Event date	Max. count Curr. count	COSPAR H_p [km]	 H_a [km]	Sat. no. i [deg]	Assessed cause Object type
Feng Yun 1 C	1999/05/10	2,944	1999-025A		25,730	Deliberate
	2007/01/11	2,860	843	863	98.64	Payload
Cosmos 2251	1993/06/16	1,300	1993-036A		22,675	Collision
	2009/02/10	1,243	843	863	98.64	Payload
Pegasus 4th stage	1994/04/19	714	1994-029B		23,106	Propulsion
	1996/06/03	64	584	819	81.97	Rocket body
Iridium 33	1997/09/14	532	1997-051C		24,946	Collision
	2009/02/10	505	776	791	86.39	Payload
Cosmos 2421	2006/06/25	510	2006-026A		29,247	Unknown
	2008/03/14	40	389	415	65.04	Payload
Ariane 1 3rd stage	1986/02/22	495	1986-019C		16,615	Propulsion
	1986/11/13	36	803	833	98.61	Rocket body
OV 1/LCS 2	1965/10/15	474	1995-082B		1,640	Propulsion
	1965/10/15	37	658	761	32.17	Payload
CZ 4B 4th stage	1999/10/14	427	1999-057C		25,942	Unknown
	2000/03/11	266	727	744	98.54	Rocket body
Thorad Agena D 2nd stage	1970/04/08	374	1970-025 C		4,367	Unknown
	1970/10/17	249	1,013	1,049	99.62	Rocket body
PSLV 4th stage	2001/10/22	370	2001-049D		26,960	Unknown
	2001/12/19	117	550	674	97.90	Rocket body

The injection orbits where SRMs were applied are to 80% associated with US missions. The size of the solid motors, in terms of propellant capacity, covers a wide range. The most frequently used SRMs are the Star 37 motors, with a propellant mass of 1,067 kg, used for instance as final stage of Delta launchers to deploy GPS/Navstar payloads, the Payload Assist Module PAM-D, with 2,011 kg, also used as Delta final stage, for instance, for GTO injections, and the Inert Upper Stage (IUS), deployed from Titan IV or Space Shuttle, for instance to inject payloads into GTO with a first stage of 9,709 kg, and subsequently deliver the payload into a circular GEO by a second stage of 2,722 kg. Another powerful SRM engine, HS-601 with 4,267 kg, is used by Long March LM-2E launchers both for LEO and GTO payload injections.

SRM combustion residues are mainly composed of aluminum oxide and residues of motor liner material. Aluminum powder is added to most solid fuels, typically with a mass fraction of 18%, to stabilize the combustion process and improve the motor performance. It is assumed that about 99% thereof is continuously ejected with the exhaust stream during the main thrust phase in the form of Al_2O_3 dust of diameters largely within 1 μm $\leq$ d $\leq$ 50 μm. Due to design constraints many solid motors have nozzles protruding into the burn chamber, causing cavities around the nozzle throats. During the burn phase, trapped Al_2O_3, molten Aluminum droplets, and parts of released thermal insulation liner material can cumulate in this pool, and form slag particles which can grow to sizes of typically 0.1 mm $\leq$ d $\leq$ 30 mm.

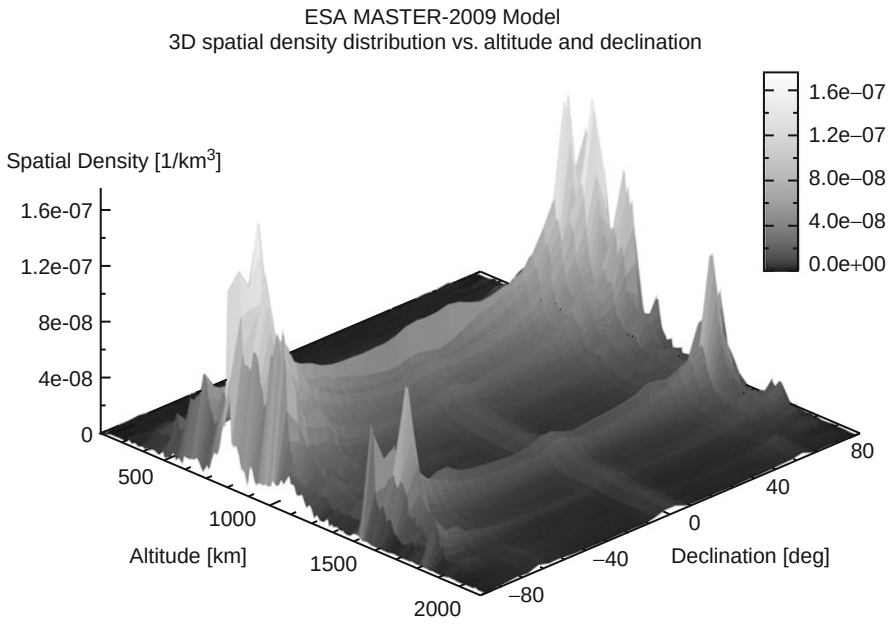


Fig. 46.5 Spatial density distribution of catalog-size objects (>10 cm) in low earth orbit (LEO) in 2010, as a function of altitude and declination

Table 46.4 Sources and their contributions to ESA's MASTER 2009 space debris model (Flegel et al. 2010) in different size regimes for epoch May 1, 2009

Diameter	>1 μm	>10 μm	>100 μm	>1 mm	>1 cm	>10 cm	>1 m
LMRO	45,919	45,919	45,919	31,138	5,827	5,814	4,174
Expl.	5.64e + 9	4.12e + 9	3.84e + 8	1.53e + 7	433,466	14,719	432
Coll.	3.58e + 9	1.13e + 9	1.18e + 8	4.46e + 6	92,677	2,927	63
MLI	22,241	22,241	22,241	22,241	15,790	5,750	773
NaK	30,162	30,162	30,162	30,162	18,410	–	–
SRM slag	4.98e + 12	4.98e + 12	2.33e + 12	1.39e + 8	177,914	–	–
SRM dust	6.07e + 14	1.18e + 13	–	–	–	–	–
Paint	1.97e + 12	1.62e + 12	2.28e + 11	–	–	–	–
Ejecta	8.61e + 13	2.70e + 13	1.08e + 12	8.00e + 6	–	–	–
Total	6.99e + 14	4.53e + 13	3.64e + 12	1.66e + 8	744,084	29,210	5,442

These slag particles are released at the end of the main thrust phase, as the internal motor pressure decreases. It can be assumed that during more than 1,100 SRM firings more than 1,000 t of propellant were released into space of which approximately 320 t were Al_2O_3 dust particles, and 4 t were slag particles formed of Al_2O_3, metallic Aluminum, and motor liner material. Due to orbital perturbations and their different effects on μm-size dust and cm-size slag, merely 1 t of Al_2O_3 dust and 3 t of SRM slag particles are believed to be still on orbit. Apart from more than 1,000

orbit insertion burns, there were also several hundred SRM burns to deorbit objects in a controlled fashion. These deorbit burns were almost exclusively performed for Russian reconnaissance satellites at very low altitudes, and the resulting SRM combustion products had a correspondingly low orbit lifetime. However, some in situ measurements (mainly from returned space hardware) show temporal increases in small-particle impact rates due to these events. At sizes of 1 μm $\leq$ d $\leq$ 1 cm SRM combustion residues dominate the space debris environment (see Table 46.4).

Apart from intact objects, fragmentation debris, and SRM residues, there are other contributors to the space debris population: (1) sodium-potassium (NaK) coolant released from 16 Russian RORSATs as they ejected their reactor cores in the 1980s, (2) multilayer insulation (MLI) material that is unintentionally released by spacecraft or rocket stages, (3) ejecta material that is released by small-particle impacts on surfaces of spacecraft and orbital stages, and (4) degradation products that are released by aging surfaces of spacecraft and orbital stages. The debris mass contribution from these sources is much less than 1% of the overall on orbit mass, and they are either too small in numbers (NaK, MLI), or too small in size (surface ejecta and degradation products) to constitute a significant risk for space missions.

Today's population of trackable and non-trackable objects can be reproduced by space debris environment models, such as ESA's MASTER-2009 model (Flegel et al. 2010). Such models consider historic launch and release events, known in-orbit fragmentations, known solid rocket motor firing events, intentional releases of NaK coolant liquid from Buk reactors of Russian RORSAT satellites, unintentional releases of surface degradation products (MLI and paint flakes), and the generation of ejecta and spall by surface impacts. Table 46.4 lists the resulting debris sources, and their contributions to the MASTER-2009 population at the reference epoch of May 2009 for the applicable size regime larger than 1 μm. From the risk point of view, the almost 170 million particles larger than 1 mm, at typical LEO collision velocities of 10–14 km/s, can disable sensitive satellite subsystems, the more than 700,000 particles larger than 1 cm can render a spacecraft dysfunctional, and the almost 30,000 objects larger than 10 cm are likely to cause a catastrophic breakup of a satellite or orbital stage.

Figure 46.6 shows the altitude distribution of MASTER-2009 objects larger than 10 cm in terms of resulting spatial densities (in objects/km^3). The contributing debris sources at these sizes are explosion and collision fragments, intact objects, and light-weight sheets of MLI. Highest concentrations are in the LEO regime, between 750 km and 900 km, with almost equal contributions from explosion fragments, collision fragments, and intact objects. In general, however, explosion fragments dominate the LEO and GEO regions, with GEO object concentrations about three orders of magnitude below the LEO maximum. When going to a 1 cm size threshold additional source terms come in, including NaK droplets and solid rocket motor slag, while launch and mission-related objects start playing a minor role. Figure 46.7 shows the individual contributions as a function of altitude. Reducing the size threshold further to 1 mm leads to the addition of ejecta particles, as shown in Fig. 46.8. With the decrease of the debris sizes from 10 cm to 1 mm the enveloping curve of spatial densities tends to flatten, due to an increasing share of

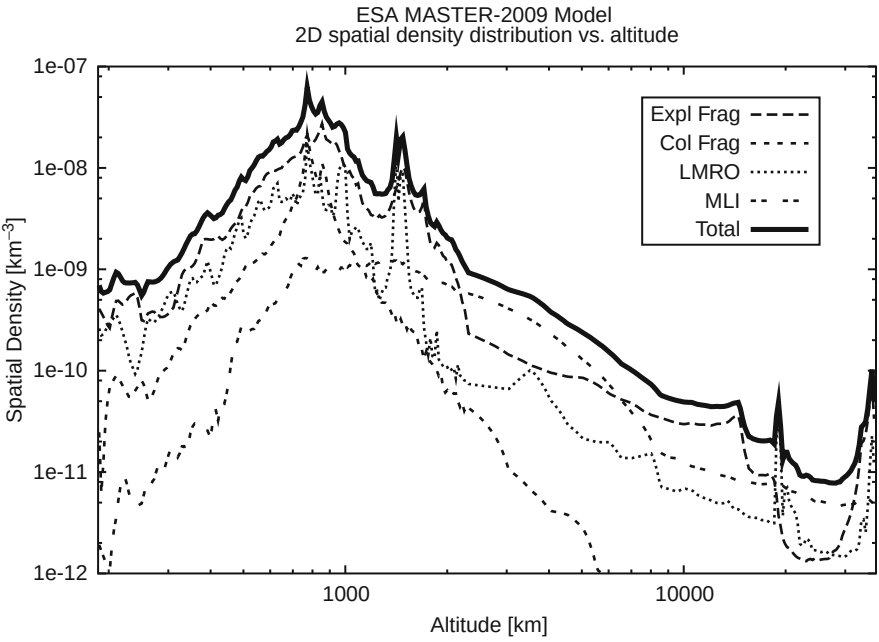


Fig. 46.6 Spatial density distribution of MASTER-2009 objects of d > 10 cm, in LEO to GEO altitudes, discriminated by sources (Flegel et al. 2010)

particles on eccentric orbits with a wider distribution over altitudes. As a consequence, the relative magnitude of the GEO concentration peak with respect to the LEO maximum reduces from three to less than two magnitudes. One cause of the increase of orbit eccentricities with decreasing object sizes lies in the area-to-mass ratio that drives solar radiation pressure and airdrag forces and is inversely proportional to the object diameter.

Spatial object densities are an essential input to debris collision risk assessments. The statistical behavior of the orbital debris population can be well represented by the laws of kinetic gas theory. Hence, the number of collisions c encountered by an object of collision cross-section A_c, moving through a stationary debris medium of uniform particle density D, at a constant relative velocity Δv, during a propagation time interval Δt is given by

$$c = \Delta v \, D \, A_c \, \Delta t \tag{46.1}$$

where $F = \Delta v \, D$ is the impact flux (in units of $\mathrm{m}^{-2}\,\mathrm{s}^{-1}$) and $\Phi = F \, \Delta t$ is the impact fluence (in units of m^{-2}). The collision probability follows a binomial law which can be well approximated by a Poisson distribution, generating the following probability $P_{i=n}$ of n impacts, and $P_{i=0}$ of no impact.

$$P_{i=n} = \frac{c^n}{n!} \exp(-c) \qquad \mapsto \qquad P_{i=0} = \exp(-c) \tag{46.2}$$

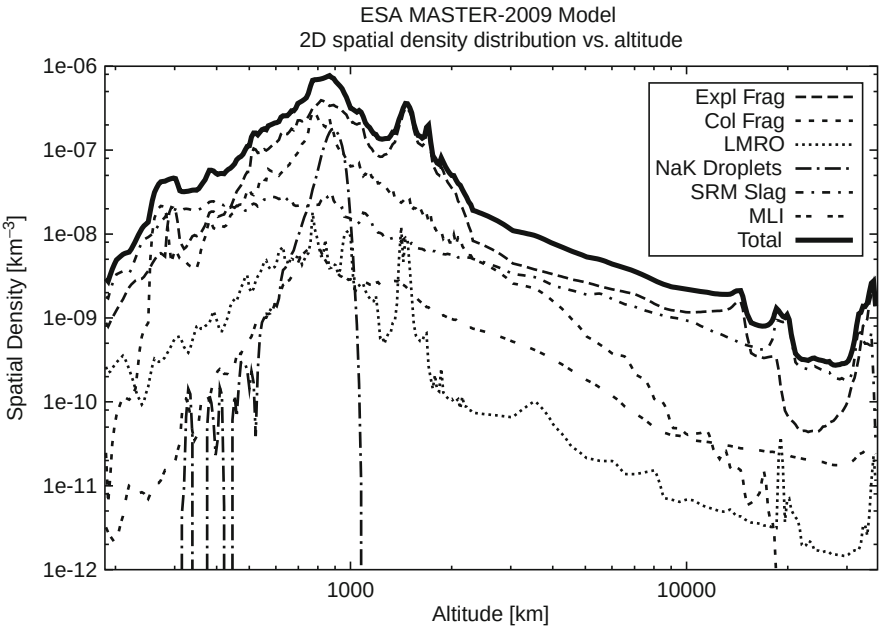


Fig. 46.7 Spatial density distribution of MASTER-2009 objects of d > 1 cm, in LEO to GEO altitudes, discriminated by sources (Flegel et al. 2010)

The probability of one or more impacts is hence the complement of no impact, given by

$$P_{i \geq n} = P = 1 - \exp(-c) \;\approx\; c \qquad \mapsto \qquad P \approx\; \Delta v\, D\, A_c\, \Delta t \tag{46.3}$$

The challenging part in the evaluation of this equation is the particle flux $F = \Delta v\, D$. In the MASTER-2009 model three-dimensional, time-dependent spatial object density distributions are established for a grid of spherical volume elements covering the entire Earth environment from LEO to GEO altitudes. Contributions from each member of the orbital debris population go into this distribution. For each of these objects, the velocity magnitude and direction is retained for each volume element passage. This information is later retrieved to determine relative impact velocities with respect to a target object passing through individual cells of the volume grid (see Klinkrad 2006). The resulting impact flux is then determined from a summation over all volume cells that are passed by the target object, with contributions from all debris objects that passed the individual cells.

When considering relative velocities between two objects on circular orbits at the same altitude, with the same orbital velocities v, but on different inclinations, then Eq. 46.4 yields the resulting collision velocity as a function of the impact azimuth A within the local horizontal plane (where $A = 0°$ denotes impacts from the flight direction).

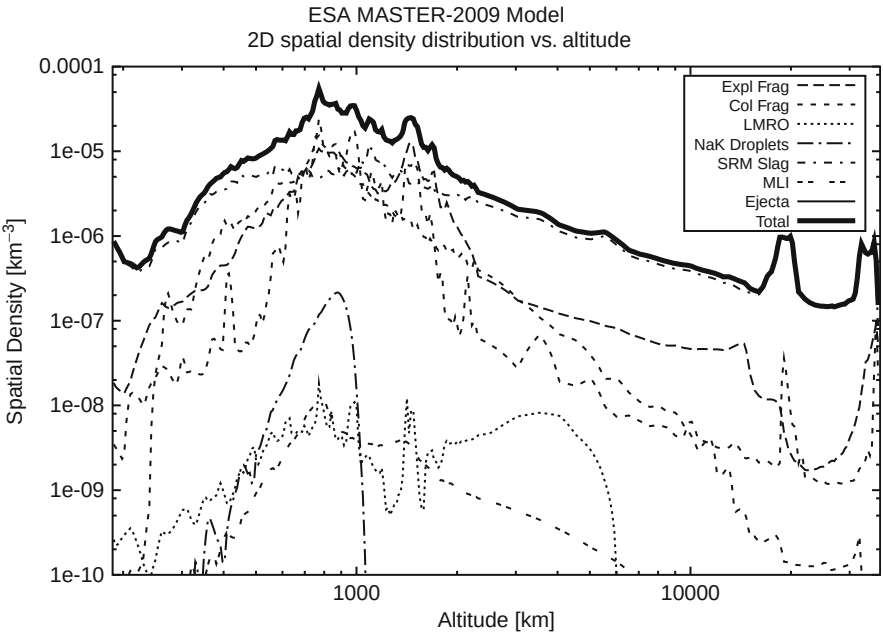


Fig. 46.8 Spatial density distribution of MASTER-2009 objects of d > 1 mm, in LEO to GEO altitudes, discriminated by sources (Flegel et al. 2010)

$$\Delta v \approx 2\ v\ \cos(A) \tag{46.4}$$

Since near-circular orbits are dominant for debris of critical sizes, Eq. 46.4 provide a good approximation of the correlation of impact velocity with impact geometry. It also states that the maximum relative velocity can be twice the orbit velocity, for an approach from the flight direction, and that the minimum relative velocity can be close to zero, for a sideways approach from ±90°. Impacts from the rear quadrants can only occur for impactors that travel on eccentric orbits, during their perigee passes. Likewise, impacts from 0° can only occur, if the impactor has an orbit with a "complementary inclination" of 180° minus the inclination of the target object. Only in that case can both objects be in the same orbit plane, on counter-rotating orbits, if their ascending orbit nodes are separated by 180°.

For typical target orbits defined in Table 46.5 the mean times between impacts by orbital debris of different sizes are listed in Table 46.6 for a common reference cross-section of 1 m^2, assuming a spherical target object. In accordance with spatial densities shown in Fig. 46.6–46.8, the highest collision risk for any of the selected sample orbits is encountered for ERS-2 on a sun-synchronous orbit of 774 × 789 km at an inclination of 98.5°. Apart from the debris concentration at this altitude, the collision frequency is also driven by the collision velocity (see Eq. 46.1). For ERS-2 it attains a most probable value of about 14 km/s, which is close to the maximum for two circular orbits at this altitude. Objects that could

Table 46.5 Sample orbits for analyzing space debris collision flux

	H_p [km]	H_a [km]	i [deg]	a [km]	e [-]	ω [deg]
ISS	356	364	51.6	6,738	0.000601	0
ERS-2	774	789	98.6	7,159	0.001096	90
Globalstar	1,399	1,401	52	7,778	0.0001	0
GPS	19,997	20,003	55	26,378	0.0001	0
GTO	560	35,786	7	24,551	0.717405	178
GEO	35,782	35,790	0.1	42,164	0.0001	0

Table 46.6 Mean time between impacts of a given debris size for a spherical target of 1 m^2 cross-section on sample orbits as defined in Table 46.5

Diameter	>0.1 mm	>1 mm	>1 cm	>10 cm
ISS	9.0 d	636 y	41,102 y	942,507 y
ERS	0.7 d	42.5 y	1,252 y	43,783 y
Globalstar	1.7 d	102 y	9,208 y	126,550 y
GPS	244.8 d	10,794 y	1.1e + 7 y	7.2e + 8 y
GTO	36.8 d	2,627 y	241,546 y	4.4e + 6 y
GEO	676.3 d	18,674 y	6.5e + 6 y	1.4e + 8 y

impact at such velocities are originating from the complementary inclination band close to 81.4° (=180°−98.6°, see Fig. 46.3 and 46.4). Since all major flux contributions are from inclinations $i \geq 65°$, the resulting collision velocities are mostly within 14 ± 2 km/s at impact azimuth angles $-30° \leq A \leq +30°$ (see Eq. 46.4), with particles mainly originating from breakup events for the size regime larger than 1 cm. The situation changes for the ISS orbit. Its lower altitude goes along with a reduction of the debris flux by about one order of magnitude, and its lower inclination of 51.6° results in a gap of complementary inclination bands at 180°−51.6° = 128.4°. The populated inclination bands only start at about 100°. As a consequence, there are no impacts from azimuth angles $-15° \leq A \leq +15°$, and most probable collision velocities are at 10 ± 1 km/s, resulting in approximately 50% of the impact energy as compared to ERS-2. In contrast to ERS-2, slag residues from SRM firings are dominating the 1 cm debris population for ISS. They mostly reside on highly eccentric orbits of low inclinations, with perigee velocities that allow low-velocity impacts also from rear quadrants of ISS azimuth angles. When looking at a typical geostationary target orbit, then the spatial density of the debris environment as compared to the LEO peak drops by about three orders of magnitude for the 10 cm population, and by about two orders of magnitude for the 1 cm population. For the GEO orbit velocity of about 3 km/s the predicted collision velocities are in the range of $0 \leq v \leq 1$ km/s, with a most probable value of 0.8 km/s, caused by old GEO objects that reached a maximum inclination excursion of 15° due long-periodic orbit perturbations with a period of 53 years. Due to the low relative velocities, the impact azimuth angles are mostly at ±80°. There are minor flux contributions from objects on GEO transfer orbits (GTO) and on 12 h Molniya orbits. They have apogee velocities of about 1.5 km/s, causing frontal impacts on the faster GEO objects.

There are different ways to mitigate the risk and/or consequences of a collision of an operational spacecraft with a space debris object. For large-size catalog objects the concept of conjunction event analysis and collision avoidance can be pursued. For sub-catalog debris that cannot be tracked, passive protection measures can be taken.

To avoid catastrophic collisions with catalog-size objects of $d \geq 10$ cm, the ISS operators perform a conjunction event screening on the basis of the US Space Surveillance Network (SSN) catalog. This screening is performed at least three times a day, for 72 h ahead, in five steps:

1. TLE-based[1] identification of approaches that fall within a 60 km radius, centered on the ISS
2. Use of more accurate, osculating orbital elements, if the approach falls within ± 10 km $\times \pm 40$ km $\times \pm 40$ km (radial $\times$ along-track $\times$ out-of-plane)
3. Consideration of orbit uncertainties, if the approach falls within ± 2 km $\times \pm 25$ km $\times \pm 25$ km
4. Determination of collision probabilities, if approach falls within ± 0.75 km $\times \pm 25$ km $\times \pm 25$ km
5. Decision on an evasive maneuver, if an accepted risk threshold is exceeded (e.g., 1 in 10,000)

In the first 4.5 years of operation the ISS performed seven debris avoidance maneuvers, with three of them executed by the visiting Space Shuttle. Due to improved procedures, based on more reliable orbit data, the subsequent avoidance maneuver was only 5.5 years later, executed by the attached ATV-1 on August 27, 2008, to avoid a fragment of Cosmos 2421. This fragment was one of 500 cataloged objects generated during three main breakup events in early 2008, just 60 km above the ISS altitude (see Johnson 2009).

As is done by NASA for the ISS, ESA maintains a conjunction event analysis service, e.g., for their Envisat and ERS-2 satellite. Once a day the entire TLE catalog of the US SSN is screened for close conjunctions with the accurately known Envisat and ERS-2 orbits for 7 days ahead. If the predicted collision probability exceeds a level of 1 in 3,000, then more precise orbit data are obtained for the conjunctor object through the processing of radar data from tasked observations. In most cases, the more accurately known conjunctor orbit with its much reduced error dispersion leads to a maneuver suppression, even if the flyby geometry is unchanged. If, however, the collision probability remains at a level above 1 in 1,000, then a collision avoidance maneuver is initiated by the relevant project team.

Envisat, launched in 2002, had to perform five avoidance maneuvers up to December 2009. Due to the Chinese FengYun 1C ASAT test in January 2007, and as a result of the collision between Cosmos 2251 and Iridium 33 in February 2009, the debris environment at the Envisat and ERS-2 orbit altitude significantly deteriorated. As a consequence, the overall avoidance maneuver frequency in the year 2010 increased to 9 (4 each for Envisat and ERS-2, and 1 for Cryosat-2). The risk of catastrophic collisions of Envisat with a 10 cm fragment from the FengYun 1 C and Cosmos 2251/Iridium 33 breakup events alone increased by +58% as compared to the rest of the US SSN catalog. The risk of a mission-terminating

impact by a 1 cm class debris object even grew by +86%, as compared with a modeled space debris population prior to these events.

To protect against non-trackable debris and meteoroids the ISS has its manned modules covered by stuffed Whipple shields. For ESA's Columbus module, for instance, they consist of a 2.5 mm bumper and a 4.8 mm back wall, separated by an 11 cm stand-off distance. Between the bumper and the back wall fabric layers of 4 mm Kevlar and 6 mm Nextel sheets are embedded as a "bullet-proof vest". The shields of the ISS manned modules can withstand impacts by objects up to 1.4 cm in size, at velocities on the order of 10 km/s. The related kinetic energy corresponds to a 1.5 t mid-size car hitting at 50 km/h, or to the energy released by an exploding hand grenade. An ISS module of 100 m^2 cross-section is expected to have impacts from debris objects of $d \geq 1$ cm at a rate of 1 in 410 years. Meteoroid impacts are negligible in this size regime. For the same module cross-section, impacts from objects of $d \geq 1$ mm will occur at a rate of 1 in 6 months, with a 90% probability that they originate from meteoroids. Whipple shields rely on impact velocities that are larger than about 6 km/s, in order to break up the impacting object into a cloud of solid, liquid, and gaseous matter that can more easily be withheld by the back wall and intermediate fabrics, due to a wider spreading and time-distributed arrival of the fragment cloud, with a resulting reduction of the pressure peak. While the volume and mass requirements of such shields are prohibitive for normal spacecraft, there are still ways of reducing their impact risk. The Canadian Radarsat, for instance, used light-weight Nextel fabric covers as external protection and rearrangements of sensitive spacecraft subsystems to improve the survivability of their 5-year mission by 50–87%. This gain was achieved for a mass penalty of 0.6% (+17 kg).

In order to increase the safety of US space assets the US Space Command is upgrading its operational surveillance network. In particular, the replacement of the UHF-based surveillance fence that extends along the 33rd parallel across the United States, to an S-band system is expected to allow a catalog maintenance down to 2 cm sizes at the ISS altitude. This could increase the SSN catalog size to more than 100,000 objects. With the full orbit knowledge of these objects one would be in a position to close the gap between avoidable and shieldable objects for ISS, and hence significantly improve the on-orbit safety for manned space flight.

The space debris environment at critical sizes above 10 cm has in the past been dominated by explosion fragments and by dysfunctional but intact remnants of previous missions. Collisions played a minor role until the FengYun 1C ASAT test in 2007, and the accidental collision between Cosmos 2251 and Iridium 33 in 2009. By 2010 these two events alone accounted for almost 40% of the US SSN catalog. In order to curtail the growth rate of hazardous space debris, particularly in the LEO regime, the international space community has identified and adopted a set of space debris mitigation measures. The main categories of recommendations can be summarized as follows:

- Reduction of mission-related objects
- Prevention of on-orbit explosions (passivation)
- Limitation of non-explosive release events

- Collision avoidance between trackable objects and operational assets
- Post-mission disposal of space systems

Mission-related objects (MROs) contribute 5.5% of the trackable catalog population with 70% of these related to launch systems and 30% related to payloads. MROs, also referred to as operational debris, are defined as objects released during nominal operations by both spacecraft and rocket bodies. This includes debris from launcher staging and payload separation (such as adapters, shrouds, and clamp bands), and objects released during spacecraft deployment and commissioning (such as parts of explosive bolts, solar array latches, and lens covers). Most of these objects are released with low relative velocities, and so they remain in close proximity to the operational orbit of the source object.

The release of MROs can be limited by system design. The best method of reducing the population of MROs is not to produce the objects in the first instance. This is reflected in most debris mitigation standards through recommendations to minimize, or to avoid the use of debris-generating systems (e.g., yo-yo de-spinners, nozzle closures of propulsion systems, and protective lens covers). System design is also encouraged to ensure that released parts (e.g., antenna deployment mechanisms, protective covers, explosive bolts, ullage motors, and heat shields) are retained with the primary object. This can be achieved through the use of lanyards, sliding or hinged covers, and special catchment devices. Moreover, materials and basic system technologies (e.g., tanks, surface materials, and structures) should be selected such that they are resistant to environmental degradation (e.g., aging by radiation, atomic oxygen and microparticle impact erosion, and thermal cycling).

Explosions of spacecraft and upper stages in orbit have been the major source of debris in the past, with 245 such events up to 2010, at a mean annual rate of about 4.5. These failures, which caused close to 4,000 cataloged fragments, might have been avoided, if onboard passivation techniques had been employed. Such procedures are a standard on many of today's launchers, and so far there are no recorded explosions of successfully passivated orbital stages. End-of-life (EOL) passivation was first considered as a design requirement at the beginning of the 1980s. All upper stages and spacecraft which were launched before then, and which are still in orbit, continue to pose an explosion hazard (a Titan III-C transtage launched in 1967 exploded after 27 years in orbit). Hence, there is a significant number of latent explosion sources still on orbit.

Space debris mitigation standards recommend that all onboard reservoirs of stored energy (e.g., propellants, pressurants, batteries, momentum control gyros) should be permanently depleted when they are no longer required for any nominal or post-mission operations. The following passivation aspects should be considered:

- Idle burn or venting of residual propellants, with valves left open
- Venting of all pressure systems, and/or activation of pressure relief mechanisms to avoid explosions due to external heating
- Discharge of batteries, shut down of charging lines, and maintenance of a permanent discharge state
- Deactivation of range safety systems
- Dissipation of energy contained in momentum control gyros

Fuel depletion or "idle" burns of orbital stages may be performed such that the resulting thrust leads to a braking maneuver, leaving the stage in a reduced-lifetime orbit. The residual lifetime should be less than 25 years to be compliant with international recommendations for space debris mitigation.

The class of non-breakup release events includes residues from SRM firings (slag and dust), sodium-potassium droplets that were generated during RORSAT reactor core ejections, or surface degradations products that are caused by aging paint coatings or multilayer insulations (MLI). All of these debris sources can be reduced or even suppressed in total through design measures.

Collision avoidance, as another debris mitigation measure, is nowadays implemented by many space operators for their operational payloads. This concept, however, can only be applied to about 5% of the catalog population, assuming that less than 1,100 of the on-orbit payloads in 2010 were operational, of which about 80% could be maneuvered. Hence, future collisions will most often occur between uncontrollable debris objects. To reduce the number of catastrophic collisions between large, intact but nonoperational objects, the use of ground-based Lasers is investigated. If a close conjunction is predicted, then a radar-guided Laser beam (see Fig. 46.10) could ablate material from one of the objects or generate photon pressure to impart a momentum that could sufficiently alter the flyby distance to a safe level.

In 2010 the mean time between two catastrophic collisions in the LEO region was on the order of 5 years. One way of reducing future collision rates is through post-mission disposal measures, i.e., through mass removal of (still) active space assets. International guidelines recommend removing spacecraft and orbital stages after their mission completion, in particular from the densely populated LEO regime and from the unique GEO ring. For GEO spacecraft disposals an orbit raise to a graveyard region at approximately 300 km above GEO is recommended. The magnitude of the altitude raise to a near-circular disposal orbit is determined by the area-to-mass ratio of the spacecraft. It is defined such that long-term orbit perturbation effects will not lead to a return of the orbit into a "GEO protected region" that extends ± 200 km around the GEO ring (which is at 35,786 km altitude). Table 46.7 shows a summary of GEO post-mission disposals over a 12-year time span. It is evident that the degree of compliance with international guidelines has gradually improved and has reached a mean level of almost 50% in 2010.

End-of-life mitigation measures for the "LEO protected zone" (that is below 2,000 km altitude) recommend an active deorbiting, or a natural decay of payloads and orbital stages into a destructive reentry within 25 years after mission completion. For typical area-to-mass ratios of such objects a timely natural decay requires an end-of-mission altitude below 600 km. Alternatively, chemical or electrical propulsion can be used to induce a direct reentry. A mono-propellant hydrazine system would need about 8.8% of the spacecraft mass for a controlled deorbit from 800 km (6.3% for a bi-propellant system). Electrical propulsion systems, due to their higher ejection velocities, can be more mass efficient by a factor of about 10. Their lower thrust levels, however, will lead to an uncontrolled reentry. An accelerated uncontrolled reentry can also be induced by thin, conductive tethers of several kilometers length that orientate themselves along the local vertical through

gravity gradient forces. As they cut through the magnetic field lines they induce a tether current that is closed through the ambient plasma and that leads to a retarding Lorentz force, acting opposite to the direction of motion of the spacecraft, with best performance for low-inclination orbits. For a mass penalty of less than 3% such systems are able to reduce orbital lifetimes of Globalstar satellites (at 1,400 km and 52° inclination) from 9,000 years to less than 2 months, and they can reduce orbital lifetimes of Iridium satellites (at 780 km and 86° inclination) from 100 years to less than 8 months.

Space debris mitigation guidelines, standards, and requirements have been developed by several space agencies since the early 1990s. In parallel, the knowledge on space debris sources increased, and the understanding of effective remedial actions improved.

A first step to a wider, international application of debris mitigation measures was taken by the Inter-Agency Space Debris Coordination Committee (IADC) in 2002, with the publication of their Space Debris Mitigation Guidelines (Anonymous 2002). This document, which was first presented at the UNCOPUOS Scientific and Technical Subcommittee in 2003, serves as a basis for the development of space debris mitigation principles in two directions: toward a nonbinding policy document, and toward applicable implementation standards. The former route was followed by a UNCOPUOS working group, while the latter direction was pursued by an Orbital Debris Coordination Working Group (ODCWG) within the Technical Committee 20 and its Subcommittee 14 of the International Organization for Standardization (ISO TC20/SC14). To a large extent these UN and ISO working groups recruit their experts from IADC member organizations.

International space debris mitigation policies and standards, based on the consensus of the IADC guidelines, could in the future facilitate and harmonize the implementation of space debris mitigation measures at a global scale. Internationally agreed standards could enforce appropriate debris mitigation measures on spacecraft operators and launch service providers through the mechanisms of conditional launch license issuance and insurance coverage, depending on the acceptance of a space debris mitigation plan by the launch authority. Fifty years after the beginning of space flight the voluntary implementation of debris mitigation and disposal measures by many space operators has become common practice. For several launching nations the compliance with national regulations or with a national space law makes debris mitigation measures even mandatory.

While debris mitigation is a necessary condition to maintain an orbital environment with a tolerable risk level for space missions, long-term forecasts of the debris environment indicate that some orbit regions may still become unstable within a few decades. Figure 46.9 illustrates the evolution of the LEO debris population larger than 10 cm for a hypothetical case of no future launches. The case corresponds to an extreme, hypothetical mitigation scenario, with immediate deorbiting of payload(s) and insertion stage(s) after orbit injection, and with no intermediate release of mission-related objects. Predictions with NASA's LEGEND model (Liou and Johnson 2008a, b) demonstrate that even for such optimistic assumptions the LEO environment will become unstable. Within 20 years collision fragments will

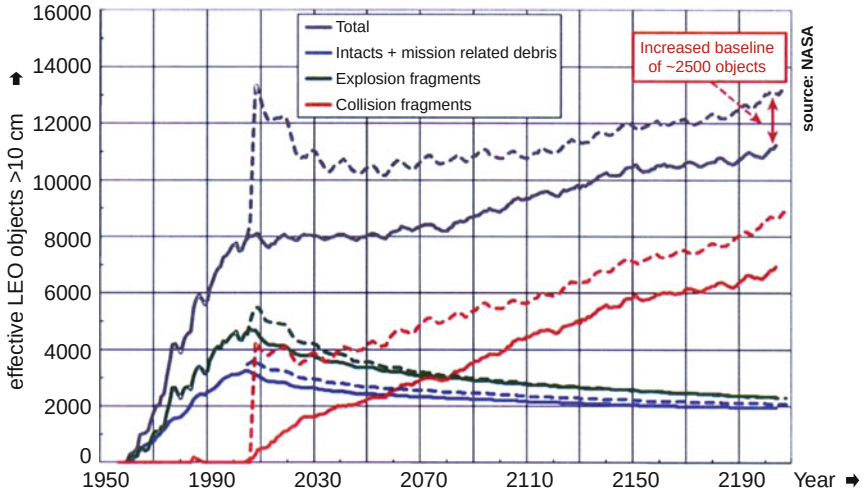


Fig. 46.9 Long-term prediction of the LEO debris environment of critical-size objects of d > 10 cm, discriminated by source terms, for a "no future launch" scenario according to (Liou and Johnson 2008a)

start to outnumber explosion fragments, and within 70 years an initially stabilizing effect from naturally reentering objects will be superseded, and the 10 cm population growth will follow the slope of the collision-induced fragment increase. In the course of the 200 year projection more and more collision fragments will collide with other collision fragments. This so-called "Kessler syndrome" is a self-maintained collisional cascading process that is fed by the LEO mass reservoir of 2,500 t in 2010. Its natural termination would be reached in the very far future when all LEO crossing objects are ground to subcritical sizes that can no more reach the specific impact energy threshold of $\geq$40 kJ/kg for causing a catastrophic breakup. As a consequence, space debris mitigation is a necessary but insufficient condition to maintain a stable orbital environment.

In order to sustain an acceptable debris risk level for future space missions, debris mitigation measures must be augmented by space debris environment remediation measures that actively remove mass from orbit, with priority on the LEO regime. The effectiveness of space debris environment remediation measures is governed by their capability to reduce the short- and long-term risk of catastrophic collisions. An initial indicator of the debris environment deterioration is the concentration of critical-size objects of 10 cm and larger that have the capability to cause catastrophic breakups. Figure 46.4 shows the distribution of the catalog objects in LEO. Highest concentrations are at 800 $\pm$ 200 km, spread over inclination bands at 65 $\pm$ 2°, 72 $\pm$ 2°, 82 $\pm$ 1°, and 97 $\pm$ 3°, with an almost equal share of intact objects, explosion fragments, and collision fragments. There is a lower, secondary LEO peak at 1,400 $\pm$ 100 km, and minor local peaks for MEO navigation satellite constellations and for GEO objects, both of which are about three orders of magnitude lower.

In order to rank priorities for mass removal from orbit, it is important to determine a risk metric. Three parts of such a metric can be defined:

- Metric #1: [catastrophic collision rate] = [10 cm collision flux] × [mean target cross-section]
- Metric #2: [short-term risk due to a catastrophic collision] = [10 cm collision flux] × [mean target cross-section] × [target mass]
- Metric #3: [long-term risk due to a catastrophic collision] = [10 cm collision flux] × [mean target cross-section] × [target mass] × [target orbit lifetime]

In these metrics, as a simplification, it shall be assumed that the target cross-section and the target mass are significantly larger than those of the impactor, and that the orbit lifetime of resulting collision fragments is similar to the orbit lifetime of the intact target object.

In 2010, the orbit environment consisted of more than 12,000 cataloged LEO objects, larger than 10 cm, of a total mass of almost 2,500 metric tons. The corresponding rate of catastrophic collisions was 0.19 per year, resulting in one such event every 5–6 years. About 45% of these collisions would have a rocket body, while 55% would have a spacecraft as their main object. For metric #1 (catastrophic collision rates), as much as 22% can be attributed to a single $2^\circ \times 50$ km bin at $87 \pm 1^\circ$ inclination and 775 ± 25 km altitude, covering 80 intact objects, of which 73 are satellites of the Iridium constellation, each with 660 kg mass and 22 m^2 cross-section. These intact objects are facing fragments from the Iridium 33/Cosmos 2251 collision of 10 February, 2009, and from the Chinese ASAT test of 11 January, 2007, as the main causes of their 10 cm collision flux. A secondary maximum of catastrophic collision rates is due to a cluster of Cosmos satellites at a bin of $83 \pm 1^\circ$ inclination and 975 ± 25 km altitude.

The short-term risk to the orbital debris environment can be expressed by the metric #2, where the dominant target object masses drive the number of critical-size collision fragments, which determine the short-term level of the debris environment deterioration. Using the same assumptions for determining catastrophic collision rates as above, the mass-weighted short-term environment risk is governed to 61% by rocket bodies and to 39% by spacecraft. Approximately 28% of the overall short-term risk is due to objects in a single bin of $2^\circ \times 50$ km, centered at $71 \pm 1^\circ$ inclination and 825 ± 25 km altitude. Most of the corresponding mass is related to Russian Zenit 2 second stages, each with dry masses of 8,900 kg and cross-sections of 33 m^2, and to 15 Cosmos spacecraft, each of 3,200 kg mass and 6 m^2 cross-sections. Of the 20 top-ranking objects according to metric #2, 19 are Zenit 2 rocket bodies, 16 of which are located in the above defined bin.

The long-term risk to the orbital debris environment can be expressed by the metric #3, where the on-orbit residence time of resulting collision fragments is applied as a weighting factor to the metric #2. As a simplifying, conservative assumption the same average orbital lifetimes shall be considered for the target object and its resulting fragments. The resulting aggregate of the individual products of catastrophic collision rate, target mass, and target orbit lifetime, over all intact LEO objects below 1,300 km, leads to a long-term debris environment risk indicator that is governed to 72% by rocket bodies and to 28% by spacecraft.

Approximately 42% of the overall long-term risk is due to the same objects that dominate the risk metric #2, stemming from a single bin of $2^\circ \times 50$ km, centered at $71 \pm 1^\circ$ inclination and 825 ± 25 km altitude. Again, most of the related mass is due to Russian Zenit 2 second stages, each with an empty mass of 8,900 kg, with a cross-section of 33 m^2, and with an orbit lifetime on the order of 700 years. Of the 23 top-ranking objects according to metric #3, 19 are Zenit 2 second stages, with 16 thereof from a single $2^\circ \times 50$ km bin. By removing these stages, the long-term risk metric #3 might be reduced by about 24%.

Long-term debris environment projections (see Liou and Johnson 2008a, b; Bastida et al. 2009; Liou 2011), based on an extreme scenario with no future launches and 90% success rates of LEO post-mission disposals indicate that the current environment will lead to a net increase of the long-lived 10 cm debris population by about 30% in the next 200 years (see Fig. 46.9). This result confirms the onset of collisional cascading in some LEO orbit regions, also known as the "Kessler syndrome." In the case of continued launch activities at today's rates, the 10 cm debris population will even grow by 60%, fueled by 24 catastrophic collisions (see Bastida et al. 2009). These collisions will almost exclusively occur between members of the previously identified, densely populated LEO inclination bands, and between orbits of low to moderate eccentricities. Further parametric studies of the long-term debris environment evolution predict that active mass removal, focusing on inclination and altitude bands with high mass concentrations in a few large objects, can reduce the number of catastrophic collisions to 14 within 200 years, and lead to a stable 10 cm object population, if 5–10 removals per year are performed (see Bastida et al. 2009; Liou 2011).

Several research groups, with different backgrounds and application targets, have devised techniques that could be used for the removal of mass from orbit. Table 46.8 shows an overview of methods that could be within technological reach. With the exception of ground- and air-based directed energy methods (mainly Lasers, see Fig. 46.10), all techniques are space based, and all of them are suited for the most critical LEO regime (with some also applicable for MEO or GEO mass removals). All methods in Table 46.8 that are restricted to debris sizes below 10 cm can contribute to space environment remediation, but at a size regime that normally does not lead to catastrophic collisions, and that hence does not fuel the collisional cascading process. The focus shall thus be on techniques that can effectively reduce the orbit lifetime of intact objects and fragments that are larger than 10 cm, including full-size satellites and orbital stages. In order to qualify as a remediation measure (as opposed to a mitigation measure), all techniques must be applicable to dysfunctional target objects, for instance, with the assistance of a remover spacecraft or through the attachment of external deorbiting devices in a rendezvous mission.

Solar sails can be used to increase the eccentricity of a target orbit. The periodic changes in perigee altitude, in combination with the increased, nonconservative drag perturbation at the perigee passes, lead to a secular decrease of the orbit energy, and hence accelerate the orbit decay. The decay rate is directly proportional to the area-to-mass ratio of the solar sail/spacecraft compound. Solar sails could be

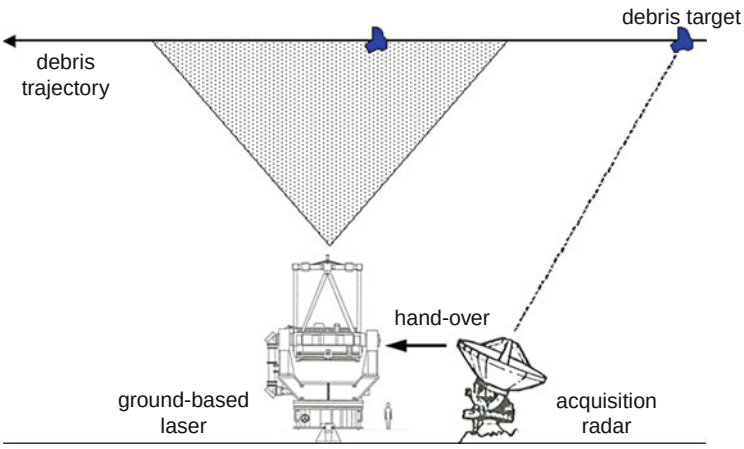


Fig. 46.10 Debris removal and/or debris orbit changes induced by a ground-based laser, tasked by a colocated acquisition and tracking radar (Klinkrad and Johnson 2009)

Table 46.7 History of post-mission disposal activities of geostationary spacecraft through 2010 ($L_1 = 75°E$, $L_2 = 105°W$, "too low" and "compliant" refer to the IADC orbit raising recommendation in (Flohrer et al. 2011))

EoL Disposal	'99	'00	'01	'02	'03	'04	'05	'06	'07	'08	'09	'10	Total	
Left at L_1	5	3	5	1	–	2	1	2	1	2	3	3	28	(15.5%)
Left at L_2	1	1	1	1	1	1	1	1	–	1	–	–	9	(5.0%)
Left at L_1/L_2	–	2	–	–	–	–	1	–	–	1	–	–	4	(2.2%)
Drift orbit (too low)	4	2	6	5	7	5	5	7	1	1	6	4	53	(29.3%)
Drift orbit (compliant)	5	3	2	4	8	5	11	9	11	6	12	11	87	(48.0%)
Annual Total	15	11	14	11	16	13	19	19	13	11	21	18	181	(100%)

inflated spheres, or arrangements of flat surfaces. They should be metallized to increase the photon-surface momentum exchange (see Fig. 46.11). Drag augmentation devices directly affect the area-to-mass ratio of an object and hence increase the airdrag that leads to an orbit lifetime reduction. For both, solar sails and drag augmentation devices, the benefit of reducing the orbit lifetime of the target object should outweigh the drawback of an increased collision cross-section.

The magnetic sail concept is using a magnetic field to deflect the plasma of the solar wind in order to accelerate or decelerate a spacecraft. The magnetic sail utilizes a loop of superconducting cable to which an electrical current is applied. The magnetic field created by the current in the loop stiffens the cable into a rigid circular shape. Charged particles encountering the magnetic field are deflected, and momentum is imparted on the loop. In the solar wind, the magnetic sail creates drag and accelerates the spacecraft in the direction of the wind. Employing the magnetic sail in non-axial configurations produces a force perpendicular to the solar wind that can be used for maneuvers. However, the technical implementation of the concept

Table 46.8 Debris removal techniques and their applicability with respect to orbit regime and debris size (credit: NASA/JSC)

Debris removal technique	Altitude regime	Debris size
Solar sail	LEO, MEO, GEO	>1 m
Magnetic sail	LEO, MEO, GEO	>1 m
Attachable deorbit/re-orbit module	LEO, MEO, GEO	>1 m
Capture/orbital transfer vehicle	LEO, MEO, GEO	>1 m
Drag augmentation device	LEO	>10 cm
Momentum tethers	LEO, GEO	>10 cm
Electrodynamic tethers	LEO	>10 cm
Airborne laser/directed energy	LEO	<10 cm
Space-based laser/directed energy	LEO, MEO, GEO	<10 cm
Ground-based laser/directed energy	LEO	<10 cm
Space-based magnetic field generator	LEO	<10 cm
Sweeping/retarding surface (balloon, film, foam ball, etc.)	LEO	<10 cm

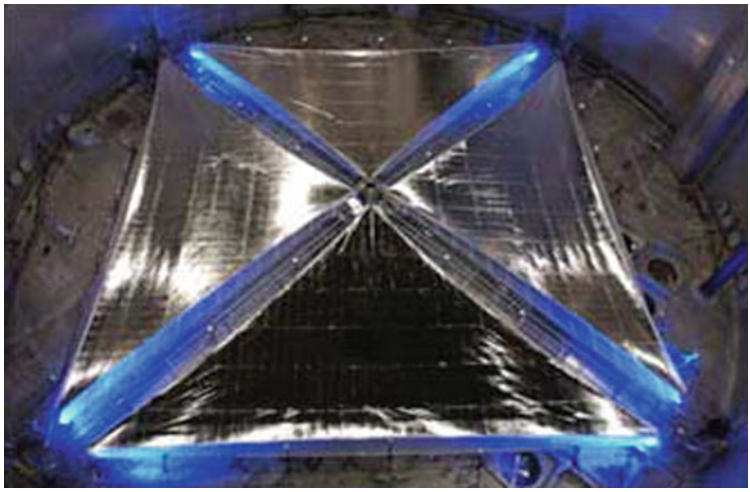


Fig. 46.11 Use of solar radiation pressure and/or drag augmentation for LEO orbit lifetime reduction (Klinkrad and Johnson 2009)

is not yet mature, and it would be vulnerable to small-particle impacts from debris and meteoroids.

Tethers can be applied in mass removal systems in two different ways: as conductive electrodynamic tethers, or as momentum-exchange tethers. When two sizeable objects are connected by a momentum-exchange tether, and if this tether is reeled out along the local vertical, then different orbit velocities and perturbing accelerations cause a swinging motion, primarily within the common orbital plane. If the tether is then cut at the time of its highest retrograde ΔV, then the lower object will obtain a lower perigee (e.g., in LEO for direct deorbit, or for release into a reduced-lifetime orbit), and the upper object will obtain a higher apogee.

Likewise, in a reverse mode, such a system can be use to re-orbit LEO or GEO objects. The related tether loads can be significant, and the tether design is technologically demanding. For a net gain, the active remover satellite would have to deorbit more than one large object, and also dispose of itself.

In the case of an electrodynamic tether (EDT) an electromotive force is generated within a conductive wire that is attached to a space vehicle as it moves through the Earth's magnetic field. If a pair of plasma contactors at either end of the tether emits and collects electrons, an electric current flows through the tether by closing the circuit via the ambient plasma. The tether then generates a Lorentz force via interaction between the tether current and the geomagnetic field. This force acts as a deceleration, opposite to the direction of flight, and it hence reduces the orbit lifetime by dissipating orbital energy. The efficiency of this method depends on the average magnetic induction, and it thus decreases with $1/r^3$ and with cos i_m, where r is the geocentric radius and i_m is the mean geomagnetic inclination of the orbit. The resulting, reduced orbit lifetime is proportional to $1/L^2$, where L is the tether length. Figure 46.12 shows how a servicing satellite attaches an EDT to a dysfunctional payload to accelerate its decay. An electrodynamic tether is a promising deorbit concept due to its relatively simple design, its low system mass, and its efficiency even at high LEO altitudes (see Pardini et al. 2006). A conductive aluminum tether with a system mass fraction of 2.5% as compared to the client object can reduce the lifetime of a high-inclination low LEO constellation at 780 km altitude from 100 years to less than 1 year (e.g., for Iridium). For a medium-inclination high LEO constellation at 1,400 km the orbital lifetime can be reduced from 9,000 years to less than 2 months (e.g., for Globalstar).

Technologically, the most mature solutions of orbit mass removal are the attachments of de- or re-orbit propulsion modules, or the capture of a target object by a space tug. Figure 46.13 shows the latter approach for a GEO tug satellite. The shown ROGER remover spacecraft of 3.5 t mass is intended to be launched as a secondary payload into a GTO orbit, from which it injects itself into the GEO ring. It then performs rendezvous operations with a preassigned client object, inspects it via video cameras, and casts a net over it. The net is then tightened via reels in the end-masses while leaving it connected to the remover spacecraft by means of a tether. The whole compound is then tugged into the GEO graveyard orbit, where the tethered net enclosing the debris object is released. With an overall propellant mass fraction of 77% and 20 disposable nets (or, alternatively, 10 disposable nets and 2 reusable, tethered gripper devices) the ROGER remover satellite is planned to perform up to 20 GEO disposal missions. While this concept is particularly attractive for GEO, it could also be employed in densely populated LEO regions. Several similar concepts, mostly based on prior studies of servicing spacecraft, are under investigation in space industry.

Mass removal from orbit has a technical, a financial, and a legal dimension. As of today, many of the solution concepts listed in Table 46.8 are not yet sufficiently advanced in their technology readiness, and even the most mature concepts would incur significant costs if they were realized. Moreover, the removal of on-orbit mass

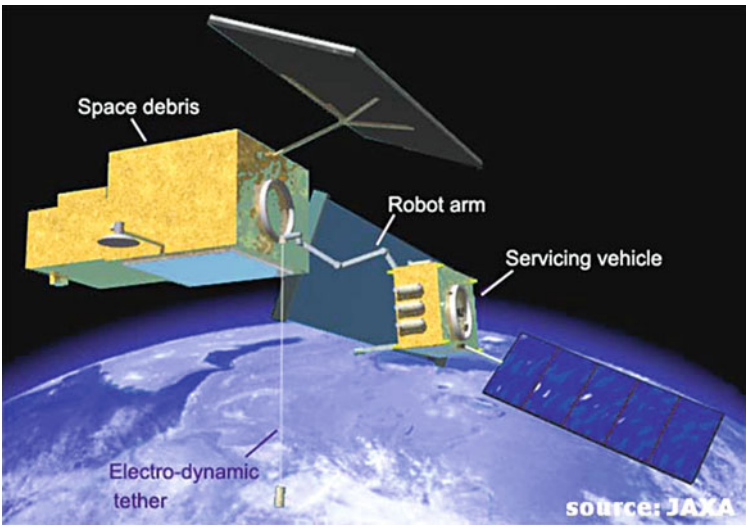


Fig. 46.12 Use of conductive tethers for LEO orbit lifetime reduction through electrodynamics forces induced by the geomagnetic field (credit: JAXA)

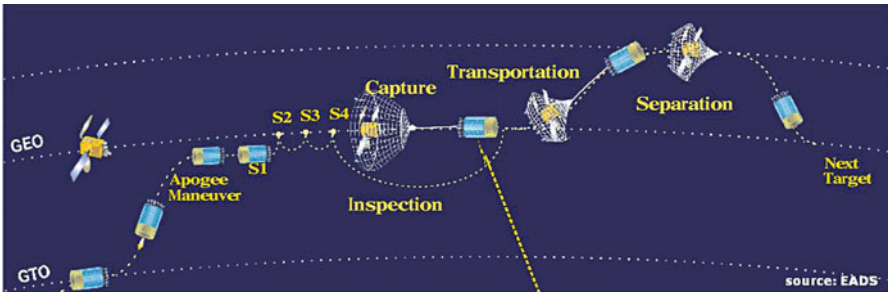


Fig. 46.13 Use of a space tug to re-orbit objects from the GEO region (credit: EADS Astrium)

that belongs to another launch authority and/or space operator requires mutual agreement on the procedure, on the cost sharing, and on possible liabilities, particularly for an uncontrolled reentry.

Conclusion

Out of almost 16,000 objects that were contained in the US Space Surveillance Network catalog in 2010, slightly more than 1,000 were operational spacecraft, of which roughly 80% could be maneuvered, and of which 400 were in the GEO ring, while more than 500 were in the LEO regime. Since LEO and GEO are of particular interest for space operators, these orbit regimes were denoted at

"protected regions" by IADC and UNCOPUOS. In order to safeguard a sustainable long-term usability of the LEO and GEO regions space debris mitigation measures must be applied rigorously by all space faring nations and supernational organizations. The necessary mitigation measures have been identified, e.g., by the 12 IADC members, and cast into international guidelines and standards, into agency-specific sets of requirements, and into national space laws. Analyzes of the long-term evolution of the space debris environment indicate that such agreed mitigation measures are a necessary but insufficient condition to maintain the space object population at a stable level. Even an extreme mitigation scenario with no future launches will result in a long-term collisional cascading (the so-called Kessler syndrome) at some LEO altitudes. This run-away process is fueled by existing mass on orbit, and the only way to stabilize the environment is through active mass removal from particularly densely populated altitude and inclination bands. This is a challenging task from a technical, economical, and legal point of view that can only be successfully implemented, if an international consensus is reached among space faring nations. In the past the Scientific and Technical Subcommittee (STSC) of UNCOPUOS, with guidance from IADC members and contributions from COPUOS members, installed a working group that developed the UNCOPUOS Debris Mitigation Guidelines (see Anonymous 2009) in the course of a multi-year work plan. Likewise, in 2010, UNCOPUOS STSC established a working group on the "sustainable use of outer space." This initiative could be a starting point for the development of an international framework that could include space debris environment remediation as one of its main objectives.

Following the publication of previous reports on "space traffic management" and "space debris mitigation," the International Academy of Astronautics (IAA) will publish a report on "space debris environment remediation". Its authorship, with more than 20 contributors from 11 different countries and many different disciplines, could further consolidate the basis for international deliberations on the technical, economical, and legal aspects of mass removal from orbits with critical mass concentrations, to allow a continued and safe use of space also in the far future.

The overarching principle of a responsible and sustainable use of space was formulated back in the 1990s by the late Joseph P. Loftus, former assistant director of NASA/JSC: "Space operations should comply with a general rule of the National Park Service: 'What you take in you must take out.'"

Cross-References

- ▶ An Examination of the Governmental Use of Military and Commercial Satellite Communications
- ▶ Appendix 1: The World's Launch Sites
- ▶ Appendix 2: Major Launch Systems Available Globally
- ▶ Appendix 3a: Global Communications Satellite Systems
- ▶ Introduction to Satellite Navigation Systems

► Operational Applications of Radar Images
► Space Weather and Hazards to Application Satellites
► Space Weather & Survivability of Spacecraft
► Telemetry, Tracking, and Command (TT&C)

Notes

1. TLE (two-line elements) are an orbit data format in terms of doubly averaged, mean orbital elements

References

H. Klinkrad, *Space Debris – Models and Risk Analysis* (Springer-Praxis, Berlin/Heidelberg/New York, 2006)

M. Oswald, P. Wegener, S. Stabroth, C. Wiedemann, J. Rosebrock, C. Martin, H. Klinkrad, P. Vörsmann, The MASTER 2005 model, in *Proceedings of the 4th European Conference on Space Debris*, Darmstadt, ESA-SP-587, 2005

S. Flegel, Maintenance of the ESA MASTER Model, final report of ESA contract 21705/08/D/HK, 2010

Anonymous, UNCOPUOS space debris mitigation guidelines. A/RES/62/217, UNCOPUOS Scientific & Technical Sub-Committee, 2009

Anonymous, IADC space debris mitigation guidelines. IADC-02-01, Rev. 1, 2002

C. Pardini, T. Hanada, P.H. Krisko, Benefits and risks of using electro-dynamic tethers to de-orbit spacecraft, in *57th International Astronautical Congress*, IAC-06-B6.2.10, Valencia, 2006

H. Klinkrad, N.L. Johnson, Space debris environment remediation concepts, in *Proceedings of the 5th European Conference on Space Debris*, Darmstadt, ESA-SP-672, 2009

J.C. Liou, N.L. Johnson, Instability of the present LEO satellite populations. Adv. Space Res. **41**, 1046–1053 (2008a)

J.C. Liou, N.L. Johnson, A sensitivity study of the effectiveness of active debris removal in LEO. Acta Astronaut. **64**, 236–243 (2008b)

B. Bastida, H. Krag, Strategies for active removal of space debris, in *Proceedings of the 5th European Conference on Space Debris*, Darmstadt; ESA-SP-672, 2009

J.-C. Liou, An active debris removal parametric study for LEO environment remediation. Adv. Space Res. **47**(11), 1865–1876 (2011)

T. Flohrer, R. Choc, B. Bastida, Classification of geosynchronous orbits – issue 13. European Space Agency, GEN-DB-LOG-00074-OPS-GR, 2011

N.L. Johnson, The International Space Station and the space debris environment – 10 years on, in *Proceedings of the 5th European Conference on Space Debris*, Darmstadt; ESA-SP-672, 2009

Space Weather and Hazards to Application Satellites

47

Michael J. Rycroft

Contents

Introduction 1176
The Sun, Our Star 1177
Energy from the Sun 1177
Sunspots and the 11-Year Solar Cycle 1177
Coronal Mass Ejections (CMEs) 1181
The Earth's Atmosphere and Near-Space Environment 1182
Five Key Aspects of the Space Environment 1183
The Residual Atmosphere in Low Earth Orbit (LEO) – Near-Vacuum Conditions 1184
Thermal Radiation – From the Sun and the Earth 1184
Plasma – The Electrically Charged Gas, the Ionosphere 1185
Energetic Charged Particles – Van Allen Radiation Belts, Solar Protons, and Cosmic Rays 1185
Micro-meteoroids and Orbital Debris 1187
How the Space Environment Affects Satellite Operations 1188
Near-Vacuum Conditions 1188
Thermal Radiation 1189
Plasma 1189
Energetic Charged Particles 1190
Micro-meteoroids and Space Debris 1191
Protection of Application Satellites Against Space Weather Effects 1192
Cross-References 1192
References 1193
Further Reading 1193

M.J. Rycroft
Cambridge Atmospheric, Environmental and Space Activities and Research (CAESAR) Consultancy, 35 Millington Road, Cambridge, CB3 9HW, UK
e-mail: michaelrycroft@btinternet.com

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_78, © Springer Science+Business Media New York 2013

Abstract

Magnetic activity on the Sun causes disturbances moving out through interplanetary space as a stronger than usual solar wind. Enhanced solar activity can also produce large fluxes of penetrating energetic protons of great destructive potential. When either of these phenomena arrives near the Earth and strikes a satellite or spacecraft, damage is likely to be done, either directly and indirectly. Here we give an overview of the several different effects which occur in low Earth orbit (LEO), in medium Earth orbit (MEO), and in geostationary orbit (GEO) as well as in high inclination elliptical Earth orbits where positioning satellites are located.

Keywords

Atomic oxygen • Coronal mass ejections (CMEs) • Galactic cosmic rays • Ionosphere • Ionospheric scintillations • "Killer" electrons • Magnetic storm • Plasma • Radio communications • Satellite drag • Satellite operations • Single event upsets • Solar activity • Solar cycle • Solar protons • Space debris • Space environment • Space weather • Sunspots • Van Allen radiation belts

The answer, my friend, is blowin' in the wind. (A line from a song by Bob Dylan 1962)

Introduction

"Space weather" is defined by the US National Space Weather Program as referring to "conditions on the Sun and in the solar wind, magnetosphere, ionosphere, and thermosphere that can influence the performance and reliability of spaceborne and ground-based technological systems and can endanger human life or health." From the viewpoint of an application satellite, the space environment is a hostile, harsh, and hazardous environment. It is thus a significant challenge to design application satellites so that they can operate reliably throughout a significant in-orbit lifetime, without the possibility of benefiting from physical repair.

Valuable background information on the science lying behind the subject is to be found in a chapter by Rycroft (2010), in the recent book of Schrijver and Siscoe (2009), and at the websites http://www.vsp.ucar.edu/Heliophysics/ and http://www-spof.gsfc.nasa.gov. There are several books at different levels on space weather, such as those by Freeman (2001), Song et al. (2001), Carlowicz and Lopez (2002), Daglis (2001; 2004), Bothmer and Daglis (2007), Lilenstein (2007), and Moldwin (2008). Singh et al. (2010) have recently written a useful overview paper; in addition, useful space weather websites are indicated in the endnotes.

First, we consider features on the Sun and solar activity as a driver of space weather phenomena. Then we discuss the five key features of the space environment which impact the design of sensors, instruments, electronic and all other equipment aboard all space vehicles, and which influence the performance of satellites or spacecraft in different orbits. Since the space age began, numerous measurements have been made of important physical parameters of the space environment; from these, useful

empirical (numerical) models of variations of those parameters have been constructed. Such valuable models are complemented by theoretical models, that is to say models which are based upon our understanding of the physical processes operating. (For numerical values, we shall use here the International System of units [SI] based upon the meter [m], kilogram [kg], second [s], and Ampere [A]).

The Sun, Our Star

Energy from the Sun

Thermonuclear reactions occur within the Sun's core. The tremendous amount of energy released allows the Sun to provide essentially all the energy which the Earth receives from the cosmos. It is this energy which, directly or indirectly, powers all phenomena occurring on Earth and in its environs. It does this in the form of radiant energy (see "Electromagnetic radiation principles and concepts" in this volume) or in a form of energy derived from this source. Although the flux of this radiant energy peaks in the visible part of the spectrum to which our eyes are tuned, much energy is emitted as ultraviolet, X-ray, and gamma-ray – i.e., high energy – photons, plus longer-wavelength radiation at infrared and radio wavelengths.

The Sun's spectrum is approximately that of a black body at 5,800 K; this is the temperature of the solar "surface," a plasma termed the "photosphere." Just above the photosphere is the chromosphere, a cooler region. The outer solar atmosphere, at heights exceeding a fraction of a solar radius, known as the corona, emits X-rays because its temperature is about one million K. As the solar wind, the Sun's outer atmosphere flows away from the Sun at speeds of several hundred kilometers per second.

Sunspots and the 11-Year Solar Cycle

From time to time, dark splodges appear on the photosphere; these are called sunspots. Two photographs of the Sun are shown in Fig. 47.1, taken on September 27, 2001 (on the right), and 2008 (on the left). The right hand image shows sunspots, generally occurring in pairs, at solar maximum conditions; the left hand image shows no visible sunspots during a very deep solar minimum. Sunspots are regions of intense magnetic fields, with complex (twisted) arched magnetic field lines above the surface connecting each pair of sunspots.

It is observed that as the Sun rotates (once in 27 days on average, but in approximately 25 days at the Sun's equator and about 32 days nearer the poles), the polarity of the leading sunspot is always in the same direction, for example outward, and that of the following sunspot is always in the other direction, or inward. This applies to all sunspots in one hemisphere of the Sun, with the situation being exactly the opposite for sunspots in the opposite hemisphere. All sunspots migrate from high latitudes toward the Sun's equator on a timescale of 11 years. Then new sunspots appear at high latitudes, with the leading sunspots having the opposite polarity; this situation marks the beginning of each new 11-year solar cycle.

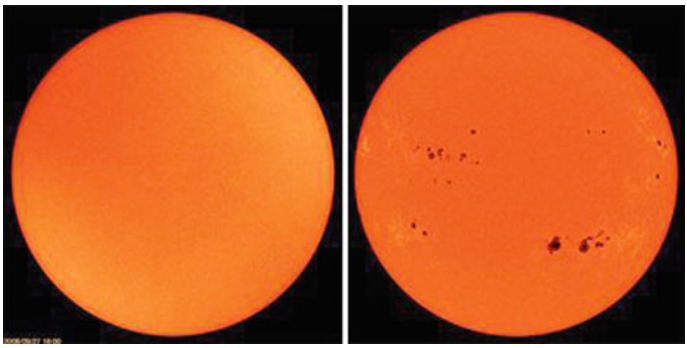


Fig. 47.1 Sometimes, the Sun has sunspots on its visible surface (Courtesy of NASA Goddard Space Flight Center (October 7, 2008). Spotless Sun: Blankest year of the space age. *Science Daily*. Retrieved June 28, 2011, from http://www.sciencedaily.com/releases/2008/10/081006184638.htm)

The positions of sunspots since the year 1874 are displayed in the so-called Maunder butterfly diagram shown in the top panel of Fig. 47.2. The lower panel gives the sunspot area as a percentage of the Sun's visible disk, with a typical value of up to 0.2%. The numbers at the bottom give the number of the solar cycle; now the Sun is beginning the new solar cycle 24, after a very deep and long solar minimum in 2008. Figure 47.3 presents a graph of the yearly average number of sunspots over the last 400 years which also exhibits the 11 ± 1 year cycle since 1750; before that time the number of sunspots was very small, during the Maunder minimum. The sunspot cycle has a faster rise to solar maximum and a slower decline from it (Hathaway and Wilson 2004). The sunspot number predicted for the maximum of solar cycle 24 in 2013 is like that shown by the arrow. The maximum sunspot number predicted for solar cycle 25 is even smaller. Thus the Sun's activity also varies on a timescale of a hundred years and more. The intensity of radio emission from the Sun at a wavelength of 10.7 cm, which can be received at the Earth's surface by a radio telescope, also varies markedly over 11 years; this quantity (known as F10.7) is often used as a proxy measure of solar activity.

The Sun's activity varies over this 11-year cycle. The brightness of the Sun, its irradiance, is $\sim$0.1% larger at solar maximum than at solar minimum even though sunspots absorb some light emitted below them. The solar flux variation over the solar cycle is approximately 10% in the ultraviolet, but varies by 10–100 times at X-ray wavelengths. Between these bands, at a wavelength of 28.4 nm, radiation is emitted by multiply ionized iron atoms (Fe XV). Figure 47.4 shows a sequence of 11 images of the Sun taken from the SOHO satellite in the extreme UV part of the spectrum at yearly intervals from 1996 to 2006; this composite figure shows clear evidence of the 11-year cycle of solar activity, with solar activity maximizing around the year 2001.

On September 1, 1859, Carrington was observing sunspots when one sunspot region suddenly brightened noticeably; this was the very first observation of a solar flare. The event occurred in the middle of a large geomagnetic storm which intensified a day later. That was the first evidence of a Sun–Earth connection,

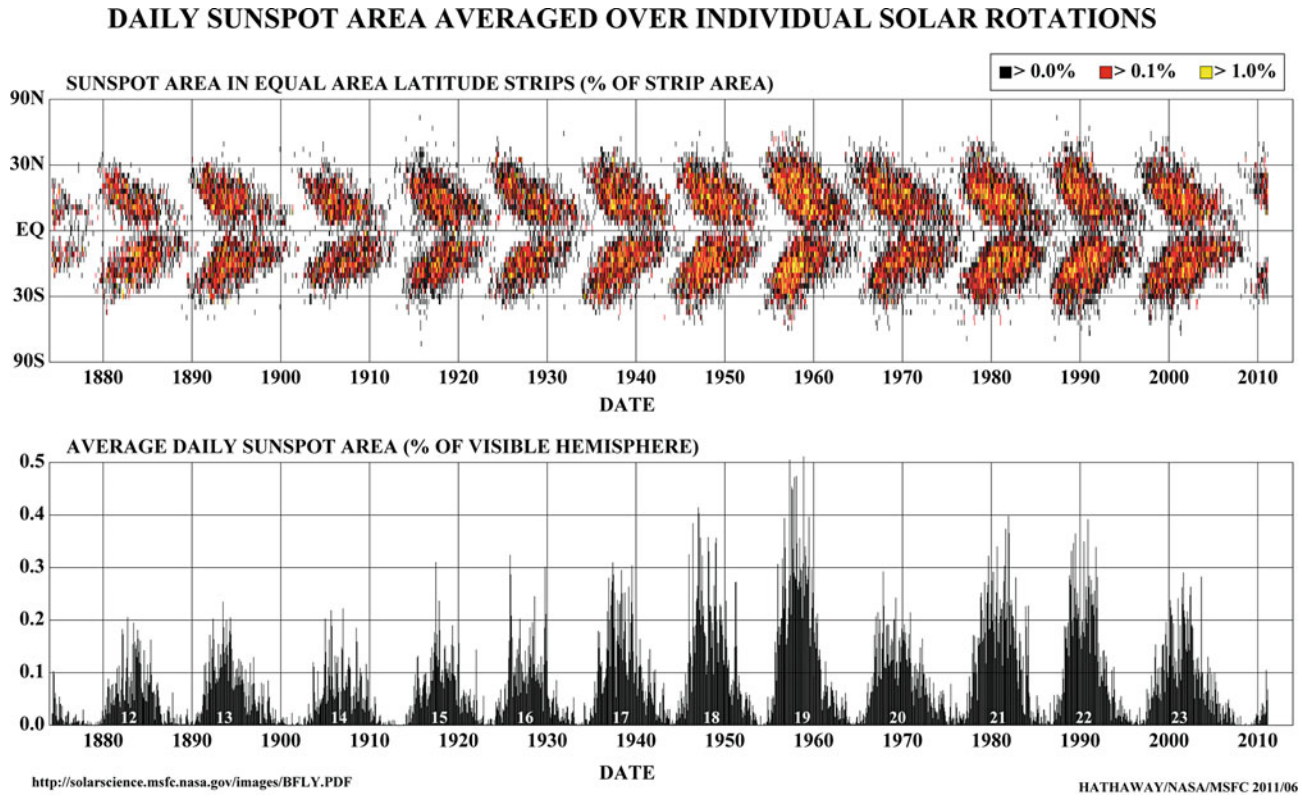


Fig. 47.2 The positions of sunspots (*upper panel*), and the area of the Sun covered by sunspots (*lower panel*) plotted since the year 1874 show the 11-year cycle of solar activity. (Courtesy of NASA, http://solarscience.msfc.nasa.gov/images/BFLY.PDF)

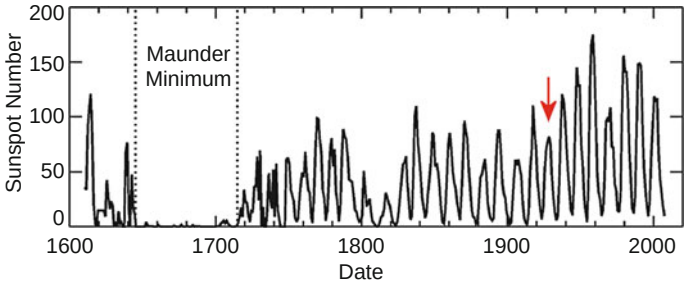


Fig. 47.3 The number of sunspots plotted since the year 1600; the 11-year solar cycle is clear since the early 1700s. (Courtesy of NASA, http://science.nasa.gov/science-news/science-at-nasa/2009/29may_noaaprediction/)

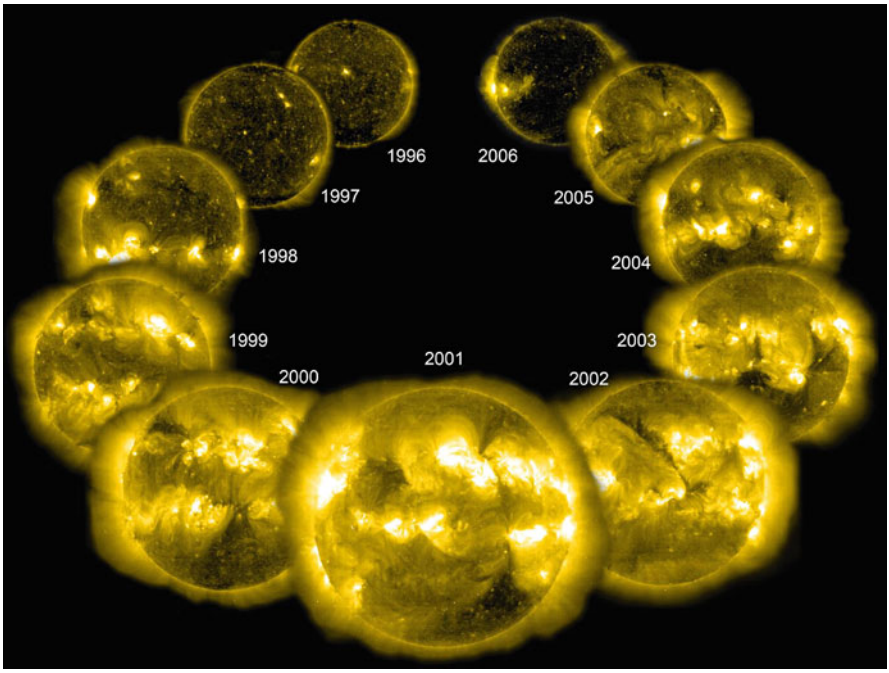


Fig. 47.4 Extreme ultraviolet images of the Sun (Courtesy of ESA and NASA, http://sci.esa.int/science-e/www/object/index.cfm?fobjectid=47710)

which led to the new research field now known as solar-terrestrial physics. As luck would have it, Carrington observed the largest solar flare ever; no solar flare observed since has been brighter. It is now believed that a solar flare happens as a result of a plasma instability which reconfigures the arched magnetic field lines in the low solar corona. This injects accelerated energetic electrons down to the chromosphere where they collide with atoms and emit X-rays. Nowadays, these X-rays are routinely detected by instruments aboard US geostationary satellites.

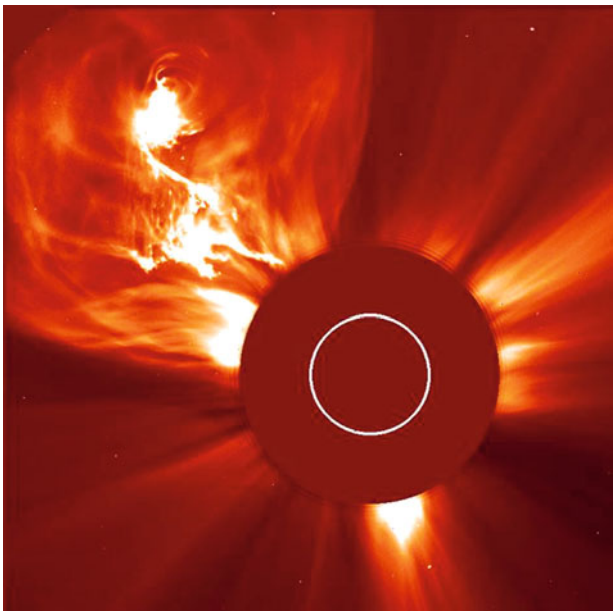


Fig. 47.5 A dramatic Coronal Mass Ejection (CME) event was produced by the Sun on January 4, 2002. (Courtesy of ESA and NASA, http://sohowww.nascom.nasa.gov/gallery/images/c2fireball.html)

Coronal Mass Ejections (CMEs)

Associated with the magnetic field reconfiguration near the Sun is the ejection of hot, dense plasma into the corona and beyond, into interplanetary space. Such a huge (up to approximately 10^{12} kg) high speed solar wind event carrying strong, twisted magnetic fields away from the Sun is termed a Coronal Mass Ejection (CME) event. Figure 47.5 is an image taken by the Large Angle and Spectrometric Coronagraph (LASCO) C2 instrument aboard the NASA/ESA Solar and Heliospheric Observatory (SOHO) spacecraft. In these images, direct sunlight is blocked by an occulting disk, so that only light emitted by the solar corona at heights greater than three solar radii is detected, in order to reveal structure in the high corona; the approximate size of the Sun is shown by the white circle in Fig. 47.5.

Dark areas in Fig. 47.5 represent coronal holes, where the low density solar wind is moving most rapidly away from the Sun along radial interplanetary magnetic field lines. The bright area at the bottom shows an arched magnetic field structure from the Sun's surface extending out to about four solar radii, and containing very hot plasma of high density. A more complex arched magnetic field region at the top left has exploded, creating an enormous volume of very hot, dense plasma ballooning out into interplanetary space and evolving as it does so. This is a dramatic example of a Coronal Mass Ejection (CME) event.

When such a CME is directed toward the Earth, it is called a halo CME; it compresses the Earth's magnetosphere, within which the Earth's dipolar magnetic

field is contained. Thus the magnetopause, the boundary of the magnetosphere (which is usually at about 10 Earth radii (R_E) upstream of the Earth on the sunward side) is compressed and moves inward to about 6 R_E. Then satellites in GEO – at a geocentric distance of 6.6 R_E – under noon time conditions will be directly exposed to the solar wind plasma. They have then to operate in a totally different plasma environment from the one for which they were originally designed to function.

This situation creates a geomagnetic storm (or, simply, a magnetic storm). Today it is said that solar activity creating a magnetic storm at the Earth is particularly "geoeffective." A CME takes from one to several days to travel from the Sun to the Earth; a CME having a faster solar wind speed generally produces a stronger magnetic storm. Magnetic storms are generally most prevalent in the declining phase of the solar cycle.

Sometimes, solar flares and CMEs occur together. At other times there can be a solar flare without a CME being generated, and sometimes there can be a CME without a solar flare. Large X-ray flares and large CMEs sometimes, but not always, occur simultaneously. This complicated situation makes the forecasting of space weather events (Bothmer and Daglis 2007) in the Earth's environment very difficult. However, space weather forecasting is an important subject, because of the potential damage not only to space assets but also to long electrical conductors on the ground, such as oil and natural gas pipelines, and also to ground-based electrical power grid systems. Pipelines can suffer increased corrosion and electrical grid systems can be disrupted by induced currents caused by space weather disturbances, which also create bright auroral displays at high latitudes. On the occasion of especially intensive events these displays are seen at much lower latitudes, as occurred during the so-called Carrington event in 1859.

The Earth's Atmosphere and Near-Space Environment

Where solar energy is absorbed, the temperature of that material thereby increases. Most of the Sun's visible radiation is absorbed by the Earth's surface, and the air at the surface is thus warmed. Ultraviolet radiation is absorbed by oxygen and ozone in the stratosphere, at heights from about 15–50 km, and the stratosphere is heated. As illustrated in Fig. 47.6, solar X-rays are absorbed in the thermosphere, at altitudes above 100 km; in the process, the gaseous molecules or atoms are split into positive ions and electrons, forming an electrically charged gas termed the "ionosphere." Radiation from the Sun (at the top) in different wave bands, shown in the center of Fig. 47.6, is absorbed at different levels of the atmosphere. On the right, typical height profiles of the density of oxygen molecules (green curve) and of oxygen atoms (yellow) are plotted; the ionospheric electron density profile (e^-) is shown in red. On the left is shown the neutral gas temperature for near solar minimum (yellow) and maximum (red) conditions.

In the upper atmosphere at 100 km altitude, there is only one ion-electron pair for every 10^8 electrically neutral molecules, whereas at ~300 km altitude for every

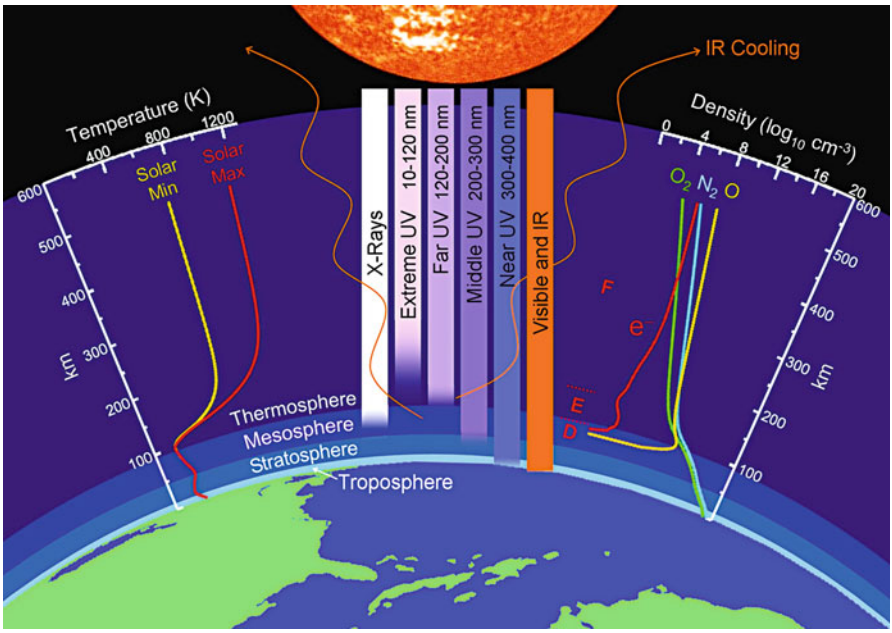


Fig. 47.6 Solar radiation of different wavelengths is absorbed at different levels in the atmosphere (Courtesy of NASA, http://science.nasa.gov/science-news/science-at-nasa/2010/15jul_thermosphere/)

ion-electron pair there are about a thousand neutrals. Thus, everywhere, the Earth's ionosphere is a weakly ionized plasma, as is evident from the right hand side of Fig. 47.6.

It is still uncertain whether the 0.1% solar irradiance variation over the solar cycle has any effect on varying the Earth's weather and climate on a decadal timescale. However, it is clear that the temperature of the stratosphere varies a little over the 11-year cycle, and other stratospheric features have long-term variations. The temperature of the thermosphere varies markedly from solar minimum to solar maximum, and so do many ionospheric parameters, the most important of which is the maximum electron density in the ionosphere. This feature occurs near 300 km altitude, at what is called the peak of the F-region.

Five Key Aspects of the Space Environment

Tribble (2003, 2010) has pointed out the five important features of the space environment as they affect most of the 800 or so active space missions today. We consider these features in some detail here.

The Residual Atmosphere in Low Earth Orbit (LEO) – Near-Vacuum Conditions

The Earth's atmosphere (having a total mass of 5×10^{18} kg) is kept close to the surface by the gravitational force of attraction on the molecules of gas in the atmosphere (mainly nitrogen, with oxygen and some minor species including argon, water vapor, carbon dioxide, and methane). This gravitational force arising from the Earth's mass (6×10^{24} kg) is often called the force of gravity. The gravitational force acting on an object of mass m kg is $9.8 \times$ m N, with $g = 9.8$ m/s^2 being the acceleration due to gravity at the Earth's surface. The mean global temperature of the atmosphere at the Earth's surface is 15°Celsius (or Centigrade), which translates to 288 K (Kelvin, or absolute).

The pressure (1,013 hPa, or 1.013 bar, at the Earth's surface) and the density (1.3 kg/m^3) both decrease exponentially with increasing altitude. At 15 km altitude the pressure and density have fallen to about one tenth of its sea level value, and at 30 km to one hundredth. At about 100 km altitude, where, conventionally, space is said to begin, there still remains about one millionth of the Earth's entire atmosphere. This is the pressure attained in quite a good vacuum system, and so conditions above 100 km altitude (the base of the thermosphere) are near-vacuum conditions. At altitudes above 120 km, the temperature can be as low as 500 K under solar minimum conditions and up to 2,000 K at solar maximum, as indicated in Fig. 47.6.

The density of the atmosphere at 500 km altitude is approximately 10^{-13} kg/m^3 at solar minimum and up to a hundred times bigger, i.e., 10^{-11} kg/m^3, at solar maximum. The thermosphere expands and contracts over the Sun's 11-year cycle of activity, causing the density at a particular altitude to change markedly. There is also a long-term change, possibly associated with increasing amounts of carbon dioxide in the troposphere (global warming – climate change – primarily due to the increasing rate of burning fossil fuels). The density at 400 km altitude at the solar minimum in 1997 was about 10% smaller than it was around 1986, and that was 10% less than 11 years earlier. During the pronounced solar minimum of 2008, the density at 400 km altitude was unexpectedly low, about 28% lower than in 1997.

Thermal Radiation – From the Sun and the Earth

In space there is no oxygen and there is no ozone to absorb the strong dose of ultraviolet radiation coming from the Sun. Therefore, a satellite's surface is bombarded by an intense flux of ultraviolet radiation. This degrades the surface materials to a greater or lesser extent; it can cause photoelectrons to be emitted (see ▶ Chap. 27, "Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing").

Direct thermal radiation from the Sun (typically 1,365 W/m^2), plus some solar radiation reflected back into space by clouds and by the Earth's surface, heat up a satellite's surface. There is an additional contribution to radiation that heats a satellite from the Earth-atmosphere system; this is terrestrial thermal infrared (IR)

radiation with a spectral peak at wavelengths near 10 μm (10×10^{-6} m); its contribution is about 240 W/m^2. When designing a satellite's thermal control system, all these sources have to be accounted for. Within the satellite heat is conducted from the sides that are illuminated to the dark side, from where it is radiated away into the blackness of space (at 2.7 K). There is no conduction of heat away from the satellite in the near-vacuum conditions of space. Surface materials with the desired absorptivity and emissivity characteristics are chosen so that the temperature inside the satellite lies within the correct operating temperature range.

A polar orbiting LEO satellite is eclipsed by the Earth for about 30 min during every 90 min orbit. This phenomenon exerts a strong thermal cycle on the satellite. A satellite in geosynchronous orbit (GEO) is eclipsed by the Earth for only about 45 min during the maximum eclipse period in the equinoxes; for most of the time there are no eclipses at all. The differential expansion of illuminated and dark satellite metallic surfaces is another effect that has to be considered.

Plasma – The Electrically Charged Gas, the Ionosphere

During the daytime, different wavelengths of solar extreme ultraviolet (EUV) and X-ray radiation cause the formation of different ionospheric layers at different heights (see Fig. 47.6). At heights above about 170 km, the ionospheric plasma moves upward to create a long lasting F-region peak. Above that the plasma density falls off with increasing height. During the night, the electrons collide with ions and recombine to form neutral molecules or atoms, the more so at lower heights. Because the flux of solar EUV and X-radiation is much larger at solar maximum than at solar minimum, there is a marked variation of ionospheric parameters over the solar cycle.

Energetic Charged Particles – Van Allen Radiation Belts, Solar Protons, and Cosmic Rays

In space we speak of radiation – that means not only electromagnetic radiation but also fluxes of energetic charged particles which constitute a radiation hazard to astronauts (especially if they are carrying out ExtraVehicular Activities, EVAs) and to electronic equipment on satellites. The most energetic charged particles are galactic cosmic rays (GCRs) coming through interstellar space from beyond our Milky Way galaxy, ions with energies exceeding 1 GeV (10^9 eV). (Here, 1 eV is the energy gained by an electron when it is accelerated through a potential difference of 1 V; one electron Volt [1 eV] is equivalent to a thermal energy of ~11,000°K.) Solar energetic particle events, also termed solar proton events (SPEs), have individual ion energies of more than 1 MeV (10^6 eV); these energetic protons come from active regions on the Sun. During intense solar flares, the fluxes of these particles increase dramatically. Almost all (about 95%) of the ions are protons, with some doubly charged helium ions and some highly charged ions of heavier elements.

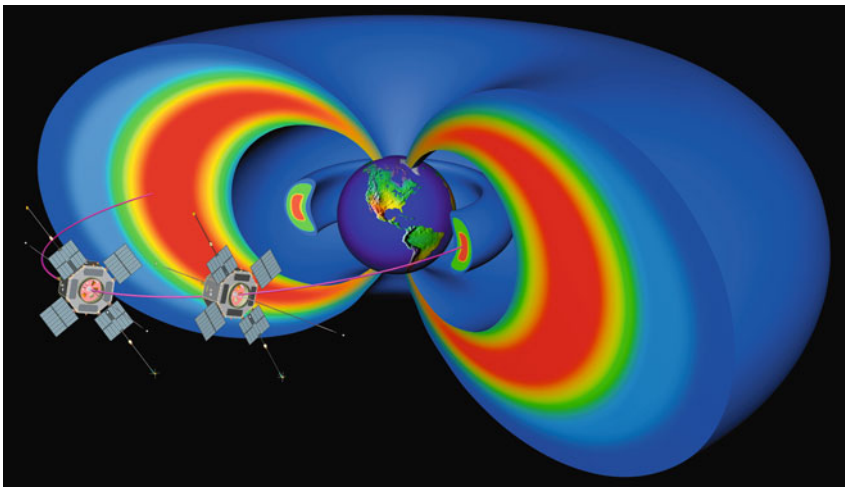


Fig. 47.7 The Radiation Belt Storm Probes (RBSP) were launched in 2012 by NASA to study the Van Allen Belts (Courtesy of NASA, http://rbsp.jhuapl.edu/gallery/artRender/pages/artRender_01.php)

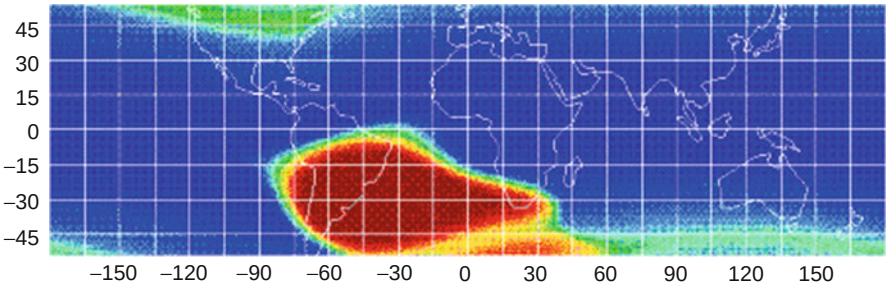


Fig. 47.8 The South Atlantic geomagnetic anomaly, shown in *red* (Courtesy of NASA, http://imagine.gsfc.nasa.gov/docs/ask_astro/answers/961004.html)

At lower energies, 1 keV (10^3 eV) charged particles from the ionosphere and from the solar wind are accelerated up to about 1 MeV by complex plasma wave–particle interactions inside the Earth's magnetosphere to create the Van Allen radiation belt electrons and ions (being mainly – 95% – protons). The Van Allen belts occur in two doughnut shaped regions around the Earth, as shown in Fig. 47.7; the strongest fluxes shown in red occur in the inner and outer zones. The actual fluxes of Van Allen charged particles can vary from typical values by up to two orders of magnitude, both up and down, from day to day.

The fluxes of energetic ions and electrons lost from the Van Allen radiation belts of the Earth's magnetosphere are largest in the vicinity of the South Atlantic geomagnetic anomaly. This is a region of reduced geomagnetic field strength which lies to the East of Brazil; it is shown in red in Fig. 47.8.



Fig. 47.9 Objects in orbit around the Earth which are tracked from the ground (Courtesy of ESA, http://www.esa.int/SPECIALS/ESOC/SEMN2VM5NDF_mg_1_s_b.html)

Micro-meteoroids and Orbital Debris

Strictly speaking, this subject lies outside the topic of space weather; however, it is valuable to mention it here. The chapter on space debris provides additional useful information on this subject.

Very many tiny dust particles of the cometary material (from which the solar system was originally formed) continuously enter the upper atmosphere from space. Known as micro-meteoroids, the total mass of these primordial particles arriving per day at speeds of about 10 km/s or more is about a hundred metric tons (10^5 kg). The larger particles (each with a mass of a fraction of 1 g (10^{-3} kg) up to 1 g) typically burn up at approximately 100 km altitude, emitting a flash of light – this is the explanation for how a meteor (a "shooting star" or a "falling star") is created. Even larger particles (having masses more than 1 g and up to a significant fraction of 1 kg or even more) do not burn up, but come down to the Earth's surface as meteorites.

This material adds to the significant amount of space debris (2,000 metric tons) already present in the thermosphere. Such debris largely consists of several thousand dead satellites and numerous spent upper stage rockets, plus various other parts and materials. Space debris has been created in situ by exploding rocket fuel tanks (due to the small amounts of fuel left in them), exploding electrical batteries or radioisotope thermoelectric generators, or by spent rockets or satellites colliding with other pieces of space debris; where the larger pieces of space debris are to be found is indicated in Fig. 47.9. Most of these objects are in LEO, a good number are

in GEO, and some are in other orbits; the objects are exaggerated in size in order to make them visible in the figure.

How the Space Environment Affects Satellite Operations

Here we consider the many effects which space mission planners, space systems engineers, and space instrument designers must consider – and the several possible problems that must be resolved – before a space mission is attempted. Otherwise, premature failure may result, with considerable human efforts expended in vain and a sizable sum of money wasted.

Near-Vacuum Conditions

Even though the atmosphere is very thin in LEO, it still exerts a drag force on a satellite moving through it at high speed and on other particles in orbit. The magnitude of this force is proportional to the density of the air, the cross-sectional area which the satellite or particle presents, its drag coefficient (note that satellites are not normally as streamlined as sports cars are!), and the square of its orbital velocity, which is large, approximately 8 km/s. Under solar maximum conditions the drag force acting on a satellite at 500 km altitude is up to a hundred times larger than at solar minimum. For a satellite orbiting near 300 km altitude the drag force is an order of magnitude larger than at solar minimum. The consequence of this drag is that the perigee of all LEO satellites is reduced more quickly near solar maximum, and so the lifetime of the mission is shorter than it would be under solar minimum conditions. For the International Space Station (ISS) orbiting at, say, 400 km altitude, the consequence is that at solar maximum much more propellant has to be transported to the ISS to boost the altitude of its orbit than at solar minimum.

A satellite surface glows when neutral atmospheric atoms and molecules impinge on its surface. This is because of its high velocity, about 8 km/s with respect to the gaseous atoms and molecules; on impact this gives the gas particles sufficient energy (about 5 eV) to react chemically with its surface, and cause a glow in the visible part of the spectrum. This glow may interfere with the performance of an optical experiment, such as carried out by a telescope, if the beam of the instrument goes through the glowing region.

The major constituent of the air at a few hundred kilometer altitude is atomic oxygen, formed by the breakdown of molecular oxygen in the thermosphere by energetic solar ultraviolet and X-ray photons. The number density of these atoms can vary by a factor of 10 (i.e., the density lies between $10^{13}/m^3$ and $10^{14}/m^3$) near 400 km altitude. These oxygen atoms are highly reactive, chemically; for example, they will oxidize a front-silvered mirror, turning it black. Further this action changes the thermal characteristics of the surface, and even weakens it physically.

In space, because of the near-vacuum conditions, conventional lubricants do not work well at all. A moving joint, such as in the elbow action required to deploy an

arm, can seize up. Therefore, special lubricants have been invented. Satellite and spacecraft surface outgas when first in space, that is to say particles and molecules of gas stuck to the surface are released into space. Thus the local pressure around the satellite builds up during the first few days in orbit. Such detached particles (even though the satellites were put together in a clean room on Earth) can stick onto the surface of an optical instrument, say a mirror, or a lens, and degrade its performance below optimum. Even a partial mono-molecular layer can ruin the performance of a front-silvered mirror. The thrusters which are used to control the orientation of a satellite or spacecraft will introduce gas in the vicinity and may, similarly, reduce the performance of an optical instrument.

There are no known effects of the residual Earth's atmosphere on the performance of positioning satellites in inclined elliptical medium Earth orbits (MEO) or on communications satellites in MEO or GEO. Neither are there any such effects acting on spacecraft operating outside the Earth's magnetosphere.

Thermal Radiation

The surface materials covering the satellite of spacecraft may have their absorptivity properties and their emissivity characteristics changed by bombardment by energetic charged particles. Both these properties may also be altered by exposure to the strong flux of solar ultraviolet radiation.

Plasma

High Frequency (HF, 3–30 MHz) radio communications under the ionosphere are strongly affected by SPE events, which lead to communications blackouts, especially in the polar regions. For radio communications between the ground and a satellite, and vice versa, the radio frequency used must exceed the maximum plasma frequency of the ionosphere. This is the F-region peak plasma frequency; its value is usually between 10 and 30 MHz, but it varies considerably with geographic location, time of day/night, phase of the solar cycle, geomagnetic activity, and several other quantities.

Small-scale spatial and temporal variations of the ionospheric plasma density cause scintillations, that is to say rapid amplitude and phase changes of radio signals received on the ground from spacecraft/satellites. These are most marked in equatorial and auroral regions, especially under conditions of enhanced solar and geomagnetic activity. Such effects require corrections to be made to GPS navigation satellite signal (1–2 GHz, L band) delays, and to spaceborne radar altimeter observations. The fact that the ionosphere is a bi-refringent medium has to be considered; this is to say that there are two values for the refractive index at each radio frequency, due to the fact that two wave modes (ordinary (O) and extraordinary (X)) propagate.

At GEO, where the ambient plasmaspheric density is generally small, below 10^7 m^{-3} (i.e., below 10 cm^{-3}) or in the interplanetary medium (when the

solar wind speed is unusually large so that the magnetopause is compressed from its usual upstream distance of 10 Earth radii (R_E) from the Earth's center to about 6 R_E), spacecraft charging can cause problems. This happens when sunlight incident on a spacecraft surface creates photoelectrons (by the photoelectric effect) so that the surface is left charged positively; the flux of thermal electrons from the surrounding plasma is insufficient to neutralize this charge when their density is low. Between the sunlit and the dark sides of the spacecraft, a strong electric field develops. If this field becomes large enough, arcing occurs – that is to say a spark, a mini lightning flash, an electrical discharge – takes place through the interior of the spacecraft where the sensitive electronic equipment is placed. If the discharge passes through a circuit it "zaps" it, so that it no longer operates.

Energetic Charged Particles

Energetic charged particles of energies exceeding a fraction of 1 MeV constitute a serious radiation hazard not only to astronauts but also to all forms of integrated circuits contained in electronic equipment aboard a satellite in LEO or GEO or on more distant roving spacecraft. This is because energetic ions cause extra currents to flow in these circuits, which can change the binary state of a computer memory; such a flip is called a single event upset (SEU). It may result in a phantom command to the satellite, which is hard for ground controllers to understand. The memory either may recover from the SEU, or the damage may be permanent, in which case it is called a single event latch-up (SEL).

Figure 47.10 shows observations made in geostationary orbit of the huge solar proton event which occurred on July 14, 2000, the so-called Bastille Day event. The proton data are shown for three energy thresholds; there is a three, or even four, orders of magnitude flux increase in a short time, and the enhanced fluxes persist for several days.

Deep dielectric charging and discharging occurs when 1 MeV "killer" electrons from the Van Allen belts penetrate the satellite walls and deposit charge on the insulating (dielectric) material of electronic circuit boards. The dielectric material can break down, causing electrical shorts unless the boards are carefully grounded.

Associated with a halo CME coming directly toward the Earth, the appearance of a coronagraph image (as shown in Fig. 47.5) suddenly changes. It becomes covered in white specks, sometimes called "snow." Each speck is produced when an energetic charged particle hits the charge-coupled device (CCD) detector in the instrument.

Another effect is that the performance of a satellite's solar array degrades with time due to solar proton bombardment. For example, the efficacy of the solar array aboard SOHO fell from 100% to 82% over 10 years. A further energetic ion effect is that of sputtering; when an ion of about 1 MeV or more hits a spacecraft surface, it kicks a few atoms from the surface. Over time a thin coating could be removed by this process, or the surface material could be weakened.

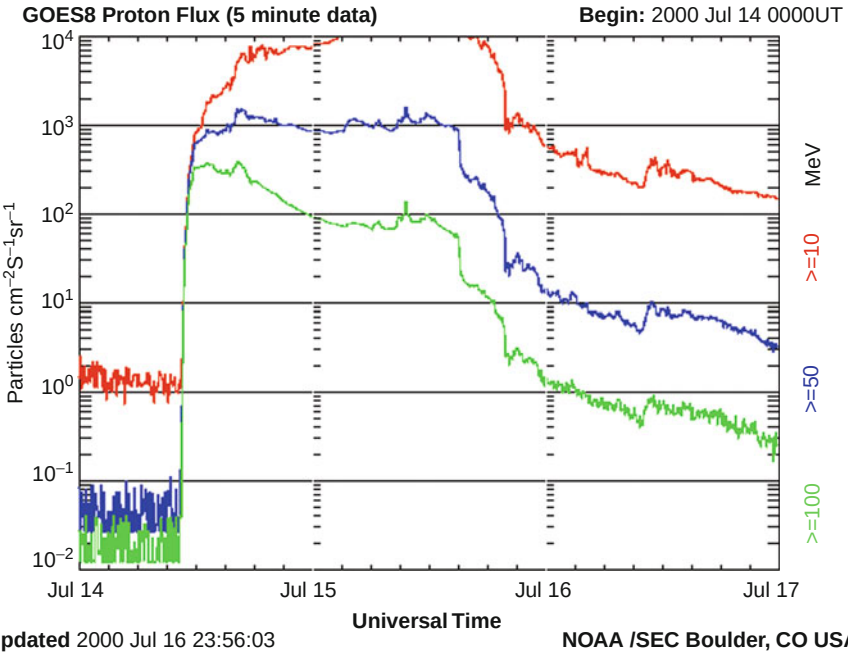


Fig. 47.10 Solar proton fluxes at GEO during a huge geomagnetic storm (Courtesy of NASA, http://pwg.gsfc.nasa.gov/istp/events/2000july14/)

Micro-meteoroids and Space Debris

Particles with diameters exceeding 10^{-2} m (i.e., 1 cm), and moving at very high speed (10 km/s or even more) with respect to a satellite in LEO, pose an immediate danger to its integrity. This is because they can pass right through its outer case, be it made of metal or some composite material. Fortunately, the 21,000 such particles shown in Fig. 47.9 can be tracked from the Earth by radar and by optical telescopes, and their orbits calculated. If a collision appears to be imminent, the orbit of, for example, the ISS or a satellite having thrusters aboard can be changed a little to avoid a collision happening.

Of greater danger is the larger flux of particles with diameters between 0.1 and 1 cm whose orbital parameters cannot be known from observations made on the ground. There are believed to be about 6×10^5 such particles in LEO which can, for example, crack the glass covers of solar cell arrays or penetrate metallic foils. However, they can be stopped using honeycomb structures which act like the fenders (bumpers) on cars. Smaller particles whose diameters are a fraction of 1 mm can damage glass (optical) surfaces. The orbits of all debris particles tend to become more circular due to atmospheric drag acting at perigee. Because this drag is largest near solar maximum, this is the time in the solar cycle when most space debris is removed.

Protection of Application Satellites Against Space Weather Effects

Application satellites are complex and expensive spacecraft that cost many millions of dollars to manufacture and launch. Clearly, the protection of these assets against space debris, coronal mass ejections, energetic charged particles, increased thermospheric gas densities, and other hazardous conditions arising from space weather disturbances makes a good deal of sense. Maneuvering spacecraft to avoid collision with space debris is one possibility. Improved tracking and sharing of data on possible conjunctions of operational spacecraft is currently being pursued. There is a global network of solar observatories as well as of solar observing satellites that monitor the Sun 24 h a day, 7 days a week. When halo CMEs or other potentially destructive events are observed, that information is communicated to satellite operators. They thus have time to "power down" satellites and take other preventive measures that give spacecraft a much greater chance of surviving these events. To date, only a small number of satellites has been lost due to solar events or to collisions with debris, but something like the Carrington event, should it occur again, could have very severe effects on space assets.

Conclusion

In this chapter we have outlined the effects of

1. The residual terrestrial atmosphere which causes a drag force to act on a satellite, and of the highly reactive oxygen atoms present
2. The heating of a satellite due both to visible radiation from the Sun and to infrared radiation for the Earth
3. The ionospheric plasma which affects the strength of radio signals propagating from the ground to the satellite and vice versa, together with scintillations
4. Energetic charged particles – both ions and electrons – from the cosmos, from the Sun, and from the terrestrial Van Allen belts, which can penetrate into the heart of a satellite and damage its solid-state electronic circuits
5. Micro-meteoroids and space debris which, if large enough (greater than 1 mm in size), can physically damage the structure of a satellite.

The efficacy of all of these effects varies markedly through the 11-year cycle of solar activity. Fortunately, the forecasting of all such effects is improving as our understanding of the very many different processes which are involved improves with the continuing research being undertaken.

Cross-References

- ► Electromagnetic Radiation Principles and Concepts as Applied to Space Remote Sensing
- ► Lifetime Testing, Redundancy, Reliability and Mean Time to Failure
- ► Orbital Debris and Sustainability of Space Operations

References

V. Bothmer, I.A. Daglis, *Space Weather: Physics and Effects* (Springer Praxis, Berlin/New York, 2007), 438 pp.
M.J. Carlowicz, R.E. Lopez, *Storms from the Sun: The Emerging Science of Space Weather* (Joseph Henry Press, Washington DC, 2002), 220 pp.
I.A. Daglis (ed.), *Space Storms and Space Weather Hazards* (Kluwer, The Netherlands, 2001), 482 pp.
I.A. Daglis (ed.), *Effects of Space Weather on Technology Infrastructure*. NATO Science Series (Springer, Dordrecht, 2004), 176 pp.
J.W. Freeman, *Storms in Space* (Cambridge University Press, Cambridge, 2001), 139 pp.
D.H. Hathaway, R.M. Wilson, What the sunspot record tells us about space climate. Solar Phys. **224**, 5–19 (2004)
J. Lilenstein (ed.), *Space Weather; Research towards Applications in Europe*. Astrophysics and space science library, vol. 344 (Springer, Dordrecht, 2007), 330 pp.
M. Moldwin, *An Introduction to Space Weather* (Cambridge University Press, Cambridge, 2008), 134 pp.
M.J. Rycroft, The plasma and radiation environment in Earth orbit, in *Encyclopedia of Aerospace Engineering* (Wiley, Chichester, 2010). Chapter eae323
C.J. Schrijver, G.L. Siscoe (eds.), *Heliophysics: Plasma Physics of the Local Cosmos* (Cambridge University Press, Cambridge/New York, 2009), 435 pp.
A.K. Singh, D. Siingh, R.P. Singh, Space weather: physics, effects and predictability. Surv. Geophys. **31**, 581–638 (2010)
P. Song, H.J. Singer, G.L. Siscoe (eds.), *Space Weather* (American Geophysical Union, Washington DC, 2001), 440 pp. Geophysical Monograph 125
A.C. Tribble, *The Space Environment: Implications for Spacecraft Design*, 2nd edn. (Princeton University Press, Princeton, 2003), 248 pp.
A.C. Tribble, The effect of the space environment on spacecraft technologies, in *Encyclopedia of Aerospace Engineering* (Wiley, Chichester, 2010). Chapter eae568

Further Reading

www.spaceweather.eu
www.spaceweathercenter.org
www.swpc.noaa.gov
www.spenvis.oma.be
http://royalsociety.org/summer-science/2011/aurora-explorer
http://sohowww.nascom.nasa.gov
http://goespoes.gsfc.nasa.gov/
http://sdo.gsfc.nasa.gov/data/
http://stereo.gsfc.nasa.gov
http://swri.org/9what/releases2011/solarwind.htm
http://www.esa-spaceweather.net
http://www.solarstorms.org
http://www.n3kl.org/sun/noaa.html
http://www.exploratorium.edu/spaceweather/

Section 7

Appendices

The World's Launch Sites

A1

Arthur N. Guest and Joseph N. Pelton

Contents

Introduction 1197
Australia 1198
Brazil 1199
China 1199
Europe 1200
India 1200
Israel 1201
Italy 1201
Japan 1201
Russian and CIS Launch Sites 1202
South Korea 1203
United States of America 1203
Multinational Sites 1205

Introduction

There are over three dozen launch sites around the world. With the growth of the newly emerging spaceplane transportation and space adventures business, there are also a growing number of "spaceports" designed for commercial liftoffs and landings. While there are an increasing number of commercial spaceports in the United States, there are also numerous other sites under consideration around the world in locations such as Singapore, the United Arab Emirates, Malaysia, and various sites in Europe.

A.N. Guest
International Space University, San Francisco, USA
e-mail: guest.arthur@gmail.com

J.N. Pelton (✉)
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: joepelton@verizon.net

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_88, © Springer Science+Business Media New York 2013

Today, perhaps the busiest launch sites for various types of spacecraft are at Cape Canaveral in California (Eastern Range) and Vandenberg (Western Range), the Baikonur and Plesetsk Cosmodromes in Russia, the Kourou launch facility of the French Space Agency (CNES) and the European Space Agency in French Guiana, the Kagoshima and Tanegashima launch facilities in Japan, the Jiuquan and Xichang launch ranges in China, and the Satish Dhawan launch facility in India. This concentration on official governmental launch sites will undoubtedly change as more and more commercial spaceports evolve in the coming years. Many of the commercial spaceports being licensed in the United States by the FAA-AST are for horizontal takeoff and landing and are intended only for the operation of suborbital flights and thus do not support commercial launch of application satellites to the Earth orbit.

For years, the space programs of the United States and Russia were comparable in size and by far the world's largest space operations. But this has changed as European (both national and regional), Chinese, Japanese, Indian, and Canadian space programs have matured. Further, there are many emerging space programs in Australia, Argentina, Brazil, Iran, Israel, North and South Korea, Pakistan, Taiwan, and the Ukraine. The following provides background information on launch sites around the world that provide launch operations or have done so in the past. Not all licensed commercial spaceports are included, particularly those that only provide suborbital flights.

Australia

Site Name: Spaceport Australia
Location: Woomera, Australia (Latitude 31.1°S Longitude 136.6°E)
Launch Vehicles Supported: Kistler K-1 (canceled)
Originally known as the Woomera Rocket Range, this site was originally created to support British and Australian missile operations. In 1967, WRESAT, an Australian science satellite, was launched on a US Redstone rocket from Woomera. The original launchpads have been dismantled, but the site is being redeveloped as Spaceport Australia to support future commercial launch opportunities. Spaceport Australia was to be used to support Kistler Aerospace's K-1 launch vehicle. The development of the K-1 rocket was canceled in 2007. The Australian Space Council is currently supporting future launch operations from this site. Spaceport Australia is located in the desert and well positioned for polar orbit launches.
Site Name: Asia Pacific Space Center (proposed)
Location: Christmas Island, Australia (Latitude 10.4°S Longitude 105.7°E)
Launch Vehicles Supported: Aurora (proposed)
The Asia Pacific Space Center (APSC) was proposed to support the Aurora rocket, a new Russian/Australian launch vehicle derived from the Soyuz launch vehicles. The site has been proposed to be located in the Indian Ocean on Christmas Island.

Brazil

Site Name: Alcantara Launch Center
Location: Alcantara, Brazil (Latitude 2.3°S Longitude 44.4°W)
Launch Vehicles Supported: VLS-1 (proposed)
Located near Sao Luis on the Atlantic Ocean coastline, this site supports launches by the proposed Brazilian VLS-1 launch vehicle as well as the Sonda 3 and Sonda 4 sounding rockets. The development of the VLS-1 launch vehicle has been delayed since an accident on the launchpad in 2003. The potential of using the Alcantara Launch Center to support other commercial launch vehicles has been proposed although no immediate plans exist. The launch center's location (2° south of the Equator) offers it substantial advantage for geosynchronous satellite launches.

China

Site Name: Jiuquan Satellite Launch Center (JSLC)
Location: Gobi desert, Inner Mongolia (Latitude 40.6°N Longitude 99.9°E)
Launch Vehicles Supported: Long March 2C/2D/2F & Long March 4B/4C
On April 24, 1970, the People's Republic of China became the fifth nation to launch an artificial satellite into Earth orbit. This satellite was named Mao-1 in honor of Chairman Mao Tse Tung and this spacecraft was launched into orbit by a Long March 1 vehicle from the Jiuquan Satellite Launch Center in Inner Mongolia. Jiuquan Satellite Launch Center (also known as Shuang Cheng Tzu) was the first launch complex to be built by China and it is located north of Jiuquan City in the Gobi desert 1,600 km west of Beijing, China. Jiuquan has been limited to south-eastern launches into 57°–70° orbits in order to avoid overflying Russia and Mongolia. Jiuquan is used for recoverable Earth observation and microgravity missions, but due to the site's isolated geographical location, most Chinese commercial flights take off from other spaceports.

On October 15, 2003, a Long March 2F rocket with the spacecraft Shenzhou 5 was launched from the Jiuquan Satellite Launch Center. Inside the spacecraft was Yang Liwei, China's first astronaut. The flight made China the third nation to send a person into space.

Site Name: Xichang Satellite Launch Center (XSLC)
Location: Xichang City, China (Latitude 28.3°N Longitude 102.0°E)
Launch Vehicles Supported: Long March 2C & Long March 3A/3B/3BE/3C
This site offers better access to geostationary orbits than Jiuquan since it is much closer to the equator. It was built 65 km north of Xichang City in 1978 but its first orbital launch was not until 1984. Xichang launches Long March 2C and Long March 3-series launch vehicles. A number of commercial communications satellites on behalf of a number of operators have been launched from this site.

Site Name: Taiyuan Satellite Launch Center (TSLC)
Location: Shanxi Province, China (Latitude 37.5°N Longitude 112.6°E)

Launch Vehicles Supported: Long March 2C/2D and Long March 4B/4C
This center was started in 1968 as a test base for missiles and rockets that were too large to fly from Jiuquan. The Taiyuan Satellite Launch Center, also known as Wuzhai, opened its single space launchpad in 1988 for launching Long March 4 launch vehicles. These launches have been primarily for placing remote sensing, meteorological and reconnaissance satellites into polar orbit. In 1996 and 1997, Long March 2C rockets carried Iridium satellites into orbit.
Site Name: Wenchang Satellite Launch Center (WSLC)
Location: Hainan Island, China (Latitude 19.7°N Longitude 111.0°E)
Launch Vehicles Supported: Long March 5 (proposed)
The Wenchang Satellite Launch Center is being constructed to support the Long March 5 launch vehicles. The site was specifically chosen for its proximity to the equator to support launching satellites into geostationary orbit. The site is expected to be completed in 2013 in order to support launch of Long March 5 in 2014.

Europe

Site Name: Guiana Space Center (Centre Spatial Guyanais)
Location: Kourou, French Guiana (Latitude 5.2°N Longitude 52.8°W)
Launch Vehicles Supported: Ariane 5, Soyuz, Vega (proposed)
On December 24, 1979, the European Space Agency became the seventh entity to launch an artificial satellite into Earth orbit. The satellite was named CAT and it was launched by an Ariane 1 rocket from the Guiana Space Center (or Centre Spatial Guyanais, CSG). CSG is owned by the French national space agency, CNES. It is used by the European Space Agency (ESA) and its commercial space launch arm Arianespace to launch ESA's Ariane 5 launch vehicles. CSG is one of the most favorable sites for launching satellites into geostationary orbit since it is only 5° north of the equator. French Guiana's coastline permits launches into both equatorial and polar Sun-synchronous orbits with inclinations up to 100.5°. Hundreds of sounding rockets and balloons and space satellites have been launched from Centre Spatial Guyanais. Most recently CSG has been expanded to support the Russian Soyuz launch vehicle, which was first launch from Kourou on October 21, 2011. Construction is currently underway to enable the support of launches of the Vega launch vehicle, which are planned to begin in early 2012.

India

Site Name: Satish Dhawan Space Center (SHAR)
Location: Sriharikota Island, India (Latitude 13.9°N Longitude 80.4°E)
Launch Vehicles Supported: PSLV and GSLV
On July 18, 1980, India became the eighth nation to launch an artificial satellite into orbit. The satellite was named Rohini 1. It was launched on the Satellite Launch Vehicle (SLV) from Satish Dhawan. The Satish Dhawan Space Center is located on Sriharikota

Island on India's east coast state of Andhra Pradesh. In India, all space activities, launcher development and satellite design and manufacture, is the responsibility of the Indian Space Research Organization (ISRO) and its commercial partners and subsidiary organizations. ISRO has launched many space research and application satellites on its PSLV (Polar SLV) and GSLV (geostationary SLV) vehicles. This launch site, that is only 14° north of the Equator, is well suited for geosynchronous launches.

Israel

Site Name: Palmachim Air Force Base
Location: Negev Desert, Israel (Latitude 31.5°N Longitude 34.5°E)
Launch Vehicles Supported: Shavit
On September 19, 1988, Israel became the ninth nation to launch an artificial satellite into orbit. The satellite was named Horizon 1 (Ofeq 1 in Israeli). It was launched on Shavit rocket from Israel's Palmachim Air Force Base. This location is still Israel's main spaceport. This site is south of Tel Aviv and is near the town of Yavne in the Negev Desert. Launches from Palmachim are done into a retrograde orbit in order to allow all launches to head west toward the Mediterranean Sea as opposed to heading east toward neighboring states.

Italy

Site Name: Broglio Space Center (San Marco Platform)
Location: Offshore near Malindi, Kenya (Latitude 2.9°S Longitude 40.3°E)
Launch Vehicles Supported: None (site not operational)
Italy's San Marco Range (renamed the Broglio Space Center in 2004) consists of a pair of platforms that float in Formosa Bay 3 miles off the coast of Kenya in the Indian Ocean. The San Marco platform is the launchpad and the Santa Rita platform holds the firing control blockhouse. The range started firing sounding rockets in 1966 and eight satellites were boosted to space from there by 1976. Italy used the offshore platform for another launch in 1988. This platform is not currently operational but could be employed for future launches. The position close to the equator makes this well suited for geosynchronous launches.

Japan

Site Name: Uchinoura Space Center (Kagoshima Space Center)
Location: Kagoshima, Kyushu Island, Japan (Latitude 31.2°N Longitude 131.1°E)
Launch Vehicles Supported: M-V (retired)
In February 1970, Japan became the fourth nation to launch an artificial satellite into Earth orbit. The satellite was named Ohsumi. This launch was on a Lambda 4S-5 rocket, and the Kagoshima site, which was constructed in 1962, was used to

support this initial Japanese orbital flight. The site is built on the southern tip of Kyushu Island to support an eastward launch over the Pacific Ocean. This site was initially employed for the launch of sounding and meteorological rockets, then later used for scientific and application satellite launches. Japan's first six satellites were launched from the Kagoshima Space Center which has been called the Uchinoura Space Center since 2003 and the establishment of JAXA. The large M-V orbital rocket was first launched there in 1997. Over 20 orbital launches have initiated from this site and much larger number of suborbital flights. The site is currently used to launch sounding rockets.

Site Name: Tanegashima Space Center (TNSC)
Location: Tanegashima, Japan (Latitude 30.4°N Longitude 131.0°E)
Launch Vehicles Supported: H-IIA and H-IIB

Japan's space agency (JAXA) operates the Tanegashima Space Center orbital launch site on the southeastern tip of Tanegashima Island some 1,050 km southwest of Tokyo. The complex's northern Osaki Launch Site can support H-IIA and H-IIB rockets. It also has static test facilities for liquid-fuel rocket engines. This site also houses the H-II Range Control Center.

Russian and CIS Launch Sites

Site Name: Baikonur Cosmodrome
Location: Tyuratam, Kazakhstan (Latitude 45.6°N Longitude 63.4°E)
Launch Vehicles Supported: Proton, Strela, Dnepr, Zenit, Rockot, and Cyclone 2

Sputnik 1, the world's first artificial satellite was launched on October 4, 1957 from this site. The first development of this launch site was in the early 1950s in the Baikonur/Tyuratam area of Kazakhstan in central Asia. The launchpad from which Sputnik 1 launched, as well as from which Yuri Gagarin, the first man in space, was launched, was constructed in 1955. It is in an isolated area about 370 km southwest of Baikonur. It was renamed by the Kazakhstan Government after the close town of Tyuratam in the 1990s. However, the global space community still refers to it as Baikonur Cosmodrome. Baikonur is a large cosmodrome with nine launch complexes encompassing 15 launchpads. All of Russia's manned space flights and interplanetary probes are launched from the Baikonur Cosmodrome. Baikonur is the only cosmodrome capable of launching Proton, Zenit, Energia, and Tsyklon SL-11 space rockets. Launches headed due east would be the most efficient launch path, but this launch orientation is not used from Baikonur because of concerns that lower stages of the rockets might fall into China.

Site Name: Plesetsk Cosmodrome
Location: Arkhangelsk Oblast, Russia (Latitude 62.8°N Longitude 40.1°E)
Launch Vehicles Supported: Kosmos 3M, Rockot, Soyuz, Start-1, Angara

The first of several pads were constructed at the Plesetsk Cosmodrome starting in 1957 to support both R7 and A-class intercontinental missile tests. The pads and various classes of ICBM rockets were moved to active duty in 1960. Initially, the Plesetsk Cosmodrome was the world's most utilized launch site. In time, the

launches moved to the Baikonur/Tyuratam Cosmodrome that was build to support newer launch systems. Today, Plesetsk can support Kosmos 3M, Rockot, Angara, and other launch vehicles. Plesetsk Cosmodrome is located in Russia at 62.8°N and 40.1°E, which allows the launch of communications satellites and surveillance satellites to polar and highly elliptical Molniya orbits. The Plesetsk Cosmodrome is currently being renovated and expanded for possible future operations.
Site Name: Svobodny Cosmodrome (not operational)
Location: Amur Oblast, Russia (Latitude 51.4°N Longitude 128.3°E)
Launch Vehicles Supported: Start-1 and Rockot
This is the most recently constructed Russian cosmodrome, started in 1996. This cosmodrome was built on a former missile site called Svobodny-18 and lies about 100 km away from the Chinese border. This site can support the launch of Start and Rockot launch vehicles.

South Korea

Site Name: Naro Space Center
Location: Goheung County, South Jeolla (Latitude 34.4°N Longitude 127.5°E)
Launch Vehicles Supported: Naro-1
South Korea's new spaceport is operated by the Korea Aerospace Research Institute (KARI) to support the launch of the Naro 1 rocket, formally known as the KSLV 1. The first two attempts to launch the Naro-1 rocket have failed.

United States of America

Site Name: Kodiak Launch Complex
Location: Kodiak Island, Alaska (Latitude 67.5°N Longitude 146°W)
Launch Vehicles Supported: Athena, Minotaur I, Minotaur IV, Taurus, Taurus II (proposed)
This site is a commercial launch facility and because of its Northern Latitude is well suited for launching satellites into polar orbit. The launch site is located on the Narrow Cape, of Kodiak Island, Alaska, south of the city of Kodiak and some 400 km south of Anchorage. Kodiak Island is a volcanic peak in the ocean located some 48 km off shore in the Gulf of Alaska. This commercial spaceport is operated by the Alaska Aerospace Corporation.
Site Name Cape Canaveral Air Force Station (CCAFS)
Location: Cape Canaveral, Florida (Latitude 28.3°N Longitude 80.3°W)
Launch Vehicles Supported: Falcon 9, Atlas V, Delta IV
This facility is operated by the US Air Force and has been the site for many launches to support US space mission involved with national security. The CCAFS is an installation of the US Air Force Space Command 45th Space Wing that is headquartered at the adjacent Patrick Air Force Base. The CCAFS is the

primary launch head of America's Easter Range and currently the facility has four launchpads capable of supporting launches. The facility is south-southeast of the Kennedy Space Center and the two launch systems are linked by bridges and roadways. There is a 10,000-ft runway at the CCAFS that is close to the launch complexes. It is available for airlift aircraft that deliver heavy and particularly large payloads associated with launches from the Cape.

Several major American space exploration "firsts" were launched from CCAFS. These launches include the first US artificial satellite the Explorer I in 1958, the first US astronaut launches with Project Mercury and the follow on Gemini launches in the 1961–1964 time period. It has also supported other crewed systems launches as well as flights to explore the solar system.

Site Name: Kennedy Space Center (KSC)
Location: Merrit Island, Florida (Latitude 28.5°N Longitude 81.5°W)
Launch Vehicles Supported: Space Shuttle (retired)
This site has served as the location for processing, launching, and landing space shuttles and their payloads, including components of the International Space Station. It has historically served as the launch site for the Apollo launches. This launch center is essentially designed to support crewed missions to places beyond Earth but also has supported other launch missions as well. KSC is located on Merrit Island. It is adjacent to the US Air Force launch facilities known as the Cape Canaveral Air Station and the newly formed Space Florida launch facility. Kennedy was built first to support the Apollo lunar landings of the 1960s. After the last Apollo lunar launch in 1972, launch complex 39 supported Skylab space station in 1973–1974, then the Apollo-Soyuz Russian-American linkup in space in 1975, and now space shuttles since the late 1970s.

Site Name: The Mojave Spaceport
Location: California, USA (Latitude 35.0°N Longitude 118.2°W)
Launch Vehicles Supported: Various proposed horizontal takeoff spaceplanes
The Mojave Spaceport was the first commercial spaceport licensed by the US Federal Aviation Administration Office of Commercial Space Transportation (FAA-AST). This facility is limited to the horizontal takeoff and landing of spaceplane systems and reusable spacecraft. This facility is located 160 km north of Los Angeles, California. Commercial operations are maintained at this site by Scaled Composites, Inter-Orbital Systems (IOS), Orbital Sciences, and XCOR Aerospace.

Site Name: Spaceport America (formerly known as the Southwest Regional Spaceport)
Location: Las Cruces, New Mexico (Latitude 32°N Longitude 107°W)
Launch Vehicles Supported: Space Ship Two
This especially designed and purpose-build facility, located 72 km north of Las Cruces, New Mexico, was constructed in close proximity to the White Sands Missile Range. This facility is designed to support the spaceplane operations of Virgin Galactic and the Space Ship Two spaceplanes. The carrier vehicle will take off and carry Space Ship Two to altitude where the two vehicles will separate. Space Ship Two will land after a suborbital parabolic flight to a maximum height of 120 km. The alternative flight center is the Mojave Spaceport.

Site Name: Vandenberg Air Force Base (VAFB)
Location: Lompoc, California (Latitude 34.4°N Longitude 120.35°W)
Launch Vehicles Supported: Delta II, Delta IV, Atlas V, Minotaur I, Minotaur IV, Taurus, Pegasus, Falcon 1
This facility, which was first established in 1941 during World War II, is operated by the US Space Command's 30th Space Wing. It is located on the central Pacific coastline 20 km north of Lompoc, California, and 240 km northwest of Los Angeles. Vandenberg is the only military installation in the United States from which unmanned government and commercial satellites are launched into polar orbit. VAFB sends satellites to polar orbits by launching them due south. The base also test fires America's intercontinental ballistic missiles (ICBMs) westward toward the Kwajalein Atoll in the Marshall Islands. Vandenberg operates the Western Range tracking network, which extends all the way into the Indian Ocean to meet the Eastern Range tracking network. Western Range sites are on the California coast and downrange in the Hawaiian Islands. Vandenberg was to have provided a base for space shuttle launches on high inclination missions, but no shuttles ever have flown from there. California Spaceport is a colocated commercial launch facility at Vandenberg Air Force Base for launching satellites to polar orbit.
Site Name: Wallops Flight Facility (WFF)
Location: Wallops Island, Virginia (Latitude 37.8°N Longitude 75.5°W)
Launch Vehicles Supported: Pegasus, Minotaur I, Minotaur IV, Minotaur V (proposed) Taurus, Taurus II (proposed)
WFF is officially a part of the NASA Goddard Space Flight Center. Wallops Flight Facility is the main launch site for the vehicles owned and operated by Orbital Sciences Corporation. WFF also supports numerous sounding rockets and balloon launches.
Site Name: Ronald Reagan Ballistic Missile Defense Test Site
Location: Omelek Island, Kwajalein Atoll, Marshall Islands (Latitude 9.3°N Longitude 167.4°E)
Launch Vehicles Supported: Falcon 1, Falcon 9 (proposed)
The Reagan Test Site has been used by SpaceX Corporation, as a spaceport for launching its Falcon 1 launch vehicles since 2006. The site is beneficial for geostationary satellite launches when compared to other US launch sites due to its proximity to the Equator. There are proposals to upgrade the spaceport to support future Falcon 9 launches starting potentially in 2012.

Multinational Sites

Site Name: Sea Launch, Odyssey Platform
Location: Mobile, typically from Latitude 0°N to Longitude 154°W)
Launch Vehicles Supported: Zenit 3SL
Sea Launch is controlled by a consortium of aerospace companies and is headquartered at Long Beach, California. The consortium was headed by the Boeing Corporation until Sea Launch claimed bankruptcy in June 2009. Since

emerging from bankruptcy in 2010, Sea Launch's majority owner is Energia Overseas Limited. Sea Launch, which has been in existence since 1999, operates its floating launch platform from a position near Kiribati (popularly known as the Christmas Islands). This position, almost exactly on the equator, is ideal for the commercial launch of satellites that are to be deployed into a geosynchronous orbit. Sea Launch operations begin at the ship's home port at Long Beach, California, where a satellite is prepared and loaded onto Sea Launch Commander, the assembly and command ship. Then, a Zenit rocket in a horizontal position is transferred to an environmentally controlled hangar on Odyssey, the partially submersible, self-propelled, launch platform. Odyssey once was a North Sea oil drilling platform. It is 436 ft. long and 220 ft. wide. After sailing to a launch point in the Pacific, a rocket is rolled out onto Odyssey's deck, erected, and fueled with kerosene and liquid oxygen. The rockets fired from Sea Launch typically carry telecommunications satellites to space on their way to geostationary orbit.

Major Launch Systems Available Globally

A2

Arthur N. Guest and Joseph N. Pelton

Contents

China 1208
Long March 2C 1208
Long March 2C/CTS1 1208
Long March 2C/CTS2 1209
Long March 2D 1209
Long March 2F 1209
Long March 3A 1210
Long March 3B 1210
Long March 3BE 1210
Long March 3C 1210
Long March 4B 1211
Long March 4C 1211
Long March 5E (Under Development: Estimated Operational 2014) 1211
Europe 1212
Ariane 5ECA 1212
Ariane 5ES 1212
Vega 1212
India 1212
Polar Satellite Launch Vehicle (PSLV) 1212
Geosynchronous Satellite Launch Vehicle (GSLV) 1212
Geosynchronous Satellite Launch Vehicle (GSLV): Mark III (Proposed) 1213
Israel 1213
Shavit 2 1213
Japan 1213
H-IIA 202 1213
H-IIA 204 1213

A.N. Guest
International Space University, San Francisco, USA
e-mail: guest.arthur@gmail.com

J.N. Pelton (✉)
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: joepelton@verizon.net

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_89, © Springer Science+Business Media New York 2013

H-IIB 304 1214
South Korea 1214
Naro-1 (KSLV-I) 1214
Russia 1214
Angara 1.1 (Proposed) 1214
Angara 1.2 (Proposed) 1214
Proton M/Breeze M 1215
Rockot 1215
Shtil 1215
Volna 1215
Start-1 1215
Soyuz FG / Fregat 1215
Soyuz-2.1a/Soyuz-ST 1216
Ukraine 1216
Zenit2 1216
Zenit 3SL 1216
United States 1216
Athena I 1216
Athena II 1217
Atlas V 1217
Delta II 1217
Delta IV 1218
SpaceX Falcon 1e 1218
SpaceX Falcon 9 1219
Orbital Sciences Pegasus 1219
Orbital Sciences Minotaur I 1219
Orbital Sciences Minotaur IV 1219
Orbital Sciences Taurus 1219

China

Long March 2C

Height: 43 m
Lift-off Mass: 245 t
Lift-off Thrust: 2,962 kN
Fairing Diameter: 3.35 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
LEO Capability: 3,850 kg
SSO Capability: 900 kg
Launch Site(s): JSLC/TSLC/XSLC

Long March 2C/CTS1

Height: 43 m
Lift-off Mass: 245 t

Lift-off Thrust: 2,962 kN
Fairing Diameter: 3.35 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
Stage-3 Propellant: Solid Propellant
SSO Capability: 2,100 kg
Launch Site(s): JSLC/TSLC/XSLC

Long March 2C/CTS2

Height: 43 m
Lift-off Mass: 245 t
Lift-off Thrust: 2,962 kN
Fairing Diameter: 3.35 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
Stage-3 Propellant: Solid Propellant
GTO Capability: 1,250 kg
Launch Site(s): JSLC/TSLC/XSLC

Long March 2D

Height: 41 m
Lift-off Mass: 250 t
Lift-off Thrust: 2,962 kN
Fairing Diameter: 3.35 m/3.80 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
LEO Capability: 4,000 kg
SSO Capability: 1,150 kg
Launch Site(s): JSLC/TSLC

Long March 2F

Height: 58.3 m
Lift-off Mass: 497.9 t
Lift-off Thrust: 5,923 kN
Fairing Diameter: 3.80 m/4.20 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
LEO Capability: 8,600 kg
Launch Site(s): JSLC

Long March 3A

Height: 52.5 m
Lift-off Mass: 241 t
Lift-off Thrust: 2,962 kN
Fairing Diameter: 3.35 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
Stage-3 Propellant: LOX/LH_2
LEO Capability: 2,600 kg
Launch Site(s): XSLC

Long March 3B

Height: 54.8 m
Lift-off Mass: 425.8 t
Lift-off Thrust: 5,923 kN
Fairing Diameter: 4.00 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
Stage-3 Propellant: LOX/LH_2
LEO Capability: 5,100 kg
Launch Site(s): XSLC

Long March 3BE

Height: 56.3 m
Lift-off Mass: 456 t
Lift-off Thrust: 5,923 kN
Fairing Diameter: 4.00 m/4.20 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
Stage-3 Propellant: LOX/LH_2
LEO Capability: 5,100 kg
Launch Site(s): XSLC

Long March 3C

Height: 54.8 m
Lift-off Mass: 345 t
Lift-off Thrust: 4,443 kN
Fairing Diameter: 4.00 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH

Stage-3 Propellant: LOX/LH_2
LEO Capability: 3,800 kg
Launch Site(s): XSLC

Long March 4B

Height: 48.0 m
Lift-off Mass: 250 t
Lift-off Thrust: 2,962 kN
Fairing Diameter: 2.90 m/3.35 m/3.80 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
Stage-3 Propellant: N_2O_4/UDMH
SSO Capability: 2,230 kg
Launch Site(s): JSLC/TSLC

Long March 4C

Height: 48.0 m
Lift-off Mass: 250 t
Lift-off Thrust: 2,962 kN
Fairing Diameter: 2.90 m/3.35 m/3.80 m
Stage-1 Propellant: N_2O_4/UDMH
Stage-2 Propellant: N_2O_4/UDMH
Stage-3 Propellant: N_2O_4/UDMH
SSO Capability: 2,950 kg
Launch Site(s): JSLC/TSLC

Long March 5E (Under Development: Estimated Operational 2014)

Height: 62.0 m
Lift-off Mass: 800 t
Lift-off Thrust: 10,640 kN
Fairing Diameter: 2.90 m/3.35 m/3.80 m
Booster Propellant: LOX/Kerosene
Stage-1 Propellant: LOX/LH_2
Stage-2 Propellant: LOX/LH_2
LEO Capability: 25,000 kg
GTO Capability: 14,000 kg
Launch Site(s): WSLC
Note: Long March 5E is the largest of six variations being developed (LM-5A to LM-5F).

Europe

Ariane 5ECA

Height: 59.0 m
Lift-off Mass: 780 t
Fairing Diameter: 5.4 m
Booster Propellant: Solid Propellant
Stage-1 Propellant: LOX/LH_2
Stage-2 Propellant: LOX/LH_2
GTO Capability: 10,500 kg
Launch Site(s): Kourou

Ariane 5ES

Height: 59.0 m
Lift-off Mass: 780 t
Fairing Diameter: 5.4 m
Booster Propellant: Solid Propellant
Stage-1 Propellant: LOX/LH_2
Stage-2 Propellant: LOX/LH_2
LEO Capability: 21,500 kg
Launch Site(s): Kourou

Vega

700 km Polar Orbit Capability: 1,500 kg
Launch Site(s): Kourou
Note: First Flight: January 2012 (Proposed).

India

Polar Satellite Launch Vehicle (PSLV)

Height: 44 m
Lift-off Mass: 294 t
SSO Capability: 620 kg
GTO Capability: 1,050 kg
Launch Site(s): Satish Dhawan Space Center (SHAR)

Geosynchronous Satellite Launch Vehicle (GSLV)

Height: 49.0 m
Lift-off Mass: 414 t

GTO Capability: 2,500 kg
Launch Site(s): Satish Dhawan Space Center (SHAR)

Geosynchronous Satellite Launch Vehicle (GSLV): Mark III (Proposed)

Height: 42.4 m
Lift-off Mass: 630 t
GTO Capability: 4,500 kg
Launch Site(s): Satish Dhawan Space Center (SHAR)
First flight: 2012 (proposed)

Israel

Shavit 2

Height: 18.0 m
Lift-off Mass: 30 t
Polar LEO Capability: 300–800 kg
Launch Site(s): Palmachim Air Force Base

Japan

H-IIA 202

Height: 53.0 m
Lift-off Mass: 289 t
Fairing Diameter: 4 m
Booster Propellant: Solid Propellant
Stage-1 Propellant: LOX/LH_2
Stage-2 Propellant: LOX/LH_2
LEO Capability: 10,000 kg
GTO Capability: 3,700 kg
Launch Site(s): Tanegashima Space Center

H-IIA 204

Height: 53.0 m
Lift-off Mass: 289 t
Fairing Diameter: 4 m
Booster Propellant: Solid Propellant
Stage-1 Propellant: LOX/LH_2
Stage-2 Propellant: LOX/LH_2

GTO Capability: 6,000 kg
Launch Site(s): Tanegashima Space Center

H-IIB 304

Height: 57.0 m
Lift-off Mass: 530 t
Fairing Diameter: 5.2 m
Booster Propellant: Solid Propellant
Stage-1 Propellant: LOX/LH_2
Stage-2 Propellant: LOX/LH_2
GTO Capability: 9,000 kg
LEO Capability: 16,500 kg (for HTV orbit)
Launch Site(s): Tanegashima Space Center

South Korea

Naro-1 (KSLV-I)

Height: 33.0 m
Lift-off Mass: 140 t
Fairing Diameter: 3 m
Stage-1 Propellant: LOX/Kerosene
Stage-2 Propellant: Solid
LEO Capability: 300 kg
Launch Site(s): Naro Space Center
Note: First two flights unsuccessful. No successful flights as of October 2011.

Russia

Angara 1.1 (Proposed)

Height: 34.9 m
Lift-off Mass: 149 t
LEO Capability: 2,000 kg
Launch Site(s): Plesetsk Cosmodrome (proposed)
First flight: 2013 (proposed)

Angara 1.2 (Proposed)

Height: 41.5 m
Lift-off Mass: 171.5 t
LEO Capability: 3,700 kg

Launch Site(s): Plesetsk Cosmodrome (proposed)
First flight: Unknown. Post-2013 (proposed)
Note: There are larger versions of the Angara launcher proposed as well.

Proton M/Breeze M

Height: 53 m
Lift-off Mass: 690 t
LEO Capability: 22,000 kg
GTO Capability: 6,920 kg
GSO Capability: 3,250 kg
Launch Site(s): Baikonur Cosmodrome

Rockot

Height: 29.2 m
Lift-off Mass: 107 t
Fairing Diameter: 2.5 m
LEO Capability: 2,140 kg
Launch Site(s): Plesetsk Cosmodrome

Shtil

LEO Capability: 160 kg
Launch Site(s): Launched from a submarine (typically from Barents Sea)

Volna

LEO Capability: 100 kg
Launch Site(s): Launched from a submarine (typically from Barents Sea)

Start-1

Height: 22.7 m
Lift-off Mass: 47 t
Fairing Diameter: 1.8 m
SSO Capability: 350 kg
Launch Site(s): Plesetsk Cosmodrome

Soyuz FG / Fregat

Height: 49.5 m
Lift-off Mass: 305 t

LEO Capability: 7,800 kg
SSO Capability: 4,500 kg
Launch Site(s): Baikonur Cosmodrome

Soyuz-2.1a/Soyuz-ST

Height: 46.1 m
Lift-off Mass: 305 t
Fairing Diameter: 2.95 m
GTO Capability: 2,760 kg
SSO Capability: 4,500 kg
Launch Site(s): Baikonur Cosmodrome, Centre Spatial Guyanais

Ukraine

Zenit2

Height: 57 m
Lift-off Mass: 445 t
LEO Capability: 13,740 kg
Launch Site(s): Baikonur Cosmodrome

Zenit 3SL

Height: 59.6 m
Lift-off Mass: 462.2 t
GTO Capability: 6,100 kg
Launch Site(s): Sea Launch Odyssey Platform

United States

Athena I

Height: 19.8 m
Lift-off Mass: 66.3 t
LEO Capability: 820 kg
SSO Capability: 360 kg
GTO Capability: 13,740 kg
Launch Site(s): Cape Canaveral, Vandenberg, Kodiak
Note: Retired 2001, proposed to return to service 2012.

Athena II

Height: 30.5 m
Lift-off Mass: 120.2 t
LEO Capability: 2,065 kg
SSO Capability: 1,165 kg
GTO Capability: 590 kg
Launch Site(s): Cape Canaveral, Vandenberg, Kodiak
Note: Retired 2001, proposed to return to service 2012.

Atlas V

Note: There are 10 variations of the Atlas V. The naming convention for the Atlas V is Atlas V XYZ, where X is the diameter of the payload fairing (4 m or 5 m), Y is the number of SRB boosters (0,1,2,3,4,5), and Z is the number of engines on the Centaury core stage.

Atlas V Variation	401	411	421	431	501	511	521	531	541	551
GTO Capability (kg)	3,460	4,450	5,210	5,860	2,690	3,900	4,880	5,690	6,280	6,860

Launch Site(s): Cape Canaveral, Vandenberg

Delta II

Note: There are numerous variations of the Delta II. The naming convention for the Delta II is Delta II VWXY-Z, where V is the type of first stage engine (7 = RS-27A engine), W is the number of first stage solid rocket motors (9,4, or 3), X is the type of second stage (2 = Aerojet AJ10-118 K engine), Y is the type of third stage (0 = none, 0H = None, heavy configuration, 5 = Star-48B solid motor, 5H = Star-48B solid motor, heavy configuration, 6 = Star-37FM solid motor), and Z = the type of fairing (9.5 ft diameter, 10 ft diameter, or 10 ft diameter long).

Two-stage		Three-stage	
LEO capability (kg)		GTO capability (kg)	
Delta II 7320-9.5	2,809	Delta II 7326-9.5	934
Delta II 7320-10	2,703	Delta II 7326-10	898
Delta II 7420-9.5	3,185	Delta II 7425-9.5	1,110
Delta II 7420-10	3,099	Delta II 7425-10	1,073
Delta II 7920-9.5	5,030	Delta II 7426-9.5	1,058
Delta II 7920-10	4,844	Delta II 7426-10	1,029
Delta II 7920-10L	4,805	Delta II 7925–9.5	1,819
Delta II 7920H-9.5	6,097	Delta II 7925-10	1,747
Delta II 7920H-10	5,959	Delta II 7925-10L	1,739

(*continued*)

Two-stage		Three-stage	
LEO capability (kg)		GTO capability (kg)	
Delta II 7920H-10L	5,899	Delta II 7926-9.5	1,660
		Delta II 7926-10	1,581
		Delta II 7926-10L	1,578
		Delta II 7925H-9.5	2,171
		Delta II 7925H-10	2,123
		Delta II 7925H-10L	2,102
		Delta II 7926H-9.5	1,981
		Delta II 7926H-10	1,934
		Delta II 7926H-10L	1,916

Launch Site(s): Cape Canaveral, Vandenberg

Delta IV

Note: There are five variations of the Delta IV.

	Number of common core boosters (first stage)	Number of solid boosters	Payload fairing diameter (m)
Delta IV medium	1	0	4
Delta IV medium + (4,2)	1	2	4
Delta IV medium + (5,2)	1	2	5
Delta IV medium + (5,4)	1	4	5
Delta IV heavy	3	0	5
	Payload capabilities (kg)		
	GEO	GTO	LEO
Delta IV medium	1,348	4,508	9,390
Delta IV medium + (4,2)	2,208	6,200	12,477
Delta IV medium + (5,2)	2,105	5,124	11,062
Delta IV medium + (5,4)	3,116	6,905	13,774
Delta IV heavy	6,573	13,248	22,977

Launch Site(s): Cape Canaveral, Vandenberg

SpaceX Falcon 1e

Height: 24.7 m
Lift-off Mass: 35.2 t
Fairing Diameter: 1.7 m
LEO Capability: 1,010 kg
Launch Site(s): Omelek Island

SpaceX Falcon 9

Height: 54.9 m
Lift-off Mass: 333 t
Fairing Diameter: 3.6 m
LEO Capability: 10,450 kg
SSO Capability: 8,560 kg
GTO Capability: 4,600 kg
Launch Site(s): Omelek Island, Cape Canaveral

Orbital Sciences Pegasus

LEO Capability: 450 kg
SSO Capability: 300 kg
Launch Site(s): Launched from an aircraft. Sites include Wallops, Canary Island, Omelek Island, Cape Canaveral, Vandenberg

Orbital Sciences Minotaur I

Height: 19.2 m
Fairing Diameter: 1.27 m
LEO Capability: 580 kg
Launch Site(s): Vandenberg, Kodiak, Cape Canaveral, WFF

Orbital Sciences Minotaur IV

Fairing Diameter: 2.05 m
SSO Capability: 900 kg
LEO Capability: 1,900 kg
Launch Site(s): Vandenberg, Kodiak, Cape Canaveral, WFF

Orbital Sciences Taurus

Height: 32 m
Lift-Off Mass: 77 t
LEO Capability: 700–1,050 kg (depending on configuration)
Launch Site(s): Vandenberg, Kodiak, Cape Canaveral, WFF

Global Communications Satellite Systems

A3a

Joseph N. Pelton

International Communications Satellite Systems[a]			
Name of Satellite System	Location of Headquarters	Service provided	Web site
Eutelsat	Paris, France	FSS and BSS	www.eutelsat.com
Globalstar	Milpitas, California, USA	MSS	www.globalstar.com
Inmarsat	London, UK	MSS	www.inmarsat.com
Intelsat	Luxembourg	FSS	www.intelsat.com
			www.intelsatgeneral.com
Intersputnik	Moscow, Russia	FSS and BSS	http://www.intersputnik.com/satellites/00023/
Iridium	Chandler, Arizona, USA	MSS	www.iridium.com
O3b (Not yet deployed)	Paris, France	FSS – broadband internet	www.O3b.com
Orbcomm	Fort Lee, New Jersey, USA	Store and forward data	www.orbcomm.com
SES Global – SES World Skies	Luxembourg	FSS, BSS, and DTH TV	http://www.ses-worldskies.com
Regional Communications Satellite System[a]			
Name of Satellite System	Location of System Headquarters	Service provided	Web site
Amos	Tel Aviv, Israel (Operated by Spacecom)	FSS and DTH TV	www.amos-spacecom.com/content.cfm
AP Star	Hong Kong, China (Operated by APT Satellite)	FSS and DTH TV	http://www.apstar.com/apt_apstar
Arabsat	Riyadh, Saudi Arabia	FSS and DTH TV	http://www.arabsat.com

(*continued*)

J.N. Pelton
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: joepelton@verizon.net

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_90, © Springer Science+Business Media New York 2013

Regional Communications Satellite System[a]			
Name of Satellite System	Location of System Headquarters	Service provided	Web site
Asia Broadcasting Satellite System (*formerly Agila*)	Bermuda, UK	BSS/DTH TV	http://www.absatellite.net/about/index.html
Asia Cellular Satellite (*formerly Garuda Satellite*)	London, UK (Managed by Inmarsat)	MSS plus terrestrial cellular	www.inmarsat.com www.acesinternational.com/corporate
Asia Sat	Causeway Bay, Hong Kong	FSS and DTH TV	http://www.asiasat.com/
GE Satellite	Bethesda, Maryland, USA	FSS and DTH TV	www.gesatellite.com
Hispasat	Madrid, Spain	FSS and DTH TV	www.hispasat.com
Hylas	London, UK (Operated by Avanti Corporation)	FSS/Broadband Data	www.avantiplc.com
ICO Global Communications	Bellevue, Washington, USA	MSS	www.ico.com
IP Star	Bangkok, Thailand	FSS	www.ipstar.com
JSAT and Horizon	Tokyo, Japan	FSS	http://www.jsati.com/satellite-services.asp
Optus	Sydney, Australia	FSS and DTH TV	http://www.optus.com
Telenor	Oslo, Norway	FSS and BSS	http://www.telenor.com/
Telstar 11, 12, 14 & 18	Palo Alto, California USA (Operated by Space Systems/Loral)	FSS and Direct to the Home Broadcast	http://www.telesat.ca/en/Telstar_Fleet
Thuraya	Abu Dhabi, United Arab Emirates	MSS	http://www.thuraya.com
Worldspace	Yazmi Holdings, Maryland, USA	DABS	biz.yahoo.com/e/100629/wrspq.pk8-k.html
XTAR	Palo Alto, California, USA, and Madrid, Spain	FSS in X-band	www.xtar.com
Yahsat	Abu Dhabi, United Arab Emirates	FSS Services	www.yahsat.com
Domestic Communications Satellite Systems[a]			
Country	Name of satellite(s)	Type of service	Web site for latest system information
Argentina	*ARSAT-1*	FSS and DTH TV (Pending service shortly)	http://www.arsat.com.ar/ingles/home_ing.html
Australia	*Optus*	FSS and DTH TV	http://www.optus.com

(*continued*)

Domestic Communications Satellite Systems[a]			
Country	Name of satellite(s)	Type of service	Web site for latest system information
Brazil	*Brazilsat*	FSS and DTH TV	http://www.embratel.com.br/
Canada	*Anik, Nimiq DBS, Telstar*	*Anik* – FSS to Canada and North America	http://www.telesat.ca
		Nimiq – BSS service to Canada	
		Telstar – FSS to Canada and Atlantic region	
China	*ChinaSat-6*	Sinosat: FSS	http://www.fas.org/spp/guide/china/comm/chinasat.htm
	Sinosat	Chinasat: BSS	info@sinosat.com.cn
Egypt	*Nilesat 1 & 2*	FSS	http://www.nilesat.com.eg/services.htm
Greece	*Nova Greece*	BSS	www.satsig.net/ivsat-europe.htm
India	*Insat*	FSS and BSS	
Iran	*Mesbah*	Store and Forward data	www.universetoday.com/24552/iran-launches-satellite-into-orbit
Israel	*Amos*	FSS and DTH TV	http://www.amos-spacecom.com/
Japan	*B-Sat*	BSS	www.b-sat.co.jp/
	JSat	FSS and BSS	http://www.jsati.com/satellite-services.asp
	N-star	MSS	http://www.nttdocomo.com/services/miscellaneous/index.html
Korea	Koreasat	FSS and BSS	www.lyngsat.com/South-Korea.html
Luxembourg	*ASTRA*	FSS and DTH TV	www.ses-astra.com
Malaysia	*Measat*	FSS and DTH TV	http://www.measat.com/satellite.html
Mexico	*SatMex*	FSS	http://www.satmex.com.mx/index1.php
	Solidaridad Satellites	BSS	http://space.skyrocket.de/doc_sdat/solidaridad-1.htm
Nigeria	*Nigcomsat* (System failure: Is being replaced)	BSS	http://www.nigerianbestforum.com/index.php
Norway	*Telenor*	FSS and BSS	www.highbeam.com/doc/1G1-110308909.html

(*continued*)

Domestic Communications Satellite Systems[a]			
Country	Name of satellite(s)	Type of service	Web site for latest system information
Pakistan	*Paksat-1*	FSS and DTH TV	www.suparco.gov.pk/pages/paksat1.asp
Russia	*Express*	FSS and DTH TV	www.intersputnik.com
	Intersputnik		
	Stationar		
	Yamal		
Spain	*Hispasat*	FSS and DTH TV	www.hispasat.com/
Thailand	*Thaicom*	FSS and DTH TV	www.thaicom.net
Turkey	*Turksat*	FSS and DTH TV	www.ts2.pl/en/Satellite-Internet/Turkey
United Arab Emirates	*Yahsat*	FSS and DTH TV	www.yahsat.ae
United Kingdom	*Sky Television*	BSS	www.sky.com
United States	DBSD	Hybrid MSS	http://www.ico.com/_about/
United States	Direct TV	BSS	www.directv.com
United States	Dish (Echostar)	BSS	www.dish.com www.echostar.com
United States	Galaxy (Intelsat)	FSS and DTH TV	www.intelsat.com
United States	GE SAT	FSS and DTH TV	www.gesat.com
United States	G Star	FSS and DTH TV	www.geo-orbit.org/westhemipgs/fgstar4p.html
United States	Hughes Net	FSS and broadband data	www.hns.com
United States	JSAT Horizon	FSS and DTH TV	www.jsati.com
United States	Light Squared Network	Hybrid MSS	www.lightsquared.com
United States	SES Americom	FSS and DTH TV	www.ses-americom.com
United States	Terrestar	Hybrid MSS	www.terrestar.com
United States	Wild Blue	FSS and DTH TV	www.wildblue.com
United States	XM-Sirius	DABS	www.xmradio.com

[a]This chart was prepared and copyrighted by Joseph N. Pelton. Copyright permission is granted to Springer Press for publication in the *Handbook for Satellite Applications*. Please check with the relevant web site to obtain the latest information

US Domestic Communications Satellite Systems

A3b

Joseph N. Pelton

Name of system (and HQ location)	Operator	Type of service	Web site for latest information
DBSD (Bellevue, Washington and Reston, VA, USA)	Former ICO Global Communications Subsidiary (one satellite plus terrestrial mobile) (Now owned by DirecTV)	Hybrid Terrestrial-Satellite Land Mobile Service	www.DirecTV.com/
DirecTV	DirecTV	Broadcast Satellite Service-DBS	www.directv.com
Dish	Dish (Echostar)	Broadcast Satellite Service-DBS	www.echostar.com www.dish.com
Galaxy and Intelsat Americas Sats	Intelsat	Fixed and DTH Service	www.Intelsat.com
GE Sat	General Electric	Fixed Satellite and DTH Service	www.gesat.com
G Star	GTE & SES Americom	Fixed Satellite and DTH Service	www.geo-orbit.org/westhemipgs/fgstar4p.html
Hughes Net	Hughes Network Systems	IP-based broadband Data Services	www.hns.com
JSAT (Horizon)	Japan Satellite	DTH and FSS Services	www.jsati.com/news.asp?DocID=180

(*continued*)

J.N. Pelton
Former Dean, International Space University, Arlington, Virginia, USA
e-mail: joepelton@verizon.net

J.N. Pelton, S. Madry, S. Camacho-Lara (eds.), *Handbook of Satellite Applications*,
DOI 10.1007/978-1-4419-7671-0_1, © Springer Science+Business Media New York 2013

Name of system (and HQ location)	Operator	Type of service	Web site for latest information
LightSquared (formerly Mobile Satellite Ventures)	LightSquared Network	Hybrid Mobile Satellite and Terrestrial Cellular Service	www.lightsquared.com/
SES Americom	SES Global	DTH Satellite and Fixed Satellite Service	www.ses-americom.com
Sirius (Now part of XM Radio)	XM Radio Inc.	Direct Broadcast Radio	www.xmradio.com/bestofsirius/index.xmc
Terrestar	TerreStar Corporation (Now owned by DirecTV)	Hybrid Mobile Satellite and Terrestrial Cellular Service	www.DirecTV.com/
Wild Blue (Formerly Ka star)	Wild Blue Communications Inc.	Fixed Satellite and DTH Service	www.wildblue.com
XM Radio	XM Radio Inc.	Direct Broadcast Radio	www.xmradio.com

Table prepared and copyrighted by Joseph N. Pelton, printed by his permission